THE PONTIFICAL ACADEMY OF SCIENCES

Scripta Varia

106

INTERACTIONS BETWEEN GLOBAL CHANGE AND HUMAN HEALTH

VATICAN CITY
2006

Working Group
31 October - 2 November 2004

INTERACTIONS BETWEEN GLOBAL CHANGE AND HUMAN HEALTH

Address:

The Pontifical Academy of Sciences
Casina Pio IV, 00120 Vatican City
Tel: +39 0669883195 / Fax: +39 0669885218
Email: academy.sciences@acdscience.va

PONTIFICIAE ACADEMIAE SCIENTIARVM SCRIPTA VARIA
106

INTERACTIONS BETWEEN GLOBAL CHANGE AND HUMAN HEALTH

31 October - 2 November 2004

EX AEDIBVS ACADEMICIS IN CIVITATE VATICANA

MMVI

The opinions expressed with absolute freedom during the presentation of the papers of this meeting, although published by the Academy, represent only the points of view of the participants and not those of the Academy.

ISBN 88-7761-085-9

Pope Benedict XVI

The Participants of the Working Group of 31 October - 2 November 2004

The Participants of the Working Group of 31 October - 2 November 2004

The Academy or *The School of Athens* by Raphael, in the Vatican Palace
'In those people you will have recognised your oldest predecessors in the investigation of both matter and spirit'
(Pius XII, Address to the Plenary Session of the Academy, 3 December 1939)

CONTENTS

Scientific Papers

OVERVIEW

'The workshop will have as its goal to identify the connections and feedbacks by which the various aspects of Global Change can affect human health (with a focus on infectious disease), and the potential mechanisms by which disease events may influence the biogeophysical environment.

As humanity is entering the 21st century, three issues related to our well-being figure prominently in the public concern: Socio-economic development, adverse changes in the environment, and human and animal health. Each of these issues is written about in the media, each has a community of scientists researching and discussing it, and each is the subject of national and international political activity. Yet, while they have usually been discussed as separate issues, they are really components of a coupled system.

The workshop has as its central purpose to analyze the feedbacks and interactions between the three components of this system. The challenge will be to hypothesize on so far unsuspected links between the three components'.

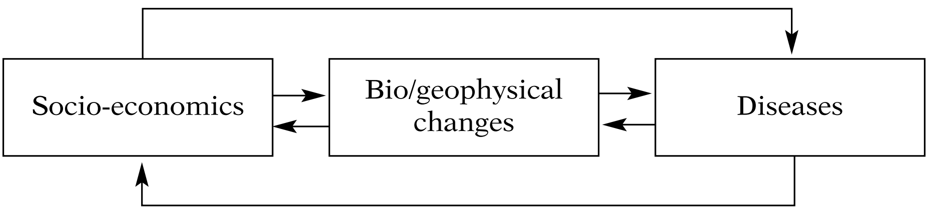
Prof. Paul J. Crutzen

INTRODUCTION

The Pontifical Academy of Sciences is sponsoring a workshop on 'Interactions between Global Change and Human Health', which will take place 31 October - 2 November 2004. This workshop will have as its goal to identify the connections and feedbacks by which the various aspects of Global Change can affect human health (with a focus on infectious disease), and the potential mechanisms by which disease events may influence the biogeophysical environment.

Workshop Goals

As humanity is entering the 21st century, three issues related to our well-being figure prominently in the public concern: Socio-economic development, adverse changes in the environment, and human and animal health. Each of these issues is written about in the media, each has a community of scientists researching and discussing it, and each is the subject of national and international political activity. Yet, while they have usually been discussed as separate issues, they are really components of a coupled system. The workshop has as its central purpose to analyze the feedbacks and interactions between the three components of this system:



The challenge will be to hypothesize on so far unsuspected links between the three components. For this purpose, the workshop will bring together experts from the three fields, carefully chosen not only for their knowledge in their own field, but also for their ability to look across boundaries of scientific disciplines and communities.

Developments in Socio-economics, Health and Environment

Socio-economic Development. It is widely acknowledged that sustainable socio-economic development is the shortest way for the promotion of well-being and good health. Such development is characterized by a healthy economy and low inflation, full employment and high incomes, efficient social security and assistance, good educational systems, adequate public infrastructure, increased material production and access to goods, and political freedom and stability. (Furthermore, a comprehensive approach to sustainable development requires that these social changes occur within the constraints of the biosphere – i.e., are in accord with the criteria of ecological sustainability). However, at the global level, this process has been an uneven achievement, and major inequalities are persisting and even increasing among regions and countries. Various factors resulting from social development (advances in medical technology, increased health awareness and efficient transmission of data and information, large-scale industrial production, and the expansion of health and humanitarian care) are available globally, at least in principle, and should enhance disease control. In reality, however, these factors have actually neither prevented the spread nor mitigated the impacts of infectious diseases worldwide. Not only the poor countries are vulnerable to the emerging/ resurging infections, but several infectious processes have emerged or have spread uncontrolled in developed countries: SARS; AIDS; West Nile Virus; Lyme disease; hantavirus pulmonary syndrome; hemolytic-uremic syndrome (E. coli H7O157) and methicillin-resistent Staphylococcus aureus. This is a clear indication that the interactions of factors other than those usually associated with poverty also play an important role in this process.

Human Health. Although infectious and parasitic diseases have been part of human life for thousands of years, as the paleopathological records have shown, the current rates of their emergence and spread and the magnitude of their impacts are unprecedented in history. At least 30 infectious agents either became known as new human pathogens and/or have increased globally in the past 25 years. Humans are spreading into the last corners of the tropical forest, where they are encountering diseases for which they have not evolved resistance. Existing national and international surveillance systems and control programs were not able to detect the emergence of new diseases quickly enough to prevent their spread. There is

evidence that this threatening prospect arises from a complex association of phenomena of different types: increasing population densities and mobility; rapid and long-distance trade; decreased biological resistance of the human host (psychological stress; malnutrition; chronic diseases and ageing; drugs); failure of public health systems; fast technological change; feeding habits; large scale environmental modifications, and genetic changes in microorganisms, to name but a few.

Global (Bio/geophysical) Change has in the public discussion been used almost synonymously with Climate Change. The issue is much broader, however. Over the past two centuries, the tremendous growth of the human population and the high resource demand of technologically developed societies have made humanity a geochemical and geophysical force, able to compete with Nature's forces and to threaten the functioning of the Earth System. Human activities are changing the composition of the biosphere, atmosphere and hydrosphere, are affecting global climate, and may even perturb the main circulation patterns of the Ocean. Because of the numerous feedbacks and teleconnections in the Earth System, the changes resulting from such perturbations are likely to be non-linear and may contain abrupt discontinuities. Examples are sudden changes in atmospheric composition (e.g., the Ozone Hole), the collapse of the Amazon forest, or the breakdown of the Gulf-Stream circulation in the North Atlantic.

Examples of Linkages Between the Three Components

Diseases can cause massive disruptions in human societies. History is full of examples: syphilis, malaria, plague, cholera, Spanish influenza, tuberculosis, and AIDS are among the most well known. Others loom on the horizon: dengue fever, Ebola, Hantavirus, west Nile fever, SARS, etc. In many instances, disease outbreaks have destroyed societies to the point where they were not able to recover. In a less globally-connected world, such effects were more or less regionally contained and could not spread globally. Nowadays, in a contiguous worldwide human population with high connectivities and few barriers to transmission, infectious agents can move about quickly. At the same time, socio-economic effects can propagate worldwide in a very short time.

Disease outbreaks do not arrive out of nowhere – in most cases agents and vectors have already been present, but specific environmental conditions and social-environmental characteristics were required before an epi-

demic or pandemic could occur. Often these are suggestive, but hard to pin down. Was the socioeconomic disruption and movement of people in World War I a prerequisite for the 1918 Spanish Flu pandemic that is thought to have cost 20-40 million lives worldwide? What were the linkages between trade and plague, travel and SARS? High population densities in close contact with animal reservoirs of infectious disease make possible the rapid exchange of genetic material. Malnutrition is of epidemic proportions in many developing countries, providing large immune-compromised populations that diseases can spread into very rapidly. Climate change may play a similar role: Warmer and wetter climate conditions may facilitate the spread of diseases, such as malaria and dengue fever.

What is the nature and importance of the environmental impacts associated with large-scale disease outbreaks? Human activity and the global environment have become inseparable, and consequently the future of the bio/geophysical Earth System is dependent on human stewardship. Conversely, the human economy and society is totally dependent on the functioning of the life support systems of the Planet to provide clean air and water, nutrition, a stable climate, etc. What would be the consequences of the economic losses, changes in human behavior, abrupt demographic change, etc. for the way humanity manages its environment? How would this influence climate change scenarios?

At present, we are already witnessing one way in which human health issues affect climate change: the concern about the health effects of pollution aerosols from power plants etc. is leading to much more efforts to control emissions than had been anticipated. Since these aerosols would have a cooling effect on climate, such emission reductions may lead to accelerated climate warming over the next century.

While the interactions between infectious disease and climate change will be the focus of this workshop, we should not fail to mention that climate change, and especially the extreme climate events associated with it, are already taking a toll in human lives, as exemplified in the substantial number of excess deaths associated with the extreme heat period in Europe during the summer of 2003.

Looking into the Future

Scientists in socioeconomics, health, environmental and climate science have all evolved scenarios for the future. In some cases they have been combined, such as in the climate change scenarios of the Intergovernmental

Panel of Climate Change. Many potential scenario parameters and feedback loops are, however, missing.

What are the prospects for the occurrence of large disease events now and in the near future? New diseases are showing up all the time, and old ones may be staging a comeback. Do we know enough about influenza and its socioeconomic drivers to anticipate a recurrence of the 1918 flu epidemic, with the potential of tens of millions of fatalities? Is society better prepared if such an outbreak occurs? Medical science may be more advanced now than in 1918, but what would be the chances to deliver its benefits to 6 billion people fast enough?

Socioeconomic behavior is as much driven by perception and psychology as by 'hard facts'. The perception of an impending pandemic may lead to rapid economic collapse in a world economy dependent on fast global exchange of goods and services – see the recent economic effects of a very few cases of SARS! On the other hand, the ongoing socioeconomic devastation by AIDS in Africa has led to relatively little action, even in some of the most affected countries.

Global change is difficult to measure – this applies to climate change, but even more so to sustainable development in socioeconomics, environment and health. What are the current indicators of well-being that could be considered markers of global vulnerability to the emergence and impacts of these outbreaks? What are indicators of potential breakpoints and thresholds in the system? Could a measure of population health serve as an index of sustainable development in a more general sense?

Prevention and Mitigation

Finally, there should be no diagnosis without at least a suggestion of therapy. What can be done to prevent health disasters from occurring or to mitigate their effects? How can we manage the Earth environment to minimize the threat to human health? What are the available resources, on a global scale, to cope with these threats, from the scientific, economic, institutional and political perspectives? These and other emerging questions will be discussed at the workshop.

HOMILY
GLOBAL INTERCHANGE AND HUMAN HEALTH

JAVIER CARD. LOZANO BARRAGÁN

Lk 19, 1-10

I thank the Academy of Sciences and its President, His Excellency Msgr. Marcelo Sánchez Sorondo, for the invitation to preside at this Eucharistic celebration.

The Gospel reading about Zacchaeus, invites us to reflect on the interchange of goods. Zacchaeus promised Christ that he would distribute all his possessions in a right way and would not defraud any more. In the context of today's work of the Academy of Sciences, that is, studying the different problems of globalisation, Jesus invites us to take similar measures. The question is, how can this be done in the sector of health and at the level of global health?

Just as globalisation concerns the political and legal fields so does it also affect the field of health. Globalisation that is only based on economics reduces the well-being of people, because it limits prosperity by concentrating it in the hands of the privileged few; pretending at the same time that there is no other way or that other democratic ways are not possible. Within the logic based on financial gains, some people would rather relegate health to a private issue, because they think it falls within social needs that do not concern economic globalisation. Hence, the idea that health care must be something private and not the concern of the State, because the public care of health will decrease the earnings. This is an approach that protects only the rich and not the poor, especially at country levels. However, such a position is totally wrong as can be seen even from the very nature of infectious diseases. There is a 'microbial' unification of the world. Infectious diseases are quickly spread everywhere by the growing mobility of populations, and this affects both rich and poor

countries. Moreover, it is not possible to fight them without a global action by all, especially the rich countries.

Today 50% of the deaths in poor countries, of people between 0 and 40 years of age, are attributed to the following diseases: 13% is due to AIDS, 11% is due to perinatal conditions, 9% is due to stomach illnesses, 6% is due to malaria, 5% is due to smallpox, and 4% is due to tuberculosis. Then the other 50% is attributed to other causes. We must realise that the UN agencies such as the WHO, UNESCO, UNAIDS, UNICEF, the World Bank and several NGOs are trying to combat these diseases. It is therefore clear that diseases as well as drugs are embraced into globalisation. Furthermore, one has to note that today economic globalisation itself endangers the health of people, because it creates an enormous gap between the rich and the poor, both at the international level between countries and at the national level within the same country.

Following the example of Zacchaues in today's Gospel reading, we must distribute the defrauded goods of health among the whole world population. How can this be done?

In the WHO, three priority areas for the globalisation of health for all are identified: 1) The need for a more effective global governance that is capable of generating consensus and political decisions that correspond to the real needs of the population in the field of health; 2) The need to generate and spread knowledge, making it more available so as to inform both the decision-makers and the general population about their health; 3) Support globalisation in the area of health and promote action at local and national levels, in order to ensure better health especially for those that are excluded in the current globalisation.

As practical principles for a worldwide action for the protection of health at the national and local levels, the WHO proposes the following: the elaboration of a scientific and factual approach to sick people, together with the participation of the population in the decision-making; the compromise of assuming health strategies based on sustainable development; the application of the global approach by the people in health development; the need to respect the specificity of each sex and promote the quality of life; the application of flexible strategies that are appropriate to permanent change; to be conscious of the relationship between cost and efficacy; and as a prior condition, that the new strategies match up with the diversity of values and cultural norms of the different peoples.

The WHO affirms that, in order to reduce the incidence of sickness, the fight against the physical and mental problems caused by morbidity

must be conducted within the framework of a sustainable development that is centred on the human being. This implies the promotion of macroeconomics and social politics founded on equity, according to the principle of cost and efficacy, not directed to protected groups, and with appropriate measures to defend the most vulnerable people.

In any case, we have to affirm that health is part of globalisation especially due to four factors: the first is the homogenisation of work. With the globalisation of work, we find ourselves in a world where you have the same working conditions in the various companies in all countries, be it in developed or developing countries. Just as work is homogenous so is the harm it causes and so are the illnesses and the restraint on health. The second factor is that of migration. When people migrate, they migrate with their illnesses, such that infectious diseases are not exclusively limited to the developing world, they are present everywhere. The third factor is that of business in health services. In the past we used to have special places or institutions in the developed world that would deal with serious problems of health. Nowadays, these services are exported to the developing world, with the same quality at a reduced cost, rendering major income to health enterprises. And finally, the fourth factor is the globalisation of the national health systems of various countries, since problems are almost the same everywhere.

Furthermore, infectious diseases are also spread everywhere by food trade. On the other hand, the globalisation of health care systems enables people to have a longer life expectancy. This offers occasion for the agglomeration of people in residential centres, which may be a favourable environment for the spread of infectious diseases, like AIDS. Yet we must not forget that the abuse of antibiotics generates resistance to them, thus rendering some of them useless.

It is clear that in the health sector, economic globalisation alone, which is ruled by the laws of the market whose only goal is profit, is unjust and does not produce well being. The predominance of business concerns in health management jeopardizes health policies, and when the latter are not guided by cultural values they are also shattered. The challenge then is to promote cultural values not through an artificial consensus, but through a consensus based on the respect for the dignity of the human nature. We shall therefore be able to build a true and beneficial globalisation of health, in a solidarity that recognizes the economical, social, political and cultural diversity in the world.

Globalisation can therefore be evangelised in this way. Like Zacchaeus we must insist on the globalisation of gratuity; not promote health only for

profit, but health for the free love of the Lord. Christ dead and risen is the central message from which we must proceed to give the true sense of illness, suffering and pain. Therefore, health for all, the true globalisation of health has to be based on the appreciation of life as a tension towards the resurrection of the Lord. And so the answer of the Church to the globalisation of health is creating a new solidarity in Christ, one that enlightens all the fields of human culture and economics. This position takes on a special meaning when it comes to speaking about the health of the sick people.

May this Eucharist, in which we celebrate the fullness of health in the death and resurrection of Christ, generate in the world the authentic globalisation of health in keeping with the sacrifice of Jesus. In this way we can go beyond profit and engender a profound culture of solidarity in a globalised health for all, especially the poor people. Just two months ago, Pope John Paul II created a new Foundation, known as the 'Good Samaritan Foundation', which is meant to help the sick people in the world who are in great need, especially due to the AIDS disease. May I suggest to the globalised economy and the health sector to let the Lord Jesus visit their houses, so that He may inspire all those responsible to compensate all who have been defrauded. Like in the case of Zacchaeus, it is time that salvation enters into this house. Amen.

Vatican City, 31 October, 2004

PROGRAMME

SUNDAY 31 OCTOBER 2004

9:00 Holy Mass celebrated by H.E. Cardinal Javier LOZANO BARRAGÁN (Altar of St Peter's Tomb)
9:45 *Coffee Break*

Speakers:
10:30 Prof. Meinrat O. ANDREAE
Welcome and Introduction to the Theme of the Workshop
11:20 Prof. William F. RUDDIMAN
Pre-Industrial Depopulation, Atmospheric Carbon Dioxide, and Global Climate
11:45 Prof. Veerabhadran RAMANATHAN
Aerosol Effects on Environment, Climate and Health
12:15 Discussion
12:45 *Lunch at the Academy*

Speakers:
14:40 Dr. Michel H. MEYBECK
Global Changes on Aquatic Systems and Their Interrelations with Human Health
15:00 Dr. Will STEFFEN
Impacts of Human Activities on the Functioning of the Biogeophysical Earth System
16:00 Discussion
16:30 *Coffee Break*

Speakers:
17:00 Prof. Andrew P. DOBSON
Interactions of Global Biodiversity Changes with Climate and Health
17:30 Prof. Ann G. CARMICHAEL
Infectious Disease and Human Impacts on the Environment – An Historical Perspective

18:20 Prof. Ulisses E. CONFALONIERI
Social-Environmental Vulnerability to Infections
18:50 Discussion
19:00 *Dinner at Domus Sanctae Marthae*

MONDAY 1 NOVEMBER 2004

Speakers:
9:10 Prof. Mary E. WILSON
Globalization of Infectious Diseases
9:40 Prof. Anthony J. MCMICHAEL
Interactions Between Infectious Disease
and the Biogeophysical and Socioeconomic Environment
10:10 Dr. Mahendra M. SHAH
Food, Water, and Infectious Diseases
10:40 *Coffee Break*
11:10 Discussion

Speakers:
11:40 Dr. Wolfgang LUTZ
Interactions of Demographic Trends and Human Health
12:10 Prof. Pim MARTENS
Consequences of Changes in Human Health
on Scenarios for the 21st Century
12:40 *Lunch at the Academy*

Speaker:
14:20 Dr. David HEYMANN
Effect of Social, Economic and Environmental Factors
on Current and Future Patterns of Infectious Disease
14:50 Discussion

Speakers:
15:20 Dr. David BLOOM
Interactions between Economics and Epidemics/Pandemics
15:45 Prof. Carlo JAEGER
Global Governance for Human Health
and a Benign Biogeophysical Environment
16:20 *Coffee Break*

16:45 Prof. Wolfgang SACHS
Climate Change and Human Rights
17:20 Discussion
17:45 Agenda Discussion to finalize Group Discussion topics.
Group Discussion
What evidence do we have for strong interactions/feedbacks between environment/socioeconomics/health in the past?
19:30 *Dinner at Domus Sanctae Marthae*

TUESDAY 2 NOVEMBER 2004

8:10 Group Discussion:
Socioeconomic and environmental drivers of large disease events
10:30 *Coffee Break*
10:50 Ismael Clark Arxer (Ambassador of Cuba)
Promoting Human Health and Facing Global Changes: The Cuban Experience
11:10 Group Discussion:
Indicators and measures of vulnerability of human populations. Future generations of models
12:45 *Lunch at the Academy*
14:15 Group Discussion:
Policies and responses to avoid undesirable ESH interactions
15:45 *Coffee Break*
16:15 Group Discussion:
Scenarios for the 21st century: How will expected global change and socioeconomic development affect human health, specifically with regard to large infectious disease events? How can human health events affect the world's socioeconomic and environmental development?
18:00 Final Group Discussion
19:00 *Dinner at Domus Sanctae Marthae*

LIST OF PARTICIPANTS

Academicians

H.E. Msgr. Prof. Marcelo SÁNCHEZ SORONDO, Chancellor
The Pontifical Academy of Sciences
Casina Pio IV
V-00120 Vatican City

Prof. Paul J. CRUTZEN, Organizer
Max-Planck-Institute for Chemistry
Department of Atmospheric Chemistry
P.O. Box 3060
D-55020 Mainz
(Federal Republic of Germany)

Prof. Rudolf MURADIAN
Universidade Estadual de Santa Cruz
Departamento de Ciências Exatas e Tecnológicas
Rodovia Ilhéus/Itabuna, km 16
45650-000 Ilhéus, Bahia
(Brazil)

Prof. Crodowaldo PAVAN
Universidade de São Paulo, USP
Instituto de Ciências Biomédicas, Lab. de Microbiologia
São Paulo, S.P. 05389-970
(Brazil)

Dr. Veerabhadran (Ram) RAMANATHAN
University of California, San Diego
Scripps Institution of Oceanography
Center for Atmospheric Sciences
9500 Gilman Drive, MC 0221
La Jolla, CA 92093-0221
(USA)

Experts

Prof. Meinrat O. ANDREAE, Director
Biogeochemistry Department
Max Planck Institute for Chemistry
P.O. Box 3060
D-55020 Mainz
(Federal Republic of Germany)

Dr. David E. BLOOM, Chair
Department of Population and International Health
School of Public Health
Harvard University
665 Huntington Avenue
Boston, MA 02115
(USA)

Dr. Ann G. CARMICHAEL
Indiana University
Department of History and of HPSc
742 Ballantine Hall
1020 East Kirkwood Avenue
Bloomington, IN 47405-7103
(USA)

Prof. Ulisses E.C. CONFALONIERI
Fundação Oswaldo Cruz – Escola Nacional de Saúde Pública (ENSP-Fiorcruz)
Departamento de Ciências Biológicas
Av. Brasil 4036/703
Manguinhos
B-21040-361 – Rio de Janeiro, RJ
(Brazil)

Prof. Andrew P. DOBSON
Princeton University
Department of Ecology and Evolutionary Biology
211 Eno Hall
Princeton, N.J. 08544-1003
(USA)

Dr. David L. HEYMANN, Representative of the Director-General
Polio Eradication Campaign
World Health Organization
Avenue Appia 20
CH-1211 Geneva 27
(Switzerland)

Prof. Carlo C. JÄGER
Potsdam Institute for Climate Impact Research (PIK)
Department of Global Change and Social Systems
Telegrafenberg A31
D-14412 Potsdam
(Federal Republic of Germany)

Prof. Eric F. LAMBIN
Université Catholique de Louvain
Faculté des Sciences
Département de Géologie et de Géographie
Place Louis Pasteur, 3
B-1348 Louvain-La-Neuve
(Belgium)

Dr. Wolfgang LUTZ
International Institute for Applied Systems Analysis (IIASA)
World Population Program
Schlossplatz 1
A-2361 Laxenburg
(Austria)

Prof. Anthony J. MCMICHAEL, Director
National Centre for Epidemiology & Population Health
Australian National University
Canberra, ACT 0200
(Australia)

Prof. Pim Martens, Director
International Centre for Integrative Studies (ICIS)
University of Maastricht
P.O. Box 616
NL-6200 MD Maastricht
(The Netherlands)

Dr. Michel H. Meybeck
Sisyphe – Université Pierre et Marie Curie (Paris VI)
4, place Jussieu – Case 123, Tour 56, Couloir 56-46, 4ème étage
F-75252 Paris Cedex 05
(France)

Dr. William F. Ruddiman
University of Virginia
Department of Environmental Sciences
Clark Hall
291 McCormick Road – P.O. Box 400123
Charlottesville, VA 22904-4123
(USA)

Prof. Wolfgang Sachs
Wuppertal Institute for Climate, Environment, Energy
Döppersberg 19 – P.O. Box 10 04 80
D-42103 Wuppertal
(Federal Republic of Germany)

Dr. Mahendra M. Shah
Senior Scientist: Land Use Change
International Institute for Applied Systems Analysis (IIASA)
Room s145
Schlossplatz 1
A-2361 Laxenburg
(Austria)

Dr. Will STEFFEN, Executive Director
International Geosphere-Biosphere Programme (IGBP)
Royal Swedish Academy of Sciences
Box 50005
S-104 05 Stockholm
(Sweden)

Dr. Mary E. WILSON
1812 Kalorama Square, N.W.
Washington, DC 20008-4022
(USA)

Observer

Dr. Bettina MENNE
Medical Officer Global Change and Health
World Health Organization (WHO) – Regional Office for Europe
Via Francesco Crispi, 10
I-00187 Rome
(Italy)

SCIENTIFIC PAPERS

INFECTIOUS DISEASE AND HUMAN AGENCY: AN HISTORICAL OVERVIEW

ANN G. CARMICHAEL

We have, because human, an inalienable prerogative of responsibility which we cannot devolve, no, not as once was thought, even upon the stars. We can share it only with each other.

(Sir Charles Scott Sherrington, *Man on His Nature*)

Introduction

The speed with which infectious diseases can now spread, the absolute number of humans at risk, and the rates of emergence of new pathogens: such parameters of global epidemics invite historical comparisons. Incurable, acute infectious diseases can now occur almost simultaneously across the globe, creating the further possibility that individuals who know what to do and how to intervene will themselves be compromised. SARS, for example, had its highest morbidity and mortality rates among medical personnel and hospital workers (NAC, 2003). The planetary ecosystem, meanwhile, may face many extensive global threats all at once, requiring a kind of global governance and compassionate collaboration of peoples that has never before existed. Finally, a substantial portion of humans alive today live in deep poverty, without access to material, social or political resources (Davis, 2006; UN-Habitat 2003).

This paper specifically reviews past human efforts in the understanding and control of infectious diseases. Within the past 500 years we do find justly celebrated successes. The imposition of barrier technologies to impede the spread and perceived contagion of plague, the invention of preventive vaccination and eventually eradication of smallpox, and the triumph of sanitation-based public health in the control of pandemic cholera epidemics: such are familiar examples among biomedical and public health communities. The evolution of effective epidemic controls

and interventions, this paper will suggest, was linked to the perception of geographical and socioeconomic differences in mortality experience even during great epidemics. However the benefits of most interventions accrued to some segments of a population and to larger geographical regions. Ironically success in epidemic control created further differentials in vulnerability and risk among peoples and nations.

Retelling success stories does lend a helpful optimism about the capabilities of science allied to human political and economic will, for the global environmental problems before us require a belief that our concerted human actions can prevent catastrophe. Lessons of the 2003 SARS epidemic, for example, should include the speed with which a complex disease threat was recognized clinically and understood epidemiologically (NAS, 2004; NAC, 2003; Weiss and McLean, 2004). The full lethal potential of the epidemic was certainly blunted through a rapid infusion of research funding, a general international public awareness aided by the Internet, and broad interdisciplinary approaches to disease control, deftly implemented. Biomedical science and international surveillance are powerful tools, exponentially increased by recently developed knowledge and communications technologies.

Understanding the risks before us now requires a sophisticated environmental perspective that cause – and cure-driven medicine does not communicate well to consumers. The very successes of modern biomedicine facilitate both a false sense of security among those who expect to benefit, and a deep resentment among those who see technological progress as a threat. The illusion of safety among the currently privileged will further distance them from worries and fears that could mobilize resources to prevent novel pandemics. On the other hand, many people, when presented with a threat or crisis, will accept simplifying, familiar explanations from the past because the political and technological solutions adopted in a crisis threaten their livelihoods and social well-being. Unfortunately, many may repeat some reactions and responses from the past because they do not understand fully how the present global changes differ from the agrarian age of great epidemics. But they can easily grasp that the benefits of future breakthrough technologies will surely be inequitably distributed, disparities which reinforce the human experience of centuries.

I still find useful the dominant Western ideology of 'progress', because it is an antidote to inertia and denial before the current predicament. Thus I offer the argument that perception of differential mortality in the past led to successes in human health and longevity. In the remote his-

torical past, mass suffering and dying were typically unseen outside stricken regions. Before the last century most humans could not see far enough in space or time to imagine their responsibilities to other humans on a global scale, even if they thought of their actions in such grand terms. Disease catastrophes progressed across regions within a time frame perceptible and measurable by humans' immediate lived experience. Progress followed improving communication networks, which highlighted disparities in health and disease experience. Recognition of differences among persons and groups led to useful interventions, if humans accepted a responsibility to act on these observations.

Seeing further than our predecessors did, in both time and space, carries ethical obligations. Not only are the economic connections among humans much stronger today than ever before, knowledge about the sufferings of others is not easily avoided. Great and catastrophic events, especially those that involve costs to people who regard themselves as non-responsible by-standers rather than as stakeholders, present disturbing ethical and economic dilemmas. However quickly we understood the biomedical parameters of HIV twenty years ago, and were then able to create new surveillance and control strategies, the longer-perspective of HIV/AIDS control is, unfortunately, already one that replicates historical patterns of inequitable distribution of resources and care, when assessed from the perspective of both incidence and prevalence of the infection (Farmer, 1999).

Residual apocalyptical thinking poses the greatest danger to the successful management of acute epidemic crises. The religion-driven acceptance that an end of days is foretold and possibly at hand has governed much human thinking about great epidemic infectious diseases over the past two millennia. Apocalypticism passively, and sometimes actively, dismantles our acceptance of an obligation to act collectively and collaboratively for the betterment and survival of all, because traditional religions tend to separate the saved from those whom they view as damned or evil. Fear is often used to rivet attention, attract resources, and shape political and economic actions, but can exacerbate converging natural catastrophes, which inevitably occur. Even though apocalyptical fears – including modern, secular environmental apocalypticism (Killingsworth and Palmer, 1996) – are often used to effect moral or behavioral reforms, historical evidence shows that living in the shadow of such fears reinforces violence and the dismissal of others' entitlements to either salvation or survival. Fear-driven assessments of current global change risk re-deployment of this dangerous ideological ordnance.

Three Historical Stages Linking Human History and Global Change

What impact has global change had upon human health in the past? With each significant change in humans' active manipulation of energy sources, initial health costs ensued, followed by aggregate improvements and significant human population increases (Szreter, 1997). These stages in human development may have happily coincided with geophysical stability, for the global climate of the last 10,000 years has for the most part supported human-driven changes-first changes to landscapes, then to the atmosphere, and finally to the oceans.

Similarly we might delineate three modes of anthropogenic production involving significant changes to the earth: hunting and gathering, agricultural, and industrial (Caldwell, 2004; Sieferle, 2001; Smil, 1994). Most recorded or archeologically accessible human experience with great epidemics lies within the agricultural era. Within this long phase of human history, the collective fate of all flora and fauna changed dramatically five hundred years ago. Regional economies, terrestrial ecosystems, and knowing scientific manipulation of the earth's resources altered the planet's landscape in this 'first global age', and altered human health and disease patterns forever. Effectively there were two worlds that humans manipulated before around 1500 CE. After that point one rapidly changing world emerged, and new human observations and ideas about infectious diseases accompanied post-Columbian global change. The underlying frameworks for analyzing and manipulating disease experience were also laid down before industrialization, justifying further attention to the interval that I call our 'first global age'.

Nevertheless the industrial transformation, beginning in Britain in the 1830s, was the single greatest change for both humanity and the planet, and most of the successful efforts at disease control lie within this ongoing era (Sieferle, 2001; Mokyr, 2002). Industrialization carried heavy initial penalties to the laboring population involved in production, much as had the transformation from hunting and gathering to agriculture (Szreter, 1997). Highly successful, combined English and French approaches to public health and sanitation overlap the industrial revolution temporally, but likely stem from the slower 'knowledge revolution' that preceded industrialization (Mokyr, 2002). These coincident developments fueled notions of 'progress', which led industrializing nations to accept their disproportionate wealth as inevitable and ordained.

The potential for global environmental catastrophe was created within advanced industrial modernity of the last fifty years, because human

activities now can directly affect the oceans and atmosphere. Yet over the last fifty years we also have created the means to analyze and anticipate adverse consequences of our actions and choices. Within the last fifty years we have devised curative and preventive strategies, not all of which involve much expense. We have the capability of maintaining and improving health for all humans. As a consequence, life expectancies in many nations have increased dramatically (Riley, 2001). Global population has doubled over this miraculous and precarious half century (Whitmore *et al.*, 1990; Maddison, 2003). Since all the great acute epidemic catastrophes occurred before 1950, the points of comparison must focus on the general predicaments, explanations, and choices made in the past.

1. Global Change before the First Global Age

1.1. *Before agriculture: human planetary dominance*

For at least 40,000 years humans have deforested the earth, altered the course of streams, and fashioned tools and weapons to expand food-getting. Even before consolidation into early agrarian societies, humans effected global change with the megafauna extinctions of the late Pleistocene and their many uses of fire (Mithen, 2004; Williams, 2003; Chew 2001; Dean, 1995). Before the transition to permanent agricultural settlement, Paleolithic humans increased and multiplied globally by 'extensification', that is, increasing through dispersal the numbers, range and varieties of similarly sized human population groups (Christian, 2004). While the causes of the transition to agricultural are unclear, Sieferle (2001) argues that humans could always resort to plants and smaller game, when large game became scarce. As obligatory omnivores, we were thus ecologically more efficient than both carnivore and herbivore competitors, leading to multi-regional human dominance. The survival of a wide variety of tools and objects of ritual importance bears witness to sophisticated cultural strategies and human adaptation to many harsh environments, evidence of the further importance of knowledge transmitted over time to our species' global dominance.

Human global population, around six million at the end of the last ice age – around 12,000 years ago – doubled every 5,600 years over the period from 40,000 BCE to the appearance of agrarian civilizations, around 8000 years ago (Livi-Bacci, 2001; Mithen, 2004). During this long interval

chronic infectious disease and disability, both traumatic and infectious, likely characterized most of the burden of illness in migratory groups (Cohen, 1989; Sieferle, 2001). There is simply no way to know how they faced, explained or managed temporally discrete disease and subsistence crises. Hunting and gathering human groups did not cook or store food, and they frequently removed to new locations. Relocations likely were prompted by the depletion of easily obtained food or game, typically through habitat destruction. Evidence from the relatively late colonization of Pacific oceanic islands often serves as a model of this much-earlier process (McNeill, 1994; Nunn, 1990).

1.2. *The Neolithic Revolution and Infectious Diseases*

The Neolithic revolution led to both agricultural and pastoral economies (Hughes, 1994; McNeill, 2004; Landers, 2003). After the agricultural transition humans multiplied in geographically concentrated localities, increasing the burden of gastrointestinal and respiratory pathogens. Evidence of chronic tuberculosis, smallpox, schistosomiasis infection can be found in ancient skeletal or mummy remains; descriptive evidence of periodic acute epidemics can be found in early written compendia (Grmek, 1989). Demographic costs in the move to settled agricultural societies were initially high, particularly in early childhood (Cohen, 1989). Rats, mosquitoes and other commensal species enhanced pathogen loads within settlements by adding vector capacity. A number of historically important zoonoses – measles and smallpox, for example – emerged with pastoralism and with the domestication of various animal species (Fiennes, 1978; McMichael, 2001). Such zoonoses mostly appeared in Eurasia and Africa, for the Americas and Micronesia had few animals to domesticate; Oceania had almost none. Whether farmers or herders, Neolithic humans transformed environments as much in their deliberate and inadvertent uses of animal species as in farming-driven alterations to landscapes. Northward migration into colder, wetter climates made the muscle power of livestock necessary to cultivation and facilitated a 'secondary products revolution' in the invention of multiple uses of large animals (Christian, 2004).

1.2.1. *Human Population Growth and Social Development: Unequal Burdens*

Global human population grew from 6 million to 50 million between 8000 BCE and 3000 BCE. The experience of recurrent acute infectious dis-

eases, mostly evolved from animal pathogens, ubiquitously appeared in a context of differential command of agricultural surpluses. Novel disease burdens accompanied the later phases of the Neolithic revolution, but they were not wholly natural occurrences. Infectious disease was understood well enough to be deliberately manipulated by developing social and economic systems. Dynastic states relied on hierarchical distribution of power and resources, which raised the average standard of living in the population at the same time as it reinforced social stratification (Hughes, 1994; Christian, 2004, chapter 9). Within great civilizations the burden of infectious diseases fell heaviest on the poorest, particularly in urban settings. Heights of skeletal remains show greater final adult height among the wealthy than among those who lived in overcrowded or swampy locations, or were subjected to recurrent food shortages (Floud *et al.*, 1990), most likely because the privileged always ate first (Sieferle, 2001). Comparison of different dynastic states shows that greater social, political and economic organization minimized the overall cost in lives and resources of epidemics, famines, and natural catastrophes (Bawden and Reycraft, 2000).

Successful dynastic states accumulated material and knowledge resources, and built long-lasting regimes around increasingly sophisticated irrigation systems. Insuring reliable food production, irrigation permitted easy dissemination of waterborne pathogens. The history of early Japan provides a particularly good example of this overall pattern of change. Wet-rice cultivation first supplemented, then replaced hunting, fishing and gathering between 200 BCE and 300 CE in the islands that had greatest contact with the Asian mainland. Increasing hydrological sophistication decreased famines, but was accompanied by domestication of animals and consolidation of power and land in warrior clans. Between 600 and 800 CE, one clan successfully overpowered rival warlords, resisted naval assaults, and then created a structure to monitor its wealth, borrowing comprehensive systems of population registration and tax collection already invented by the Chinese. The written records of epidemics, in particular of widespread, costly smallpox epidemics first appear clearly in these court records (Farris, 1985). All aspects of agrarian state power were mobilized during epidemics – food relief, a central bureau of medicine dispatching medicines and doctors to the provinces, tax relief, rice loans, and religious ceremonies to appease angry deities were all used to combat the epidemic of the 730s CE, the first recorded smallpox epidemic in Japan.

Earliest strong evidence of the adaptive capacities of human societies comes from imperial China. A Chinese political ideology dating from the

Han Dynasty [c. 200 BCE] held that that rulers were legitimated by a 'Mandate of Heaven', and climatic or other environmental catastrophe provided the evidence that heaven had withdrawn its support for a ruling dynasty. Anxiety about the weather was thus longstanding, and Chinese phenological records exist for more than a millennium longer than do European climate records (Marks, 1998; Elvin, 1998; Kuriyama, 2000). The permission that heaven provided to a dynasty reinforced strongly hierarchical social organization and led to both unprecedented expansion of human population and great hydraulic engineering projects that transformed East Asian landscapes forever (Elvin, 2004, chapter 6). When the 'mandate' was withdrawn, through heaven-sent invaders or environmental catastrophe, hundreds of thousands died, vital irrigation structures went untended, and epidemics served mostly to accelerate demographic collapse (Elvin, 1993).

Using surviving population registers, tax collections, and conscription records, Deng (2004) has recently challenged earlier conservative estimates of Chinese population growth intervals, showing that patterns of unparalleled population growth in some periods before the seventeenth century are even more characteristic of Chinese society than are massive collapses secondary to disasters. Moreover, most of the known disasters were environmental or military reversals, rather than caused by exogenous novel disease organisms. This Chinese path to collapse is important to understand, because at least for the last four millennia, Chinese population has accounted for one quarter of humanity (Lee and Wang, 1999). The Chinese path is also a useful contrast to explanatory frameworks favored by Western historians of global catastrophes, for collapse was linked to converging (rather than linear, sequential) ecological disasters, followed by out-migration and political upheaval. Disaster responses in Chinese history illustrate the enormous buffering capacity of human social organization, as well as the rapid collapse after some tipping point within.

1.3. *Explanations and Patterns of Acute Epidemic Disease Experience in Early Agrarian Civilizations*

1.3.1. *Where are we Now? Apocalyptical Versus Scientific Discussions*

Most historians, social scientists, and scientists explain reversals of large, multi-century and/or multi-regional human demographic expansion by citing epidemics, climate change, the exhaustion of accessible natural resources, or by some combination of these factors. Many great

religious traditions, however, include a view of time bounded by a predetermined beginning and end of days. The power of such ancient religious ideas cannot be easily excised from current scientific discussion. Not only did apocalyptical perspectives inspire much of the written evidence that investigators today use to discern patterns and explain collapse in earlier times. Also adherents to fundamentalism, steeped in ideologies created long ago and under very different historical conditions, have increased considerably within rich and powerful, as well as poor and struggling, political regimes (McGinn *et al.*, 2003).

Within the last two decades more extensive interdisciplinary study and collaboration helped to refine scientific understandings of distant times (Bawden and Reycraft, 2000). Scientifically we are beginning to know better what actually did happen in the more remote past. Yet both well-meaning environmentalist reformers and anti-technological, anti-scientific religious authorities have appropriated apocalypticism over the last half-century (Killingsworth and Palmer, 1996). Some hope to recapitulate ideas expressed in religious tradition; others hope to reverse or brake adverse global change. At issue in both cases is a view of time and human development that opposes the simpler versions of 'progress' in Western tradition.

1.3.2. *Historical Explanations of Great Epidemics and Environmental Disasters*

Great epidemics and uncontrollable climatic reversals are attractive explanatory events to account for sudden collapses of dominant societies. The particularly influential publications of William H. McNeill (*Plagues and Peoples*, 1976) and Alfred Crosby (*The Columbian Exchange*, 1972; and *Ecological Imperialism*, 1986), gave prominent place to plagues. More recent work has emphasized the extent to which ecological degradation eclipsed or stressed elites' exercise of power. Demographically successful cultures exhausted both forests and productive farmland, expanding the geographical compass of ecological degradation, making it difficult to separate the causal mechanisms and the timing of catastrophes (Chew, 2001; McNeill, 2004; Landers, 2003).

Urbanization, common in successful agrarian civilizations of the last 5000 years, effected profound environmental change over their hinterlands. Epidemics spread most briskly in cities and could economically destabilize distant regions dependent on these urban markets. This model

resonates with new research on the way Native American agrarian empires collapsed. Haug *et al*. (2003) find climate change, specifically desertification, related to the collapse of Mayan civilizations in the 800-1000 CE, exacerbating social and environmental stress.

Chronic warfare was the motor behind social stratification, as well as the most common accelerant of both disease and ecological degradation. The retention of elite power relied on a parasitic approach to both food producers and forests. Forests in Eurasia were essential to the manufacture of metal weapons; the best agricultural land and the rights to its produce were secured through military and economic dominance. Urbanization required the maintenance of such military power, which in turn determined access to and acquisition of natural resources, even in regions, such as the Americas, where neither the wheel nor metallurgy had been invented. In some rare instances, warfare led humans of the agrarian period to go beyond control and exhaustion of organic resources (plants and animals), leading to the sporadic invention and use of chemical energy (Landers, 2003).

1.3.3. *Modern Malthusian Models of Catastrophes in the Agrarian Age*

Some modern economic historians and historical demographers place a high value on early epidemic accounts because they resonate with a widely-held Malthusian model of the more remote past. Ever since Thomas Malthus's first *Essay on Population* in 1798, many western scholars have seen depopulation crises as indicating perennially unwise choices of humans, principally their failure to control reproduction in the face of diminishing resources. Such a model gives a prominent explanatory role to great epidemics and subsistence crises. Watkins and Menken (1988), however, argue that demographically famines do not follow Malthusian patterns; Morineau (1998) debates the extent to which historical data ever supported Malthus's generalizations. Historians of Asia note the extent to which Malthus, and later Marx, were instead drawing a chauvinistic contrast with 'Oriental despotism'; Europeans in Malthus's day were still stunned by the tremendous wealth of non-European despots (Chaudhuri, 1986; Lavely and Bin Wong, 1998). In fact, the recent Eurasia Project shows, in full contradiction to Malthus's observations, that European harvest failures had greater impact on mortality than did harvest failures in China and Japan (Bengtsson *et al*., 2004). Finally, the Malthusian model well explains mortality in great epidemics. Wrigley and

Schofield, in their path-breaking *The Population History of England*, 1541-1871 (1981, pp. 466-73) conclude that increasing population size is only weakly related to increases in mortality. Recent study of the horrific and extensive Asian famines of the 1870s and 1890s, in both cases associated with ENSO-related delay or absence of monsoon rains, has emphasized the extent to which Western acceptance of famines as a 'natural' consequence of unbridled population growth led those who could help to the conclusion that they were justified to do nothing (Morrison, 2000; Davis, 2001).

Malthusian models are not even universally applicable to the more remote past. Imbedded in western culture, they poorly explain population dynamics outside Europe, as Deng (2004) illustrates with data about Chinese population growth and decline. Even in western context, imbalance between resources and population size is insufficient to explain who dies in great subsistence crises. The earliest great famine for which we have fairly good social/historical evidence, the great European famine of 1315-1322 CE, does not follow a Malthusian pattern well. Mortality reached as high as ten percent, but affected the poor and powerless rather than the elite, and proceeded to cannibalism and other atrocities only in localities where the warrior elite persisted with military campaigns. The famine remained in popular memory three hundred years later (Jordan, 1996). Amartya Sen's model of modern famine survival through socially determined food entitlements thus seems to apply to this great medieval, agrarian famine (Sen, 1981).

1.3.4. *Exogenous Climate and Other Events Explaining Large-Scale Depopulation*

A few palynologists, dendrochronologists and historical climatologists have tried to discern intervals of global change that might explain large-scale regional depopulation events in frameworks linked to geophysical change. Impacts of meteors, changes in monsoon rains or northern oceanic storm patterns, and volcanic eruptions have all been postulated as potential disruptors of life systems across cultural and geophysical divides (Baillie, 1999; Atwell, 2001). Historical environmentalists have also reconstructed past El Niño patterns looking for evidence of the autonomous climate change in human history (Nunn and Britton, 2001; Grove, 1998). The disappearance of record-keeping and of evidence of economic and cultural production has also been ascribed to great epidemics. Only fairly recently has interdisciplinary research begun to sepa-

rate and examine the extent to which plagues and pestilences reported as great and universal indeed had demographic, regional significance. The contribution of Ruddiman and Carmichael (2006) in this volume addresses this general issue in greater detail.

1.4. *Explanations of Epidemics in the Agrarian Age*

1.4.1. *Epidemics and Environmentalism*

Secular ideas about the meaning and significance of epidemics in the long agricultural era of human history were linked to social and religious systems that reinforced social hierarchies and controlled their uses of available natural resources. The tight interconnectedness of humans to the natural environment remained central to both medical and religious ideologies of the premodern past (Tuan, 1979). Traditional medical systems explained death from commonly-observed misfortunes with reference to individualized circumstances. These medical systems offered some useful generalizations about disparities in risk of ill health according to season, age, geographic particulars of one's residence, or variations in one's nutrition (Hughes 1994; Grmek, 1989; Glacken, 1967; Carmichael and Moran, 2002). Interventions in ill health and general health maintenance relied on strategies that individuals could take to manipulate their personal environments and choices. Great epidemics instead prompted collective action, including the mobilization of religious authorities. The serendipity of fortune, the capricious will of a deity, the result of sin or emotions, the violation of a taboo, or the individual's failure to observe accepted dietary or hygienic norms accounted for deaths that did not follow expected patterns.

1.4.2. *Apocalypticism*

Occurring rarely, catastrophes and natural disasters were also viewed as manifestations of divine displeasure, and were assigned religious meaning. While such explanations for an epidemic or environmental catastrophe could easily foster larger, apocalyptical predictions, blaming deities also exacerbated the feelings of helplessness, in both explaining and coping with an unexpected set of events (Rohr, 2003). Moreover reliable communications beyond a regional level were not possible, and therefore an epidemic could seem universal and world-annihilating. Viewing the world as coming

to end mobilized efforts at religious conversion. Apocalypticism can, however, disable concerted human efforts in the face of calamity, or encourage believers to dissociate themselves from those assumed to be damned. Thus to the extent that such ideas have currency today, believers may be disinclined to take ameliorative actions when confronted with multiple serious epidemics or environmental challenges.

Ancient apocalypticism, particularly strong among Mediterranean cultures and deeply imbedded within early Christianity, imagined collapse as a battle lost, the divine actively in combat with mortals (Clifford, 2003). Religious devotion then and now powerfully motivates compassion, altruism, and philanthropy, as well as supplies stabilizing resolve to face fear and challenges. But many of the great global religions in the past encouraged resignation to catastrophe, or envisioned scenarios of an apocalyptical finale that separated the saved from the damned. Ancient apocalypticism also suffused some texts that have been used to support arguments about a late-Roman era of widespread plagues and pestilences. Stathakopoulous (2004) and Schamiloglu (2004) independently reviewed evidence related to the famous plague of Justinian in 541-2 CE, showing how this epidemic and its recurrences had geographically and demographically extensive effects. In the centuries before this pandemic, however, the evidence for widespread depopulation is not as clear, and apocalyptical texts should not be understood to reflect demographic collapse. Recurrent bubonic plague traveled along the Silk Road during the 600s (Twitchett, 1979; Deng, 2003). Schamiloglu (1993) claims that the disease was novel to steppe peoples.

2. European Responses to Plague: Epidemic Controls and Anti-Apocalypticism

2.1 *The Black Death: a Multi-Regional Epidemic with Global Implications*

In many accounts of global change, the great plague pandemic of 1347-1350 appears as one of the greatest demographic catastrophes of human history. Its initial appearance has been viewed both as an instance of Malthusian check on a population that had outstripped its agricultural capacity, and as an outcome of global climate changes. The ‘Black Death’ of the mid-fourteenth century was also geographically limited, and may not have extended to China. But it did sever frequent overland connections to East Asia, by causing extensive depopulation and political

chaos among steppe-region successors to the Mongol Golden Horde (Schamiloglu, 1993). Dramatic depopulation simultaneously occurred in northern China during the Yuan, or Mongol, dynasty (1264-1368), though disease may not have been the primary cause (Deng, 2003). The 'Black Death' surely ravaged the Middle East, northern Africa, and all of western and central Europe (Benedictow, 2004). Nevertheless the brutal pandemic did not afflict peoples in the Western Hemisphere, in sub-Saharan Africa, or in South Asia (Chaudhuri, 1985), although Islamic cities and regions of the eastern Mediterranean and North Africa may have been even harder hit than western Europe (Dols, 1977).

2.2. *New Responses to Great Mortality*

Despite its less than universal geographical compass, the Black Death pandemic remains an historical event with global importance. The effect of plague among Eurasian steppe peoples was to re-separate Chinese agrarian civilization from European and Islamic agrarian civilizations (Schamiloglu, 2004; Dols, 1977; Chaudhuri, 1985, 1990). But above all, from the time of the Black Death in Europe, intellectual inquiry and practical responses to epidemics confronted the issues of local, differential mortality in epidemics alongside explanations for universal pestilence. Europeans living after this catastrophe invented novel ways of dealing with epidemics, developing technologies and ideologies that reshaped their perceptions of disease. Before this watershed, European medicine was but one of many traditional medical systems that depicted health as balance between humans and their environment. Dealing with recurrent plagues – which were always regional epidemics and could cause staggering loss of life in a few months' time – prompted new practices in European medicine.

The appearance of secular explanations to account for large-scale epidemic disease detached both traditional environmentalism and religious tradition from communal responses to epidemics, and led to aggressive local control policies. In Renaissance Italy, notions of a capricious or punishing deity were gradually replaced with explanations built around observable variations in risk and survival. The 'whole world' was not at risk in plagues. After the pandemic, Europeans increasingly saw and even created patterns of differential mortality, inventing ways of dealing with recurrent pestilence assured and reinforced differential risk of disease and death within communities (Carmichael, 1986). The deliberate manipulation of infectious disease experience begins with new practices in epidemic control.

However brutal the Black Death was, 60 to 65 percent of the population of Europe survived, also disconfirming apocalyptical expectations. Life went on. By the mid 1350s plague reports were scattered and survivors knew that the world had not ended. The mortality rates, ubiquity, symptoms, and suddenness of the Black Death distinguished it from previous epidemics and mortality crises. In these respects plague could thus be understood to reflect God's will and judgment (a punishment for human sins), but also an event explicable within the natural, created cosmos (Smoller, 2000; Rohr, 2003). Europeans also adopted the conviction that human actions to combat epidemics were both necessary and consistent with God's plan for humanity.

2.2.1. *Local Barrier Technologies*

The plague recurred in western Europe for over 350 years, and led to the invention of surveillance bureaucracies, which in turn supported the most enduring response to epidemics: flight. Advanced warning systems were predicated on the assumptions that plague could be predicted and that plague could be avoided. They obviously believed that not every person and not every place presented a risk of plague; thus information gathering was vitally important. Increasingly sophisticated study of the natural circumstances associated with plague appearances led Europeans to a complex array of proactive, protective, and self-defensive strategies. Together these anti-plague practices are the foundation of modern control of feared infectious diseases (Biraben, 1976), despite the fact that flight is no longer the principal security for the privileged.

Barrier technologies were designed to create boundaries between the sick and the well. They were not developed in a particular fixed sequence, nor under historically similar circumstances or time frames. Instead, these familiar public health practices were created and adopted piecemeal. Quarantine practices, public health boards, surveillance of travelers from infected areas, disinfection of goods and houses that the ill touched, isolation of the ill, protective garments for service personnel, and publicized reports of causes of death and illness are all social technologies invented in Europe to control the spread and the tremendous costs of recurrent plague. Bourdelais (2003) emphasizes that governing elites created most of these barrier technologies, and that little innovation accrued through medical advice or theory.

As political actions, plague controls addressed the personal, social and economic objectives of people in power. Carmichael (1986; 1998) argued

as well that plague controls served to create strong differentials in plague risk within cities, such that plague was increasingly seen as a disease of the poor. By the sixteenth and seventeenth centuries, elite consensus was that plague was not a universal leveler. Even when the rich could not or did not flee to safer country retreats, their chances of dying were not as great as their poorer contemporaries (Champion, 1993). Aggressive uses of fire, perfumes, street cleaning, sea water, gunpowder and tobacco became popular in the seventeenth century (Biraben, 1976).

2.2.2. *Regional Barrier Technologies: Quarantine and* Cordons Sanitaires

Quarantine, invented in 1377, in Ragusa (now Dubrovnik, Croatia), is perhaps the most interesting of these anti-plague practices, and the one industrializing nations most needed to dismantle. Initially in 1377 well individuals and seemingly useable goods were passively sequestered, in order to determine whether plague lurked within (Grmek, 1980). Imbedded within the quarantine were thus modern ideas of healthy carriers and incubation times. But quarantine practices were not incorporated into a system of plague controls until the sixteenth century and later, when principal Mediterranean metropoles began to use passive quarantine isolation in combination with port lazarettos to confine the ill (Ciano, 1976). Concerted regional and national political control of plague spread, such as the imposition of regional *cordons sanitaires*, seems to have been the single most successful human intervention leading to the disappearance of plague from Europe. Brockliss and Jones (1997) illustrate that the systematic use of quarantine within early modern France effectively restricted appearances of plague to the port cities.

By the 1700s Europeans essentially concluded that plague was neither an act of God nor a naturally European problem, and so they built ever more elaborate regional security boundaries. Panzac (1986) and Rothenberg (1973) show the intensity of quarantine and lazaret systems along the Austrian Empire's boundary with the Ottoman Empire. Ottoman lands, both Egypt and Turkey, effectively became the pest houses for all of Europe. Baldwin (1999) examines quarantine practices and the *cordon sanitaire* in the wake of the first European cholera epidemic, 1831-1832, illustrating that western European nations benefited from observing the failed quarantine practices of Russians and eastern Europeans, for quarantine even rigorously enforced did not impede the spread of cholera. Echenberg (2002) has more recently described the retrieval of older quarantine practices in the plague pandemics of 1894-

1905, illustrating the extent to which re-used quarantine practices reinforced economic and racial differences in mortality around the globe.

Indeed, quarantine always imposed significant personal hardships, high economic costs, and political unrest. Opposition to quarantine from its inception stemmed from the absence of any medical rationale for the practice and the enormous costs involved, principally in housing, feeding, and policing those whose lives and livelihoods were temporarily suspended. Quarantine and other barrier technologies did contribute to the disappearance of plague from western Europe, and have a lasting presence in the control of epidemics to this day. International reaction to SARS in 2003 began to move this kind of barrier control away from national governments to international health authorities (Fidler, 2004).

2.2.3. *Anti-Apocalyptical Secularization*

The appearance of secular explanations to account for large-scale epidemic disease detached both traditional environmentalism and religious tradition from communal responses to epidemics, leading to a more aggressive pursuit of policies and interventions that focused on particular people or places. Francesco Petrarca, often called the 'father' of the European Renaissance or of European humanism, provides an interesting and original voice. He imagined a prosperous 'posterity', for whom accounts of the great epidemic would seem unreal. Petrarca also imagined that his own generation had obligations to these unborn generations, rather than they to their ancestors. Finally, Petrarca examined and then effectively rejected contemporary views of the world nearing the end foretold in religious texts. His perspective is elegantly written, and dismisses an apocalyptical view of the mortality:

> How can posterity believe that there was once a time without floods, without fire either from heaven or from earth, without wars, or other visible disaster, in which not only this part or that part of the world, but almost all of it remained without a dweller? When was anything similar either seen or heard?… Oh happy people of the future, who have not known these miseries and perchance will class our testimony with the fables. We have, indeed, deserved these [punishments] and even greater; but our forefathers also have deserved them, and may our posterity not also merit the same…
> (Petrarca, *Letter to his 'Socrates'*, Book VIII, no. 7 [tr. Bernardo, 1975]).

Petrarca thus rejected the bounded time of religious tradition in favor of a stance of hope and expectancy toward humans' future, temporal lives. His less articulate contemporaries parsed the universal with study of the particular, noting areas and peoples that had escaped the universal plague, rejecting notions of a capriciously acting, inexplicable deity in favor of humanly verifiable, terrestrial explanations of variations in survival. Thus for many historians the European Renaissance remains, in retrospect, an optimistic, watershed moment in human history. At the time it was, as Bowsma (1980) describes, an age of anxiety. Because anxiety is fundamentally related to time and the perception of an uncertain future, clocked time and a secular, scientific search for order and certainty about the knowable world were culturally new dimensions of recovery from near-catastrophe.

3. Global Change During the First Global Age

Christopher Columbus, often described as the discoverer of a 'New World', instead initiated the creation of one new world from two old worlds. Significant biological and cultural transformation of the planet dates from this expanded moment of European overseas trade and exploration. Western hemisphere flora first transformed great agrarian societies in Africa and Asia, through their gradual adoption of nutrient-rich staples (peanuts, potatoes and sweet potatoes, maize, and manioc) that grew well enough in poor soils and supplemented meager grain-based diets in subtropical zones (Ho, 1955; Crosby, 1986). East and South Asian populations rose between 1500 and 1800, as a direct result of the adoption of such foodstuffs outside the western hemisphere, while in Africa American crops prevented any net continental population losses from the accelerating Atlantic slave trade (Maddison, 2003). Interestingly as well, many of the addictive plants commonly used throughout the globe today – tobacco, coffee, tea, coca, chocolate – spread even more quickly than did new staple foods.

During the period from 1500 to 1850, human disease patterns and global environments changed rather dramatically and irreversibly. The change was both qualitatively and quantitatively distinctive in human history. Historical climatologists know this era as the Little Ice Age, reflecting global environmental changes, but not all regions were negatively impacted. Despite disruptions to longstanding Indian Ocean trade, the

period from 1500 to 1750 was one of great expansion (Chaudhuri, 1985; Richards, 2003).

Historically this first global age is further important because humans began to take systematic notice of differences in disease experience, and from such observations they were able to invent new strategies of disease control and health improvement. Europeans, because they traded and traveled globally, were systematic observers of health disparities, and in many localities all written observations about disease and epidemics date from this period in time.

3.1. *Depopulation in the Americas and Oceania*

Numerous infectious diseases, particularly viral pathogens adapted to humans in Eurasia and Africa over the previous three millennia, spread unimpeded among indigenous peoples in great western hemisphere civilizations. Often European settlers regarded the disappearance of native Americans as providential, that is, reflecting God's intention that they increase and multiply in these newly discovered lands (Jones, 2003; Chaplin, 2001).

Recent research shows that brutal 'virgin soil' epidemics accompanied or followed, rather than instigated, demographic collapse. In the Caribbean and central Mexico smallpox did not arrive until population had already begun a precipitate decline, (i.e., in 1518, when first contact was 1492) (Livi-Bacci, 2003). To some extent the economic wellbeing of indigenous peoples may have been weakened before Europeans were actually present to witness the great die-offs. Because long-distance trade routes were crucial to the Aztec and Incan civilizations, economic stagnation likely had increased hardships among many indigenous peoples long before they encountered the novel viral infectious diseases. Alteration of trade patterns and re-orientation of North American native production (especially fur trading) toward French and English markets similarly altered livelihoods along the St. Lawrence river and seaway, all before the viral disease epidemics accelerated collapse (Richards, 2003).

Cooler global temperatures coupled to increasing frequency of extreme weather events in the seventeenth century destabilized subsistence economies even without the introduction of unfamiliar pathogens (Fagan, 2000). In sum, we need not invoke the particular virulence of discrete pathogens, nor the accidental vulnerability of a population to novel infectious diseases. But demographic collapse in the Americas had global economic (Dean, 1995) and environmental/atmospheric (Ruddiman, 2003) consequences, whatever the pathways.

3.2. *European Perceptions of New Disease Environments*

The unfolding global environmental transformation shaped Europeans' perceptions of disease and disease environments. In many colonial zones natives were fast disappearing without any comment from European observers. Dean (1995, p. 65) described such silences as 'an endless chain of complicity that enabled the neo-Europeans to claim the inheritance of an empty land'. Jesuit missionary accounts in New England and in Brazil are some of the few to express sustained regret and concern about the steady decline of native American populations. Pacific island changes, which for the most part began in the later eighteenth century, illustrate the massive and aggressive role of western traders, extracting resources and spreading chronic infectious diseases (syphilis and tuberculosis, especially) even before the viral pathogens accelerated population decline (Bushnell, 1993; McNeill, 1994; Igler, 2004). Interest in the mechanisms of ecological and demographic collapse in the Pacific islands has increased over the last decades because archaeological and historical analysis has been particularly collaborative, and because the various archipelagos offer documentation of a variety of starting and accelerating conditions.

3.2.1. *Observed Disease Differentials: a Stimulus to Disease Control*

Disease globalization during the first global age is easier to discern from our vantage point than it was at the time. Much as trade capitalized on differences in prices, commodities, and markets, less demographically significant differences in mortality and morbidity were heightened for contemporary observers during the first global age. Persisting across the first three centuries of European overseas trade and colonization were the high rates of shipboard mortality among sailors, and among early colonists outside Europe. Disability and death from 'sea scurvy' compromised military and exploratory missions from the sixteenth through the eighteenth centuries. Some of the earliest clinical trials of different therapies for scurvy, made by English ships' surgeon James Lind in the 1750s, stemmed from debates about the causes of mortality peculiar to those long at sea (Carpenter, 1986). The higher social status of officers did not necessarily protect them from this blight, but officers did have access to more food on board, as well as began these voyages better nourished.

During these centuries, some physicians focused on the ways in which particular built environments and occupations produced distinctive disease patterns. Traditional medical advice relied upon the avoidance of

unhealthy environments and the consumption of food appropriate to the individual's humoral complexion, neither of which was an acceptable strategy for dealing with the increasingly observed diseases of miners, chimney sweeps, peat diggers, hatters, or cesspool cleaners – to name just a few of the European occupational groups that attracted novel medical observations at the time. In Great Britain dependence on coal for domestic uses was widespread by the late sixteenth century (Wrigley, 2000) and further contributed to the sense of human created, rather than naturally occurring, patterns of disease experience (TeBracke, 1975).

Native Americans' dramatic vulnerabilities to smallpox and measles led Europeans to early racial theories of human differences (Chaplin, 2001). When European colonists enslaved Africans to replace indigenous laborers, racialized theories of varying mortality and morbidity were further reinforced, because Africans were (wrongly) believed to be less vulnerable to 'intermittent fever' (malaria) and to yellow fever. Sugar cultivation in particular created new disease environments for all mosquito-borne diseases, which McNeill links to broader changes in early modern geopolitics (McNeill, 2005).

3.2.2. *Novel Research on the Medical Uses of Plants*

Commercially the most systematic observations of biological differences began in these centuries with the search for marketable botanical drugs among the remedies and plants found outside Europe. American guaiacum, hailed as a cure for the new 'French pox' was widely distributed in European markets before 1520, and facilitated the association of syphilis with the newly discovered lands (Arrizabalaga *et al.*, 1997). Peruvian cinchona bark, containing alkaloids from which quinine would be isolated in the 1820s and thus effective against malarial fevers, was transported to Rome and used in clinical experiments within the hospital of *Santo Spirito* (associated with the Vatican and the *Collegio Romano*) before 1630 (Freedberg, 2002; Rocca, 2003). These and other examples illustrate the useful observations of differences in disease experience as a stimulus to novel therapeutic approaches and commercial pharmaceutical success (Schiebinger, 2005).

3.3. *Realities of Disease Differentials in the First Global Age*

Hoffman *et al.* (2002) analyze the extent to which strong variations in standards of living occurred also on the European home front. While the

per capita incomes of middle and upper class Europeans rose steadily, buoyed by trade wealth and slave labor colonies, in general the prices of luxury goods fell relative to the prices of staples, fuel and rents. The rich got richer, the urban poor visibly poorer (or at least shorter) (Floud *et al.*, 1990; Steckel, 2004). Some anthropometric and economic data from the eighteenth and early nineteenth centuries suggest that rural poverty, particularly if individuals had access to milk products and could consume some reliable portion of the crops they produced, may not have been as pernicious to overall health as was the poverty of urbanization (Komlos and Baten, 2004).

3.4. *Smallpox: Inoculation to Vaccination*

The most important therapeutic innovation of the first global age was the European adoption of first inoculation (or variolation) and then invention of vaccination against smallpox. Inoculation reinforced notions that particular diseases might be caused by particular substances, rather than by factors inhering within individuals or their specific environments. By the 1760s, physicians in French and English medical societies undertook statistical investigations of the utility of smallpox inoculation to populations, calculating survival and risk in cost-benefit terms (Rusnock, 2002). In 1796, English physician Edward Jenner invented the procedure of vaccination, offering a way to prevent creation of smallpox epidemics in protecting populations. The real benefits of smallpox vaccination occurred in the nineteenth century. Bantha and Dyson (1999) analyze the gradual and sequential imposition of variolation and vaccination in British India during the nineteenth century, and are able to show how smallpox alone caused between 16 and 24 percent of all deaths without such intervention. As with plague epidemics in Europe, recurrent smallpox epidemics made a significant contribution to demographic patterns independent of Malthusian pressure on resources.

3.5. *Eighteenth-Century Sanitary Campaigns*

Europeans were increasingly unwilling to accept starkly demonstrable mortality gradients. Rapid urbanization in Europe accompanied wealth from trade and overseas colonies, and cities were sites of observably higher morbidity and mortality than were rural. In the eighteenth-century elites launched great environmental campaigns to lower mortali-

ty and confront both epidemic and epizootic diseases (Riley, 1985, 1989; Wilkinson, 1992). The British also studied the population-level effects of smallpox inoculation, demonstrating by the 1760s that lives were saved and epidemic mortality reduced with inoculation (Rusnock, 2002). French and Italian engineers at this time were especially skilful in studies of water management, much of it in the service of urban street cleaning and refuse removal (Guillerme, 1988). Throughout Europe both cities and nation states systematized the collection of mortality statistics, causes of death, and records of disease occurrence.

The word statistics meant state management, and in an organic economy people and animals were essential components of a nation's wealth. European nation-states were not as successful in cooperative interregional epidemic or famine management at this time as the Qing Chinese state (Marks, 1998). But by the late eighteenth century limited initial successes in the reduction of death and illness rates in Europe convinced leaders that health was both a benefit and a measure of civilization. Actual reduction in mortality was elusive, however, because increased rural to urban migration with industrialization in the early nineteenth century outpaced the benefits of early, rudimentary public health interventions. Alter *et al.* (2004) show how industrial boom-towns of eastern Belgium were overwhelmed by recurrent epidemics, because public health housing and sanitation was not systematic until the later nineteenth century.

4. Global Change in the Industrial Age: Public Health and Medical Management of Epidemics

Industrial production created unprecedented wealth, quickly distinguishing human standard of living in industrializing nations from that in the great agrarian empires. Industrial production also began irreversible changes to global environments, first to the atmosphere, then to the oceans (Kates *et al.*, 1990; Crutzen and Steffen, 2003; Sieferle, 2001; Andreae *et al.*, 2004). This combined Western advantage was apparent to all before the end of the nineteenth century, and education and sanitary campaigns together reversed the longstanding demographic advantage that rural communities had over towns and cities by the early twentieth century; rural localities in the industrialized nations were the last to benefit from great sanitation projects (Bourdelais, 2003). But within this

western, industrial context, sub-populations that were strongly disadvantaged socially and economically, such as African Americans in US cities during the early twentieth century, were able to benefit disproportionately from public sanitary campaigns, because it was not technically feasible to parse distribution of this social good (Troesken, 2004). Even British laboring classes enjoyed significant rise in overall health status during World War I, due to government investments in public health infrastructure and in well infant care (Winter, 1977).

Wealth alone did not insure health improvements even in the nineteenth and early twentieth centuries: there is only a very weak association between rising real wages and falling mortality levels. Kunitz and Engerman (1992) argue that rising per capita income was not the underlying explanation for falling mortality rates. Szreter (1997) points out that initially every industrializing region saw significant decline in population health. Instead from 1830 to 1880, the fall in mortality and morbidity from great infectious diseases and epidemics was the result of a national sanitary reforms and internationally orchestrated interventions to prevent the spread of feared epidemic diseases (Bourdelais, 2003; Baldwin, 1999). Large, public investments in sanitation and health education paid great dividends in health improvement.

These aggregate improvements in life expectancy partly masked persisting health inequalities caused by imbalances in the distribution of resources within rich nations. Riley (2001) emphasizes that industrial wealth and Western approaches to epidemic control (barrier technologies, vaccination and other biomedical technologies, and public programs based on a filth theory of disease) were not necessary routes for many lesser-developed nations in creating parity in life expectancy. Health education could be effective, even in poor nations (Riley, 2001). But as they had done for centuries, great pandemics exposed underlying disparities in risk and survival, even while generating the old fear that everyone was now suddenly at risk. Various public and personal strategies to control expected, understood risks might improve the prospects of longevity even in poor nations. In wars, natural disasters, great epidemics, larger differentials in health and disease were not as easily overcome. As Davis (2001) illustrates, the profoundly different health experience of an emerging 'Third World' came sharply into view in colonial settings. Only after the Second World War did rich nations make any concerted, large-scale efforts to improve the health of these lesser-developed nations. Great urban slums presently recreated the pre-conditions for health disasters as

were seen in early nineteenth-century rapid urbanization (UN-Habitat, 2003; Davis, 2006).

None of the twentieth-century epidemics, including the great influenza pandemic, carried the mortality rates of recurrent plagues or even the costs from endemic, virulent smallpox. Wealthy societies, buoyed by the late nineteenth century optimism stemming from the germ theory of disease and successful public sanitary interventions, mobilized resources to further the analysis of cause and transmission. In this section of the paper, brief histories of pandemic cholera in the nineteenth century (at the beginning of industrialization) and the influenza pandemic of 1918 (concluding the first, industrialized global war) illustrate both the mobilization of collective resources to impede disease spread and the experience of differential mortality that both stimulated research and reflected longstanding disparities in health.

The histories of other great epidemics of the twentieth century, such as the histories of poliomyelitis vaccination and the WHO's malaria and smallpox eradication campaigns, might also be used to reflect biomedical gains that built on perceptions of geographical and seasonal differentials in mortality, to control as well as understand epidemics. But other contributions to this volume will explore in greater details the extent to which modern, successful control of epidemics has not relied on anti-microbial cures. Importantly, the successes and improvements of the human condition subsequent to industrial production and the health transition thus far have statistically overshadowed losses from the ancient four horsemen: war, famine, plague, and conquest. Nevertheless the twentieth century was one of sickening destruction and loss of life from warfare. Although wartime losses cause but brief reversals in the aggregate demographic patterns, wars provide atypical and extreme environments for the spread of infectious disease epidemics such as the 1918 influenza. Modern war and epidemics is also a topic beyond the scope of this overview.

4.1. *Cholera Pandemics*

We now know that *Vibrio cholerae* is globally distributed, particularly through infection of copepods, the food source for much marine life. But there is no written historical evidence of early, widespread cholera epidemics in the Indian Ocean trade zone (Arnold, 1991). Cholera suddenly appeared in temporally discrete pandemics during the early nineteenth century after the British consolidated their power in the Indian subcontinent.

Outbreaks initially called back into practice regional quarantine and other barrier defenses, which authorities in Britain and France resisted as scientifically-outdated and trade-unfriendly instruments of disease control (Baldwin, 1999). These plague-style epidemic controls also required international political cooperation, which proved more difficult to achieve than local and national improvements to general health (Bynum, 1993).

Cholera's clear origins in British India led the British to oppose many traditional barrier technologies during cholera years (Maglen, 2002). After the Suez Canal opened in 1869, and after further decreases in travel time that railroads and steamships allowed, Western powers increasingly saw the issues of cholera control as a matter of modifying quarantine and isolation practices to minimize mercantile costs (Hardy, 1993). Again they relegated application of such barriers to the Muslim nations that geographically separated South Asia from Western Europe (Howard-Jones, 1975; Fidler, 1999). Political solutions to global cholera pandemics, imposed by the rich, industrializing nations along lines similar to later plague controls once again created the sense of extra-territorial pest houses, rather than a global disease and environmental phenomenon.

Cholera also was viewed as an Asian invasion, rather than as either a global phenomenon or a consequence of British trade development in the Punjab. The evidence for global climate interconnectedness coalesced later in the nineteenth century, in part as recurrent cholera pandemics seemed to have some association with variations in monsoon rain patterns (Grove, 1998). ENSO [*El Niño Southern Oscillation*], first identified in the nineteenth century by English surgeons and climatologists in collaboration with Indian scholars, were noticed in part because epidemics of cholera could not be easily associated with either of the two presumed determinants of epidemics: trade and urban crowding. In the first half of the nineteenth century disruptions of seasonal monsoon rains could be linked to epidemics of cholera. Later in the century, simultaneous, virulent malaria epidemics in East Africa, South Asia, East Asia, and western South America coupled the rain and oceanic patterns of the Indian and Pacific Oceans to the phenomenon of monsoon-related epidemics (Davis, 2001).

4.2. *Public Health, Cholera, and Novel Sanitary Reforms within the Industrializing Nations*

Cholera's appearance in western Europe in 1830-32, evoked heated popular and scientific discussions of the origin and control of epidemic diseases. First-affected regions (Russia, Prussia, and Eastern Europe)

deployed draconian quarantine procedures to impede the spread of the new disease, so that by the time the disease reached Paris and London, public officials knew that costly, politically destabilizing quarantines were not necessary (Baldwin, 1999). Knowing the futility of quarantining cholera, French hygienists used the epidemic to test their hypothesis that poverty caused higher morbidity and mortality. French public health had a strong commitment to state responsibility for the health of citizens, so hygienists heroically collected data door-to-door, as well as participated in care of the stricken (Bourdelais, 2003; La Berge, 1992). Despite more than 18,000 deaths from cholera in the city and suburbs of Paris, in 1832, the epidemic reinforced their conviction that polluted environments and non-hygienic behaviors exacerbated mortality, and that without their efforts the epidemic costs would have been much higher.

Neither did the English want to see the return of plague-style controls. Coming to power just after the first cholera epidemic, arguably the single most important architect of modern public health control, Edwin Chadwick, embraced a more restrictive view of the responsibilities of government than did the French. Advocating government-sponsored sanitation projects, Chadwick resisted any notion that the state had to solve the perennial problem of poverty in order to achieve protection from cholera (Hamlin, 1998; Baldwin, 1999; Bourdelais, 2004). When the second pandemic of cholera returned to Europe in 1848-1854, new surveillance mechanisms were in place. Cause-of-death reporting had become systematic, with debate settling now upon how those causes were to be assessed and assigned. Chadwick and his allies were instrumental in delimiting a focused approach on the provision of clean water, and the construction of sewage systems.

The second global cholera pandemic, 1848 to 1854, appeared at the same time as new microscopes, fueling interest in specific cholera-causing substances. Political and social debate about the causes of cholera mortality multiplied published data and discussion, which facilitated the elegant early epidemiological studies of an otherwise obscure London physician, John Snow (Carmichael and Moran, 2002). Snow's implication of water sources infected with a specific cholera substance was not widely appreciated until later in the century, with the more rigorous bacteriological proof that Robert Koch provided in 1884 (Carter, 2003; Hardy, 1993). Meanwhile, sanitary reforms in Britain and France reinforced the political and economic cost-benefit judgments of public health authorities (Hamlin, 1998). In 1851 the French and English called into being the

first International Sanitary Commission to orchestrate Europe-wide responses to epidemic threats from outside Europe; and the first epidemiological societies were created. By the 1860s, before the third cholera pandemic of 1866, the English sanitarians created a radically modified 'English quarantine system', further elaborated into the 1880s (Hardy, 1993). Municipal sanitation created dramatic gains in survival for individuals in the wealthy, industrializing West (Baldwin, 1999).

Public health controls, by creating different expectations of health and risk, reinforced Westerners' sense of cultural superiority, when they witnessed economically struggling, unsanitary, famished backwardness of non-Western regions (Davis, 2001). Plague from 1894 to 1907 circled the globe, illustrating with the return of this dreaded, ancient scourge, that adherence to sanitation principles protected Europeans even when they were living in great squalid colonial metropolises (Echenberg, 2002).

Some important global consequences resulted in the way that cholera was handled in emerging international law. The first International Sanitary Congress of 1851 had placed oversight of cholera surveillance in the bailiwick of Ottoman Turkey and Egypt, regimes struggling to maintain economic prosperity in the rapidly changing world. Western nations became increasingly deaf to the political and religious conditions facilitating epidemics, and excluded non-Western nations from participation in international epidemic controls. Afkhami's (1999) analysis of late-nineteenth-century Iranian attempts to implement Western public sanitation and vaccination programs and his further work on cholera pandemics in Iran shows how important programs sensitive to local culture and political traditions were to the control of epidemics. Iranians not only did not see cholera as a disease threat from the potentially revolutionary, unwashed urban underclass, the importance that they placed on keeping open pilgrimage routes for devout Muslims created absolute resistance to Western plague protocols (Afkhami, 2003).

Between 1851 and World War I, International Sanitary Congresses met periodically to address feared, internationally spread epidemic diseases, focusing on bubonic plague, tropical yellow fever, and cholera (Howard-Jones, 1975; Bynum, 1993; Fidler, 1999). International legislation and protocols in the management of global epidemic threats were little revised even by the League of Nations Health Organization or the subsequent World Health Organization, and as a consequence stood as a somewhat archaic scaffolding underlying political and legal approaches, only recently overcome through WHO management of pandemic SARS (Fidler, 2004).

4.3. *Influenza 1918*

The severity and extent of the influenza pandemic in late 1918 came as a considerable surprise to medical and public health authorities throughout the world. Over 25 million persons died worldwide, and in rich, industrialized nations, the deaths among young adults were striking (Johnson and Mueller, 2002; Langford, 2002; NAS, 2004). Despite an exponential increase in 'population flux' – Cliff and Haggett (2004) create this measurement of population mobility to describe the increasing transmission of infectious diseases in the late nineteenth century – mortality rates in Europe had plummeted during the half century before the pandemic. Microbial causes and modes of transmission had been identified for most bacterial and protozoan diseases responsible for the adult deaths. Unanticipated measles epidemics among some of the World War I armies challenged medical and public health authorities, but otherwise epidemic infections among troops and on the home fronts were due to expected increases of sexually transmitted diseases, plus waterborne and vectorborne diseases (Smallman-Raynor and Cliff, 2004). In the US, wartime research on infectious diseases was meanwhile viewed as an opportunity for young investigators and clinicians, who might otherwise have traveled to exotic locations for such experience (Byerly, 2005; Barry, 2005).

Even today debate about the origins of the lethal strain of the influenza virus in 1918-19 is unresolved within both historical and scientific literature, providing a sophisticated and modern veneer to a story analogous to that of syphilis in the early sixteenth century. Shrouded in military secrecy, the focal appearances of virulent influenza cannot be discerned at the population level until late August, 1918. By then the name 'Spanish influenza', a French epithet for late spring cases of a severe, pneumonia-associated influenza reported uniquely in Spanish newspapers, had taken hold. Spanish journalists blamed the epidemic on French troops (Echeverri, 2003). The US civilian population was largely unaware of the Army camp deaths and high morbidity from pneumonia during the late winter and early spring of 1918, and military and public health authorities similarly suppressed reports of the widespread practice among camp commanders, transferring morbidly ill soldiers by train to other training camps (Barry, 2005; Byerly, 2005).

Most authorities have accepted three distinct waves of influenza infection, beginning with sudden deaths in a Kansas (USA) army camp in March 1918 (Patterson and Pyle, 1991; Johnson and Mueller, 2002). The origins of

a second, more virulent mutation occurred somewhere during the late summer, and then spread from Brest, France (where US troops disembarked for the Western Front); from Sierra Leone; and from a military camp outside Boston. Crosby's (1989) history linked the origins of the severe influenza strain to the extraordinary survival environments of trench warfare and chemical warfare, an explanation that has long carried moral weight. A direct connection to the war has seemed plausible because of the degree of filth and degradation to which front-line soldiers were subjected. As recently as a decade ago epidemiologist Paul Ewald (1994) argued that the extraordinary virulence of the 1918 influenza strain is inconceivable without the Western Front conditions. More recently virologist John Oxford (2001) returned to accounts of severe pneumonia outbreaks in army camps from 1915 to 1917, implicating one British camp in France as the area of the new virus's emergence, thus much earlier than the 'three-wave' model postulates (Oxford, 2001). The balance of evidence about the pandemic's origins, however, still points to recombinant virus emergence in farming communities of central North America, particularly in rural Kansas, and thus is quite similar to instances of emergent infectious diseases today. In particular, the route of a novel epizootic from rural settings to urban markets, and from there global distribution is quite similar to SARS ecology and epidemiology (NAS, 2004).

The war, nevertheless, exercised a considerable role in the spread of the infection globally. 'Population flux' exaggerated mortality among young adult males: all combatant nations had a strong gender imbalance in influenza mortality. In all regional studies of the pandemic the spread of infection (as noted in the timing of observed cases) flowed from military enclaves and venues to civilian populations (Phillips and Killingray, 2003). The fundamental role of human contacts in the transmission of the disease has been studied rather closely using evidence from the South Pacific island nations. New Zealand and many of the island chains were seeded through trade and returning troop ships. While larger ports took some precautions, the young native men who provided labor services for merchant ships were largely invisible to the rudimentary public health observation, and so were permitted to travel from port to port as well as return home once they became ill. Communities with few medical services, such as the Maori of New Zealand and the Society Islanders, were especially hard hit in early November 1918 (Rallu, 1990). Australia's public health authorities, however, took advantage of the time differential between information about the epidemic and the arrival of crowded troop

ships. Specifically targeted maritime quarantine, followed by household quarantines after soldiers disembarked, and military-style blockades along thoroughfares to the Australian outback successfully broke both the speed and the extent of the epidemic (Smallman-Raynor and Cliff, 2004; McCracken and Curson, 2003).

Secondly the war undermined civilian health indirectly, in the redirection of medical resources and personnel to the war effort (van Hartesveldt, 1992). Often the quantifiable aspects of a pandemic, including relief efforts orchestrated by those in power, give an incomplete picture of both risks and mitigating interventions. Previously impoverished areas remote from the conflict often suffered staggeringly high mortality and morbidity, unless the obstacles to mobility were great. Thus detailed, multidisciplinary local studies are now beginning to reveal these geographical and gendered differentials in influenza experience. For examples, access to female nursing care insured higher survival rates for many victims in the United States (Bristow, 2003); but women closest to the war front in France had much higher death rates than their male counterparts at any age (Zylberman, 2003). Data summarized by Johnson and Mueller (2002) illustrate further that overall mortality rates were much higher in lesser developed nations and regions.

The global reach of this epidemic had at least one positive effect, useful to note in the current century. Military and civilian scientists in the richest nations finally recognized the global interconnectedness of human populations confronting some epidemics: remote outbreaks became potential risks rather than research opportunities, and thus necessitated surveillance and investments in health.

Conclusion

This essay has offered a temporally very broad overview of human experience with globally significant acute epidemic infectious diseases. Past global climate changes, to the limited extent that they can be pinpointed with available evidence, appear not to determine the appearance and severity of identifiable global pandemics in the past. Before industrialization the link between climate and great pandemics instead seems the reverse: great depopulations may be linked to global cooling (Ruddiman, 2003; Ruddiman and Carmichael, 2006). But epidemics tend to play an important role in the ability of human societies to recover, once the bal-

ance between accessible food and natural resources on the one hand, and accessible human labor to exploit resources on the other, has been compromised. The realm of human ideas about catastrophic epidemics and natural catastrophes, however, has played an even more significant role in the gradual abandonment of religious ideas of a world-ending apocalypse, and the acceptance of secular, scientific ideas of human progress.

Global geophysical change is linked to the two radical changes in the mode of production over human history, agricultural/pastoral and industrial. Each transition came with initially high health costs, and in both instances those costs were redistributed and reinforced through warfare and differential access to resources. Here again mortality and morbidity experience in the past is not neatly aligned to changing ideas of epidemic causes. Since the Neolithic revolution, humans have largely assumed poverty and wealth occurred naturally. Thus the privileged paid greater attention to epidemics that could destabilize their power, particularly acute epidemic diseases that struck rich and poor alike. But experience first with recurrent plague in Europe, and then during the 'first global age' when European overseas ventures created multiple new and striking circumstances to observe disparities and differentials in mortality risks, propelled scientific observation, analysis, and experimentation with interventions. The rapid differences in wealth that industrialization subsequently created reinforced a sense of cultural, rather than mere economic superiority, among Westerners subscribing to ideas of secular 'progress'. Large-scale public health interventions within the West were undertaken because rich individuals learned that their own health in cholera epidemics could not be assured without such investment in sanitation. Aggressive surveillance, quarantines, vaccine development, and enhanced barrier technologies – the methods of the plague centuries – instead became the mode of epidemic response in western nations after the murderous influenza of 1918. After World War II, new international investments in disease eradication, economic development, education, and environmental manipulations returned to the cholera lessons of necessary public investment for private security.

The current challenges require cognizance of all these past approaches and fears. The most important historical and ethical issues to confront today relate to the ways that humans with power over resources make choices which determine the survival and well-being of those without power (Davis, 2004; Farmer, 1999). Over the last ten thousand years, human societies have displayed disappointing resistance to any truly

equitable distribution of resources and risks. But success in the control and mitigation of demographically punishing infectious diseases was linked to ideas born in the perception of inequities and differentials in health and disease experience.

Most catastrophic declines in global population in the past were regional, because access to resources was regional. Exhaustion of accessible resources and the chronic attempts to address shortfalls in production through warfare in the more remote past destabilized trade networks, exacerbating great epidemics. Economic stability is now determined globally. Thus re-establishing the means of supporting very large populations after any acute, disruptive epidemic, even if does not reach the levels of the great pandemics and epidemics of the agricultural era, can pose protracted economic and public health challenges to recovery. Also apocalypticism, re-emergent among fundamentalist religions as well as in some ecological thinking, not only encourages inaction, it often carries an unexpressed agenda for a future that does not acknowledge the rights of all persons to exist within it.

Human choices and human actions have mattered. Pre-twentieth-century European history has a special place in the larger story of infectious diseases and global change in many well-known aspects: industrial production, the scientific method of inquiry, surveillance, isolation and containment practices in the management of epidemics, and the environmental sanitation approach to ongoing disease control. All of these were successes in part because governments subscribed to the public health investments and invested resources beyond those that individuals could make. In democracies of Europe and North America, tax payers rationally concluded that their own, individual well-being was more precarious if such investment in sanitary infrastructure was not made.

Finally, this essay has emphasized some novel early modern underlying ideas preceding industrialization: the rejection of bounded time, leading to the acceptance of obligations to future generations rather than the past; the idea of contagion, leading to the creation of boundary technologies to impede the rapid transmission of infectious disease; and the stepwise awareness and analysis of differential patterns in mortality and morbidity, seen first when Europeans understood that plague was not universal and elaborated as they expanded during the first global age, 1500 to 1850 CE. These novel ideas link also in these centuries to political ideas of justice and liberty that could not be discussed here.

REFERENCES

I would especially like to express my gratitude to the organizers of this conference, and to my colleagues George Alter and James C. Riley, for their careful and patient comments and suggestions.

Afkhami, Amir Arsalan (1999), Defending the guarded domain: epidemics and the emergence of an international sanitary policy in Iran, *Comparative Studies of South Asia, Africa and the Middle East*, 19:1, 123-134.

Afkhami, Amir A. (2003), Disease and water supply: the case of cholera in 19th century Iran, *Yale F&S Bulletin*, 103: http://www.yale.edu/environment/publications/bulletin/103pdfs/103afkhami.pdf

Alter, George, Muriel Neven and Michael Oris (2004), Mortality and modernization in Sart and surroundings, 1812-1900, pp. 173-208 in T. Bengtsson *et al*. (eds.), *Life under Pressure: Mortality and Living Standards in Europe and Asia, 1700-1900*, Cambridge, MA: MIT Press.

Andreae, M.O., L. Talaue-McManus, and P.A. Matson, Anthropogenic modification of land, coastal and atmospheric systems as threats to the functioning of the Earth System, *Earth System Analysis for Sustainability*, H.-J. Schellnhuber, P.J. Crutzen, W.C. Clark, M. Claussen, & H. Held (eds.), pp. 245-264, The MIT Press, Cambridge, Mass., 2004.

Arnold, David (1991), The Indian Ocean as a disease zone, 1500-1950, *South Asia*, 14:2, 1-21.

Arrizabalaga, Jon, John Henderson, and Roger French (1997), *The Great Pox: the French Disease in Renaissance Europe*, New Haven: Yale University Press.

Atwell, William (2001), Volcanism and short-term climate change in East Asian and world history, c. 1200-1699, *Journal of World History*, 12:1, 29-98.

Baillie, Mike (1999), *Exodus to Arthur: Catastrophic Encounters with Comets*, London: Batsford.

Baldwin, Peter (1999), *Contagion and the State in Europe, 1830-1930*, Cambridge: Cambridge University Press.

Bantha, Jayant, and Tim Dyson (1999), Smallpox in nineteenth-century India, *Population and Development Review*, 24:4, 649-680.

Barry, John M. (2005), *The Great Influenza: The Epic Story of the Deadliest Plague in History*, New York: Penguin Books.

Bawden, Garth and Richard M. Reycraft (eds.) (2000), Environmental Disaster and the Archaeology of Human Response, Maxwell Museum of Archaeology, *Anthropological Papers No. 7*, Albuquerque: University of New Mexico Press.

Benedictow, Ole J. (2004), *The Black Death, 1346-1353: The complete history*, Woodbridge, Suffolk, UK: Boydell Press.

Bengtsson, T., C. Campbell, and J.Z. Lee (2004), *Life under Pressure: Mortality and Living Standards in Europe and Asia, 1700-1900*, Cambridge, MA: MIT Press.

Biraben, Jean-Noel (1976), La lutte contre la peste, Chapter 6, v. 2, pp. 55-181, *Les Hommes et la peste en France et dans les pays européens et méditerranéens*, Paris: Mouton,

Bourdelais, Patrice (2003), *Les Épidemies terrassées. Une histoire de pays riches*, Editions de La Martinière.

Bowsma, William (1980), Anxiety and the formation of early modern culture', pp. 215-246 in B.C. Malament (ed.), *After the Reformation: Essays in Honor of J.H. Hexter*, Philadelphia: University of Pennsylvania Press.

Bristow, Nancy K. (2003), You can't do anything for influenza: Doctors, nurses, and the power of gender during the influenza pandemic in the United States, pp. 58-69 in H. Phillips and D. Killingray (eds.), *The Spanish Influenza Pandemic of 1918-19: New Perspectives*.

Brockliss, Lawrence and Colin Jones (1997), *The Medical World of Early Modern France*, Oxford: Clarendon Press.

Bushnell, O.A. (1993), *The Gifts of Civilization: Germs and Genocide in Hawai'i*, Honolulu: University of Hawai'i Press.

Byerle, Carol R. (2005), *Fever of War: the Influenza Epidemic in the US Army during World War I*, New York: New York University Press.

Bynum, William E. (1993), Policing hearts of darkness: aspects of the International Sanitary Conferences, *History and Philosophy of the Life Sciences*, 15, 421-34.

Caldwell, John C. (2004), Demographic theory: a long view, *Population and Development Review* 30:2, 297-316.

Carmichael, Ann G. (1986), *Plague and the Poor in Renaissance Florence*, Cambridge: Cambridge University Press.

Carmichael, A.G. (1998), Epidemics and State Medicine in Fifteenth-Century Milan, pp. 221-47 in R. French *et al*. (eds.), *Medicine from the Black Death to the French Disease*, London: Ashgate.

Carmichael, A.G. and Millicent F. Moran (2002), Historical connections between climate, medical thought and human health, pp. 18-51 in P. Martens and A.J. McMichael (eds.), *Environmental Change, Climate and Health*, Cambridge: Cambridge University Press.

Carpenter, Kenneth J. (1986), *The History of Scurvy and Vitamin C*, Cambridge: Cambridge University Press.

Carter, K. Codell (2003), *The Rise of Causal Concepts of Disease: Case Histories*, Burlington, VT: Ashgate.

Champion, Justin A.I. (1993), Epidemics and the built environment in 1665, in J.A.I. Champion (ed.), Epidemic Disease in London, *Centre for Metropolitan History Working Papers Series*, no. 1. Accessible online: http://www.history.ac.uk/cmh/epichamp.html

Chaplin, Joyce E. (2001), *Subject Matter: Technology, the Body, and Science on the Anglo-American Frontier, 1500-1676*, Harvard University Press.

Chaudhuri, K.N. (1985), *Trade and civilisation in the Indian Ocean: an economic history from the rise of Islam to 1750*, Cambridge: Cambridge University Press.

Chaudhuri, K.N. (1990), *Asia before Europe: Economy and Civilisation of the Indian Ocean from the Rise of Islam to 1750*, Cambridge: Cambridge University Press.

Chew, Sing C. (2001), *World Ecological Degradation: Accumulation, Urbanization, and Deforestation, 2000 B.C.-A.D. 2000*, New York and Oxford: Rowman and Littlefield.

Christian, David (2004), *Maps of Time: An Introduction to Big History*, Berkeley and Los Angeles: University of California Press.

Ciano, Cesare (1976), *La sanità marittima nell'età medicea*, Pisa: Pacini.

Cliff, Andrew, and Peter Haggett (2004), Time, travel and infection, *British Medical Bulletin*, 69, 87-99.

Clifford, Richard J., S.J. (2003), The roots of apocalypticism in near eastern myth, pp. 3-29 in B.J. McGinn *et al*. (eds.), *The Continuum History of Apocalypticism*, New York: Continuum.

Cohen, Mark (1989), *Health and the Rise of Civilization*, New Haven: Yale University Press.

Crosby, Alfred (1972), *The Columbian Exchange: Biological and Cultural Consequences of 1492*, Westport, CT: Greenwood Press.

Crosby, Alfred (1986), *Ecological Imperialism: the Biological Expansion of Europe, 900-1900*, Cambridge: Cambridge University Press.

Crosby, Alfred (1989), *America's Forgotten Pandemic: the Influenza of 1918*, Cambridge: Cambridge University Press.

Crutzen, Paul J. and Will Steffen (2003), How long have we been in the anthropocene era?, *Climate Change*, 61, 251-257.

Davis, Mike (2001), *Late Victorian Holocausts: El Niño Famines and the Making of the Third World*, New York and London: Verso.

Davis, Mike (2006), *Planet of Slums*, New York: Verso.

Dean, Warren (1995), *With Broadax and Firebrand: The Destruction of the Atlantic Brazilian Forest*, Berkeley: University of California Press.

Deng, Kent G. (2003), China: Tang, Song and Yuan Dynasties, in J. Mokyr (ed.), *The Oxford Encyclopedia of Economic History*, 5 volumes (New York and Oxford: Oxford University Press), 2, 423-28.

Deng, Kent G. (2004), Unveiling China's true population statistics for the pre-modern era with official census data, *Population Review*, 43:2, section 1, 38 pages.

Dols, Michael (1977), *The Black Death in the Middle East*, Princeton: Princeton University Press.

Echenberg, Myron (2002), Pestis Redux: the initial years of the third bubonic plague pandemic, 1984-1901, *Journal of World History*, 13:2, 429-449.

Echeverri, Beatriz (2003), Spanish influenza seen from Spain, pp. 173-190 in H. Phillips and D. Killingray (eds.), *The Spanish Influenza Pandemic of 1918-19: New Perspectives*.

Elvin, Mark (1993), Three thousand years of unsustainable growth: China's environment from archaic times to the present, *East Asian History*, 6, 7-46.

Elvin, Mark (2004), *The Retreat of the Elephants: An Environmental History of China*, New Haven: Yale University Press.

Elvin, Mark (1998), Who was responsible for the weather? Moral meteorology in late imperial China, *Osiris*, 13, 213-237.

Ewald, Paul W. (1994), *Evolution of Infectious Disease*, Oxford: Oxford University Press.

Fagan, Brian (2000), *The Little Ice Age: How Climate Made History, 1300-1850*, New York: Basic Books.

Farmer, Paul (1999), *Infections and Inequalities: the Modern Plagues*, Berkeley and Los Angeles: University of California Press.

Farris, William Wayne (1985), *Population, Disease, and Land in Early Japan, 645-900*, Cambridge, MA: Harvard University Press.

Fidler, David P. (1999), *International Law and Infectious Diseases*, Oxford: Clarendon Press.

Fidler, David P. (2004), *SARS, Governance and the Globalization of Disease*, New York: Macmillan/Palgrave.

Fiennes, Richard N. T-W. (1978), *Zoonoses and the Origins and Ecology of Human Disease*, London and New York: Academic Press.

Floud, Roderick, Kenneth Wachter, and Annabel Gregory (1990), *Height, Health and History: Nutritional Status in the United Kingdom, 1750-1980*, Cambridge: Cambridge University Press.

Freedberg, David (2002), *The Eye of the Lynx: Galileo, His Friends, and the Beginnings of Modern Natural History*, Chicago: University of Chicago Press.

Glacken, Clarence (1967), *Traces on the Rhodian Shore*, Berkeley: University of California Press.

Grmek, M.D. (1980), Le Concept d'infection dans l'antiquité et au moyen age, les anciennes measures sociales contre les maladies contagieuses et la fondation de la première quarantaine à Dubrovnik (1377), *RAD Jugoslavenske akademije znanosti i umjetnosti* (Zagreb), no. 384, 27-54.

Grmek, Mirko Drazen (1989), *Diseases in the Ancient Greek World*, Baltimore: Johns Hopkins University Press.

Grove, Richard (1998), The East India Company, the Raj, and the El Niño: the critical role played by colonial scientists in establishing the mechanisms of global climate teleconnections, 1770-1830, pp. 301-323 in R.H. Grove, V. Damodaran and S. Sangwan (eds.), *Nature and the Orient: the Environmental History of South and Southeast Asia*, Delhi: Oxford University Press.

Guillerme, André (1988), The age of water: the urban environment in the North of France, A.D. 300-1800, *Environmental History Series*, n. 9, College Station: Texas A&M University Press.

Hamlin, Christopher (1998), *Public Health and Social Justice in the Age of Chadwick: Britain, 1800-1854*, Cambridge: Cambridge University Press.

Hardy, Anne (1993), Cholera, quarantine and the English preventative system, *Medical History*, 37, 250-69.

Ho, Ping-ti (1955), The introduction of American food plants into China, *American Anthropologist*, n.s. 57:2, 191-201.

Hoffman, Philip T., David Jacks, Patricia A. Levin, and Peter H. Lindert (2002), Real inequality in Europe since 1500, *Journal of Economic History*, 62:2, 322-55.

Howard-Jones, Norman (1975), *The Scientific Background of the International Sanitary Conferences, 1851-1938*, Geneva: WHO.

Haug, Gerald H. *et al*. (2003), Climate and the collapse of Maya civilization, *Science*, 299:5613, 1731-1735.

Hughes, J. Donald (1994), *Pan's Travail: Environmental Problems of the Ancient Greeks and Romans*, Baltimore: Johns Hopkins University Press.

Igler, David (2004), Diseased goods: global exchanges in the eastern Pacific basin, 1770-1850, *American Historical Review*, 109:3, 693-719.

Johnson, Niall, and Juergen Mueller (2002), Updating the accounts: global mortality of the 1918-1920 'Spanish' influenza pandemic, *Bulletin of the History of Medicine*, 76, 105-115.

Jones, David S. (2003), Virgin soils revisited, *William and Mary Quarterly*, 60:4, 703-742.

Jordan, William C. (1996), *The Great Famine: Northern Europe in the Early Fourteenth Century*, Princeton: Princeton University Press.

Kates, Robert W., B.L. Turner, II, and William C. Clark (1990), The great transformation, pp. 1-17 in B.L. Turner, II *et al*. (eds.), *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the past 300 Years*, Cambridge: Cambridge University Press.

Killingsworth, M. Jimmie and Jacqueline S. Palmer (1996), Millennial Ecology: The Apocalyptic Narrative from Silent Spring to Global Warming, pp. 21-45 in Carl Herndl and Stuart Brown (eds.), *Green Cultures: Environmental Rhetoric in Contemporary America*, Madison, WI: University of Wisconsin Press.

Komlos, J., and J. Baten (2004), Looking backward and looking forward: anthropometric research and the development of social science history, *Social Science History*, 28:2, 191-210.

Kunitz, Stephen J., and Stanley L. Engerman (1992), The ranks of death: secular trends in income and mortality, *Health Transition Review*, 2, supplementary issue, 26-46.

Kuriyama, Shigehisa (2000), Epidemics, weather, and contagion in traditional Chinese medicine, pp 3-22 in L. Conrad and D. Wujastyk (eds.), *Contagion: Perspectives from Pre-Modern Societies*, Aldershot: Ashgate.

La Berge, Anne (1992), *Mission and Method: the Early Nineteenth-Century French Public Health Movement*, Cambridge: Cambridge University Press.

Landers, John (2003), *The Field and the Forge: Population, Production and Power in the Pre-industrial West*, Oxford: Oxford University Press.

Langford, Christopher (2002), The age pattern of mortality in the 1918-19 influenza pandemic: an attempted explanation based on data for England and Wales, *Medical History*, 46, 1-20.

Lavely, William, and R. Bin Wong (1998), Revising the Malthusian narrative: the comparative study of population dynamics in late imperial China, *Journal of Asian Studies*, 57:3, 714-748.

Lee, J.Z, and Wang, F. (1999), *One Quarter of Humanity: Malthusian Mythology and Chinese Realities, 1700-2000*, Cambridge, MA: Harvard University Press.

Livi-Bacci, Massimo (2001), *A Concise History of World Population*, Third edition, trans. C. Ipsen, Oxford: Blackwell.

Livi-Bacci, Massimo (2003), Return to Hispaniola: reassessing a demographic catastrophe, *Hispanic American Historical Review*, 83:1, 3-51.

Maddison, Angus (2003), *The World Economy: Historical Statistics*, Paris: OECD.

Maglen, Krista (2002), The first line of defence: British quarantine and the port sanitary authorities in the nineteenth century, *Social History of Medicine*, 15:3, 413-28.

Marks, Robert B. (1998), It never used to snow: Climatic variability and harvest yields in late-imperial south China, 1650-1850, pp. 411-446 in Mark Elvin and Liu Ts'ui-jung (eds.), *Sediments of Time: Environment and Society in Chinese History*, ed. Cambridge University Press.

McCracken, Kevin, and Peter Curson (2003), Flu downunder: a demographic and geographic analysis of the 1919 epidemic in Sydney, Australia, pp. 110-151 in H. Phillips and D. Killingray (eds.), *The Spanish Influenza Pandemic of 1918-19: New Perspectives*.

McGinn, Bernard J. *et al*. (eds.) (2003), *The Continuum History of Apocalypticism*, New York: Continuum.

McMichael, Anthony J. (2001), *Human frontiers, environments and disease: past patterns, uncertain futures*, Cambridge: Cambridge University Press.

McNeill, John R. (1994), Of rats and men: a synoptic environmental history of the Island Pacific, *Journal of World History*, 5, 299-349.

McNeill, John R. (2004), Woods and warfare, *Environmental History*, 9:3, 388-410.

McNeill, John R. (2005), Yellow Jack and geopolitics: environment, epidemics, and the struggles for empire in the American tropics, 1650-1900, pp. 193-206 in J.M. Diefendorf and K. Dorsey (eds.), *City, Country, Empire: Landscapes in Environmental History*, Pittsburgh: University of Pittsburg Press.

McNeill, William H. (1976), *Plagues and Peoples*, Garden City, NJ: Anchor Press.

Mithen, Steven (2004), *After the Ice: A Global Human History, 20,000-5,000 BC*, Cambridge, MA: Harvard University Press.

Mokyr, Joel (2002), *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton: Princeton University Press.

Morineau, Michael (1998), Malthus: there and back from the period preceding the Black Death to the Industrial Revolution, *Journal of European Economic History*, 27:1, 137-202.

Morrison, Kathleen D. (2000), Naturalizing disaster: from drought to famine in southern India, pp. 21-34 in Bawden and Reycraft (eds.), *Environmental Disaster and the Archaeology of Human Response*, Albuquerque: University of New Mexico Press.

[NAC] Public Health Agency of Canada's National Advisory Committee on SARS and Public Health (2003), *Learning from SARS: Renewal of Public Health in Canada*, Ottawa: Health Canada.

NAS. (2004), *The Threat of Pandemic Influenza: Are We Ready?*, Washington: National Academies Press.

[NAS/IOM] (2004), National Academy of Sciences/Institute of Medicine, *Learning from SARS: Preparing for the Next Disease Outbreak*, Washington: National Academies Press.

Nunn, Patrick D. (1990), Recent environmental changes on Pacific islands, *The Geographical Journal*, 156:2, 125-140.

Nunn, Patrick D. and James M.R. Britton (2001), Human-environment relationships in the Pacific Islands around AD 1300, *Environment and History*, 7:1, 3-22.

Oxford, John S. (2001), The so-called Great Spanish Influenza Pandemic of 1918 may have originated in France in 1916, *Philosophical Transactions of the Royal Society of London, B,* 356, 1857-59.

Panzac, Daniel (1986), *Quarantines et lazarets: l'Europe et la peste d'Orient (XVIIe-XXe siècles)*, Aix-en-Provence: Édisud.

Patterson, K. David, and Gerald E. Pyle (1991), The geography and mortality of the 1918 influenza pandemic, *Bulletin of the History of Medicine*, 65, 4-21.

Petrarca, Francesco (1975), *Rerum familiarum libri I-VIII*, trans. Aldo S. Bernardo, Albany: State University of New York Press.

Phillips, Howard, and David Killingray (eds.) (2003), *The Spanish Influenza Pandemic of 1918-19: New Perspectives*, London: Routledge.

Rallu, Jean-Louis (1990), *Les populations océaniennes aux XIXe et XXe siècles*, Institut National d'Études Démographiques, Paris: Presses Universitaires de France.

Richards, John F. (2003), *The Unending Frontier: An Environmental History of the Early Modern World*, Berkeley and Los Angeles: University of California Press.

Riley, James C. (2001), *Rising Life Expectancy: a Global History*, Cambridge: Cambridge University Press.

Riley, James C. (1989), *Sickness, Recovery and Death: a History and Forecast of Ill Health*, Iowa City: University of Iowa Press.

Rocca, Fiammetta (2003), *The Miraculous Fever-Tree: Malaria and the Quest for a Cure that Changed the World*, New York: Harper Collins.

Rohr, Christian (2003), Man and natural disaster in the late Middle Ages: the earthquake in Carinthia and northern Italy on 25 January 1348 and its perception, *Environment and History*, 9, 127-49.

Rothenberg, Gunther (1973), The Austrian sanitary cordon and the control of the bubonic plague, 1710-1871, *Journal of the History of Medicine and Allied Sciences*, 28, 15-23.

Ruddiman, William F. (2003), The anthropogenic greenhouse era began thousands of years ago, *Climate Change*, 61, 261-293.

Ruddiman, W.F., and A.G. Carmichael, Pre-industrial depopulation, atmospheric carbon dioxide, and global climate, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 158-194, The Pontifical Academy of Sciences, Vatican City, 2006.

Rusnock, Andrea A. (2002), *Vital Accounts: Quantifying Health and Population in Eighteenth-Century England and France*, Cambridge: Cambridge University Press.

Schamiloglu, Uli (1993), Preliminary Remarks on the Role of Disease in the History of the Golden Horde, *Central Asian Survey*, 12:4, 447-457.

Schamiloglu, Uli (2004), The Rise of the Ottoman Empire: The Black Death in Medieval Anatolia and its Impact on Turkish Civilization, pp. 255-279 in *Views From the Edge: Essays in Honor of Richard W. Bulliet*, Neguin Yavari, Lawrence G. Potter, and Jean-Marc Oppenheim (eds.), New York: Columbia University Press.

Schiebinger, Londa (2005), Prospecting for drugs: European naturalists in the West Indies, pp. 119-133 in L. Schiebinger and Claudia Swan (eds.), *Colonial Botany: Science, Commerce, and Politics in the Early Modern World*, Philadelphia: University of Pennsylvania Press.

Sen, Amartya Kumar (1981), *Poverty and famines: an essay on entitlement and deprivation*, Oxford: Oxford University Press.

Sherrington, Charles S. (1940), *Man on his Nature, The Gifford Lectures, Edinburgh, 1937-8*, Cambridge: Cambridge University Press.

Sieferle, Rolf Peter (2001), *The Subterranean Forest: Energy Systems and the Industrial Revolution*, Oxford: White Horse Press.

Smallman-Raynor, M.R., and A.D. Cliff (2004), *War Epidemics: an Historical Geography of Infectious Diseases in Military Conflict and Civil Strife, 1850-2000*, Oxford: Oxford University Press.

Smil, Vaclav (1994), *Energy in World History*, Boulder, CO: Westview Press.

Smoller, Laura A. (2000), On earthquakes, hail, frogs, and geography: plague and the investigation of the Apocalypse in the later Middle Ages, pp. 156-187 in C.W. Bynum and P. Freedman (eds.), *Last Things: Death and the Apocalypse in the Middle Ages*, Philadelphia: University of Pennsylvania Press.

Stathakopoulos, D.C. (2004), *Famine and Pestilence in the late Roman and Early Byzantine Empire: a systematic survey of subsistence crises and epidemics*, London: Ashgate.

Steckel, Richard H. (2004), New light on the 'Dark Ages': the remarkably tall stature of northern European men during the medieval era, *Social Science History*, 28:2, 211-29.

Szreter, Simon (1997), Economic growth, disruption, deprivation, disease and death: on the importance of the politics of public health for development, *Population and Development Review*, 23:4, 693-728.

TeBracke, William H. (1975), Air pollution and fuel crises in preindustrial London, 1250-1650, *Technology and Culture*, 16:3, 337-59.

Troesken, Werner (2004), *Water, Race and Disease*, Cambridge, MA: MIT Press.

Tuan, Yi-Fu (1979), *Landscapes of Fear*, Minneapolis: University of Minnesota Press.

Twitchett, D. (1979), Population and pestilence in T'ang China, pp. 35-68 in Wolfgang Bauer (ed.), *Studia Sino-Mongolica: Festschrift für Herbert Franke*, Wiesbaden: Franz Steiner Verlag.

UN-Habitat (2003), *The Challenge of Slums: Global Report on Human Settlements, 2003*, United Nations Human Settlements Programme, London and Sterling, VA: Earthscan Publications.

Van Hartesveldt, Fred R. (ed.) (1992), *The 1918-1919 Pandemic of Influenza: the Urban Impact in the Western World*, Lewiston: Edwin Mellen Press.

Watkins, Susan Cotts, and Jane Menken (1988), The role of crises in historical perspective: reply, *Population and Development Review*, 14:1, 165-170.

Weiss, Robin A. and Angela R. McLean (eds.) (2004), Emerging infections: what have we learnt from SARS?, a special issue of *Philosophical Transactions of the Royal Society of London, B (Biological Sciences)*, 359, no. 1447, 1047-1140.

Wilkinson, Lise (1992), *Animals and Disease: an Introduction to the Study of Comparative Medicine*, Cambridge: Cambridge University Press.

Williams, Michael (2003), *Deforesting the Earth: From Prehistory to Global Crisis*, Chicago: University of Chicago Press.

Winter, J.M. (1977), The impact of the first world war on civilian health in Britain, *The Economic History Review*, n.s. 30:3, 487-507.

Whitmore, Thomas M. *et al*. (1990), Long-term population change, pp. 25-39 in B.L. Turner II *et al*., *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*, Cambridge: Cambridge University Press.

Wrigley, E.A. (2000), The divergence of England: the growth of the English economy in the seventeenth and eighteenth centuries, *Transactions of the Royal Historical Society*, ser. 6, 10, 117-41.

Zylberman, Patrick (2003), A holocaust in a holocaust: the Great War and the 1918 Spanish influenza epidemic in France, pp. 191-201 in H. Phillips and D. Killingray (eds.), *The Spanish Influenza Pandemic of 1918-19: New Perspectives*.

ATMOSPHERIC BROWN CLOUDS: HEALTH, CLIMATE AND AGRICULTURE IMPACTS

VEERABHADRAN RAMANATHAN

1. Aerosols or Particles in the Air

Aerosols or atmospheric particles are either in solid form, such as dust or sea salt, or in liquid form, such as sulfates, nitrates or organics dissolved in water. Nature produces aerosols in the form of sea salt, mineral dust, sulfates and nitrates. Human activities either emit gaseous precursors, such as SO_2 and NO_x which get converted into particles through chemical processes, or directly emit particles such as soot, which is a mixture of elemental carbon and organics. The concentrations of natural aerosols in pristine air are typically around 100 to 1000 particles per cc in continental air and about 100 to 500 per cc in marine air. It is now increasingly difficult, if not impossible, to find such pristine air. In most regions of the northern hemisphere, including over the oceans, the concentrations are larger by factors ranging from two to ten. Humans add particles to the air by a variety of actions, including the burning of biofuels and fossil fuels for cooking purposes, the consumption of fossil fuel for energy, the clearing of forests, and land use modification, to name a few.

2. Atmospheric Brown Clouds

Coincident with greenhouse gas warming is the appearance of atmospheric brown clouds. If greenhouse gases, such as CO_2, are the ultimate end product of fossil fuel burning, then particulates in the air represent an intermediate phase. The brownish color in the haze (Fig. 1, see page 397) is due to the absorption and the scattering of visible solar radiation by black carbon, fly ash, soil dust and NO_2. During the Indian Ocean

Experiment (INDOEX), widespread brown clouds, about 3 km thick, were found over most of the Arabian Sea, Bay of Bengal and N. Indian Ocean (Ramanathan *et al.*, 2001a). Subsequently, a new NASA satellite instrument (MODIS) has identified (Kaufman *et al.*, 2002 and Ramanathan and Ramana, 2003) such widespread brown clouds over most industrialized regions of the world (see Figure 2, page 397; also see Molina and Molina, 2002).

Particle trajectories using observed winds clearly reveal how particles emitted in one continent can travel to another continent across the oceans in about a week (Fig. 3, see page 398).

This is significant because the lifetimes of most aerosols are in the order of one to two weeks. For example, model simulations (Fig. 4, see page 398) show that black carbon emitted in East Asia can travel to N. America and increase BC concentrations over N. America by as much as 75%. The fact that indeed such transport was occurring was shown by Roberts *et al.*, (2006) using aircraft data (Fig. 5). The aircraft data for the

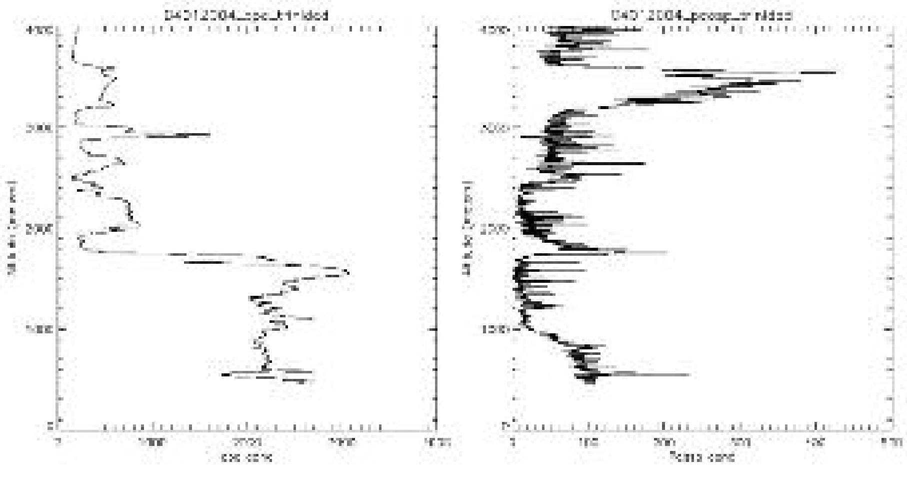


Figure 5. Long-range transport of E. Asian aerosols as measured off the coast of Trinidad in N. California on April 1, 2004 from a research aircraft during a campaign conducted by the author. The left hand panel shows the total aerosol number concentration, while the right hand panel shows the aerosol concentration between 0.1 micrometer to 3.5 micrometer. The unit for the horizontal axis is number/cm3. The vertical axis is altitude in meters. The particles due to long-range transport are seen in the layer between 3000 m to 3700 m altitude.

concentration of aerosol number concentration was collected over the west coast of the USA adjacent to the Sierra mountain range (shown on the background photo) and reveals that the transport occurs in narrowly confined layers (see the concentration spikes seen at 3 km in the left panel and at 3.5 km in the right hand panel).

In February 1999, over 200 scientists from Europe, India and the USA gathered in the islands of the Maldives to conduct the Indian Ocean Experiment (INDOEX; http://www.indoex.ucsd.edu/publications/). The interdisciplinary and comprehensive data gathered from several aircraft, ships, surface stations and satellites (Ramanathan *et al.*, 2001a) helped forewarn the world (UNEP Report 2002) of a potentially major environmental problem facing the Asian region and the world. INDOEX revealed the so-called brown cloud phenomenon (see Fig. 1 from Ramanathan and Ramana, 2002) spreading from the Himalayas all the way over the N. Indian Ocean region. INDOEX collected direct evidence that manmade particles travel several thousand kilometers across the ocean (Fig. 6, see page 399). Particles were captured on filter papers from aircraft flying into the plume and subsequently analyzed in the laboratory (courtesy of J. Anderson, Arizona State Univ.). Photographs from corresponding regions were taken from the aircraft. Fig. 6 shows clearly that the haze layer extends from the Arabian sea (7°N in the top panel) into the southern Indian Ocean up to about 6.6°S, and microscope pictures reveal soot attached to the particles in the brown clouds, whereas particles from the pristine southern Indian Ocean (bottom panel) south of 8°S, which is free of haze (as revealed by the blue skies seen in the photo) is also free of soot as revealed by lack of irregular soot clusters on the particles. The particles on this electron microscope image are natural sea salt particles from the sea surface.

The major finding of INDOEX was that the brown clouds reduced the seasonal averaged solar radiation reaching the sea surface by as much 10% (Satheesh and Ramanathan, 2000) and this phenomenon is discussed later in this paper under the 'Dimming Section'. The brown cloud induced dimming has a large impact on the radiative heating of the region (Ramanathan *et al.*, 2001a and b). The impact on the regional gas phase chemistry of the region (Lelieveld *et al.*, 2001) was also large. The fundamental message from INDOEX is that due to long-range transport, what we normally associate with urban haze can span an entire continent plus an ocean basin (see Fig. 2, page 397). Both fossil fuel and biomass burning contribute to the particles in the haze. The persistence of the haze during the long dry season from November to May, its black carbon

content, the large perturbation to the radiative energy budget of the region and its simulated impact on the monsoon rainfall distribution have significant implications to the regional and global water budget, agriculture and health (Ramanathan *et al.*, 2001b). The logical implication is that air pollution and climate changes are intricately linked and should be addressed under one common framework, which gave rise to the creation of the Atmospheric Brown Clouds project by UNEP (see Ramanathan and Crutzen in http://www.atmosphericbrowncloud.ucsd.edu/).

Why is the air pollution problem more widespread and acute in the Tropics? The brown haze is a particularly severe problem over the tropics due to a variety of reasons. The large increase in emissions of aerosols and their precursors is an important reason. For example emissions of SO_2 in South Asia have increased by a factor of 3 to 4 since 1970. However, SO_2 emissions from S. Asia are only 25% of the US emissions. Hence other factors have to be invoked to account for the thickness and extent of the haze. Organic and black carbon and fly ash contribute more than SO_2 to the haze in Asia. The other important contributor is the unique meteorology of the tropics and the subtropics (including S. Asia) which leads to a long dry season extending from late fall and winter until spring. The dryness is caused by subsidence, which precludes the wet removal of haze particles by rain. In the mid and high latitudes, the absence of a long dry season, and seasonally distributed rainfall (and snow fall) cleans the atmosphere more efficiently.

3. Direct Impacts of ABCs

3.1. *Human Health*

The most direct and important influence of aerosols is on human health. Aerosols are inhaled and ingested in the lungs and cause acute respiratory infections, chronic obstructive pulmonary disease and lung cancer. Public health experts have long suspected that particulates cause respiratory and other health problems. What is new is that they are now linked convincingly with fatalities. Worldwide, aerosols resulting from indoor and outdoor air pollution lead to about 1.6 million premature deaths annually (Smith, 2002). In India alone about 400 to 550 premature deaths are attributed to inhalation of air pollution and particles from the use of biomass fuels for cooking (Smith, 2000).

3.2. *Reduction of Photosynthetically Active Radiation (PAR) at the Surface*

Another potentially important environmental effect of ABCs is their large effect in reducing the total (direct + diffuse) PAR (Fig. 7 from Meywerk and Ramanathan, 2002). The brown clouds over the Arabian Sea decreased direct PAR by 40% to 70%, but enhanced the diffuse PAR substantially, with a net reduction in total PAR by as much 10% to 30%. The potential impact of large reductions in direct PAR and corresponding enhancements in diffuse PAR accompanied by net reduction in total (direct + diffuse) PAR on marine and terrestrial photosynthesis and on agriculture productivity (Chameides *et al.*, 2002; Stanhill and Cohen, 2001) have not been adequately studied. Similar changes in the UV spectrum due to ABCs need to be established which may have health effects due to the potential importance of UVB in producing Vitamin D (Garland and Garland, 1980; Gorham *et al.*, 1989).

Nucleation of more cloud drops and suppression of precipitation efficiency: Aerosols, in particular sulfates, larger than 0.1μm provide the necessary nuclei for all cloud drops and ice crystals. Organic aerosols also serve as cloud condensation nuclei in par with the role of sulfate particles. Trace amounts of soluble gases and organic substances in air pollution can ampli-

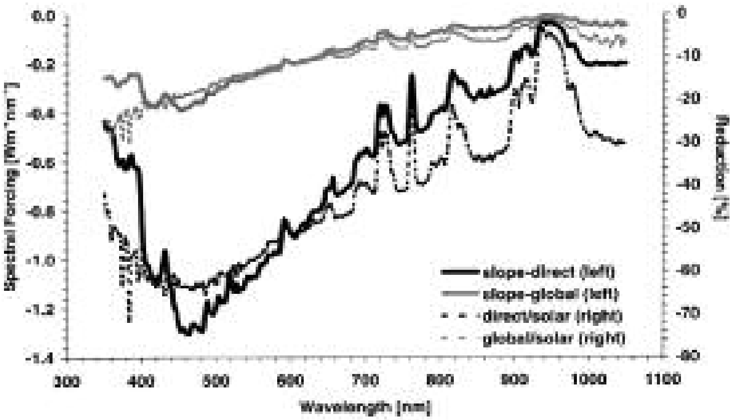


Figure 7. Decrease in solar radiation as a function of wavelength due to brown clouds as measured directly by a grating spectrometer in the Arabian sea during INDOEX. As shown the maximum decrease is seen in the PAR (400 to 700 nm) region. Source: Meywerk and Ramanathan, 2003.

fy the CCN activity of small aerosols and have led to an increase in the number of cloud drops. There is now a substantial body of in-situ aircraft observations of continental and marine clouds that show that anthropogenic aerosols enhance CCN and cloud drop number density (Fig. 8, see page 399). By nucleating more cloud drops, aerosols increase the reflection of solar radiation by clouds, which adds to the surface cooling effect. This effect is known as the *indirect forcing*. In INDOEX it was shown by direct aircraft measurements that the trade cumulus and strato-cumulus clouds in the polluted Arabian Sea had six times as many cloud drops as the pristine clouds south of the ITCZ. Another major effect of increasing the concentrations of cloud droplets is a reduction in the size of drops since they are competing for the water available for condensation. Increase in aerosols can nucleate copious amounts of small droplets. The small drops coalesce very inefficiently into raindrops, which can inhibit the formation of larger raindrops and decrease precipitation efficiency (Andreae *et al.*, 2004), one consequence of which is suppression of rain over polluted regions. This was recently confirmed by satellite observations from the Tropical Rainfall Measuring Mission (TRMM), showing tracks of reduced cloud particles emanating from forest fires and from pollution sources such as coal power plants, refineries, smelters and urban areas (Rosenfeld, 2000).

3.3. *Wet Deposition (Acid Rain)*

Aerosols such as sulfates, nitrates and organic acids are incorporated into raindrops either as nuclei or by scavenging, and are removed from the atmosphere during rainfall. A raindrop in a clean atmosphere has a pH of about 5.6, but during wet deposition of pollution aerosols the pH drops, i.e., rain becomes acidic. There is a vast literature on environmental damages due to acid rain (e.g., see review in Seinfeld and Pandis, 1998), and the acid rain problem constitutes another major environmental threat from brown clouds (e.g. see Molina and Molina, 2002).

3.4. *Global Dimming*

Aerosols, by scattering/absorbing solar radiation and emitting/absorbing long wave (IR) radiation, change the radiation fluxes at the surface and the top of the atmosphere, thereby significantly perturbing the atmospheric absorption of solar radiation. These aerosol-induced changes in the radiation budget are referred to as the *direct forcing*. At the surface, aerosols decrease

the direct solar beam and enhance the diffuse solar radiation, and both of these effects have been measured during INDOEX (Ramanathan *et al.*, 2001a). Black carbon, which strongly absorbs solar radiation, plays a major role in the forcing by partially shielding the surface from the intense tropical solar radiation. This shielding effect of BC amplifies the surface radiative forcing due to all other manmade aerosols (sulfates, organics, nitrates, fly ash) by a factor of two or more in cloudy skies. BC over the northern Indian Ocean and the Arabian Sea (during the INDOEX measurement campaign) contributed as much as 10% to 14% to aerosol mass (Mayol-Bracero *et al.*, 2002), compared with about 5% in the suburban regions of Europe and N. America. The black carbon and other species in the haze reduce the average solar radiative heating of the ocean by as much as 10% and enhance the atmospheric solar radiative heating by 50 to 100%. For comparison, the perturbation to the regional IR radiation budget (the greenhouse effect) by the observed increase in CO_2 and other greenhouse gases is about 3 W m^{-2}, equivalent to a 1.5% increase in the solar heating of the ocean.

Large reduction of the seasonal averaged solar radiation of the order of 10% or larger (due to anthropogenic aerosols) is not restricted to S. Asia. It has now been observed in many regions of the planet including E. Asia, the Atlantic, Western Pacific, Mediterranean, in Europe, N. and S. America and Africa (see summary in Ramanathan and Ramana, 2003). More recently we have observed reduction of 10 to 15% in the Himalayas. Given such large observed reduction in solar radiation due to brown clouds, we would anticipate that the planet has become dimmer with time, since emissions of brown cloud particles (notably sulfates and black carbon) have changed substantially with time (Fig. 9, see page 400). Long-term negative trends in surface solar irradiance have been observed by surface radiometers worldwide over land. The reported trends in the annual mean irradiance vary from -5% (10 W m^{-2}) between 1958 to 1985 for all land stations (*48*) to about -1 to -3% per decade for the last four decades over many of the 1500 stations in the GEBA data sets (Stanhill and Cohen, 2001). More recently it has been observed (Wild *et al.*, 2005) that several stations in N. America and western Europe have seen a reversal in the dimming trend since the 1980s, consistent with reductions in sulfur and BC emissions (Fig. 9). However, India and China are witnessing an unabated continuation of the dimming trend into the twentieth century. Surface solar insolation over Indian stations has decreased by 10 W m^{-2} from 1970 to 2004 (Ramanathan *et al.*, 2005), and in China it has decreased by about 15 W m^{-2} from 1954 to 2001 (Qian *et al.*, 2006).

The effect of ABCs on surface solar radiation is prevalent over most parts of the globe today and has been estimated by us for the period 2000 to 2004 using modern satellite and surface data gathered around the world (Fig. 10, page 400). It ranges from a reduction of 3 to 12 W m^{-2} over many parts of N. America and Europe and from 6 to 25 W m^{-2} over Asia, Africa and S. America. The potential consequences of this dimming effect on climate are discussed next.

4. Impact on Climate and Water Budget

4.1. *Masking of Global Warming*

The surface dimming due to ABCs has a surface cooling effect, which should have masked some of the surface warming due to greenhouse gases. The magnitude of this masking effect is one of the central problems in climate change with significant implications for future climate changes and policy responses to global warming (Andreae *et al.*, 2005). The extent of global warming is not fully reflected in the Earth's observed surface temperatures. The additional heat trapped by the increase in greenhouse gases from the late nineteenth century to the present time should have committed the planet to a global warming in the range of 1°C to 3°C (see Ramanathan, 1988 for an explanation of the term *committed warming*). The observed global surface warming is only about 0.6 K, i.e., only about 20% to 60% of the committed warming (depending on whether we use 3 K or 1 K for the committed warming). Some of this warming has been masked by the dimming due to brown clouds and the remaining heat is stored in the depths of the ocean to be released in the coming decades to centuries. Through the process of convective overturning, oceans transfer infrared energy to their deepest layers and hold the heat, delaying the impact of global warming. Whether this stored heat will warm the atmosphere in a few decades or a few centuries is unknown.

Current estimates of the ABC masking effect range from 30% to as high as 75% (Crutzen and Ramanathan, 2003; also see Andreae *et al.* in this issue for a more detailed consideration of the aerosols' masking effect). The fundamental issue with this large range of uncertainty of the masking effect is that, policies to reduce ABCs (due to their effects on health, acid rain and agriculture) will unmask the cooling effect of ABCs quickly and have a potentially large effect in the acceleration of global warming. This is

because the lifetime of ABCs is a few weeks and hence their masking effect will diminish as soon as their emissions are curtailed, whereas GHGs will respond to emission reduction policies on decadal to century time scales. Without a better quantification of the masking effect of ABCs, we would not know whether warming in the coming decades (if it were to occur) is due to the unmasking effect of ABCs (by efforts of cleaning up in Asia for example), or due to release of stored greenhouse heating in the oceans, or continued increase in emission of greenhouse gases. In the meantime, every decade we delay in taking action, we are committing the planet to additional warming that future generations have to deal with.

4.2. *Global Scale Drying*

Global mean land average precipitation (P) decreased by about 2.4% from 1947 to 1996 (Hulme *et al.*, 1998). The long-term precipitation trends for the global oceans are unknown due to lack of data. The large warming witnessed during the last 50 years was not accompanied by an increase in precipitation. On the other hand, climate models which include greenhouse forcing suggest an increase in global rainfall with surface warming since the warmer oceans evaporate more moisture which subsequently reaches the surface as rainfall. Increased rainfall does not automatically imply increased availability of surface water, however, since in a warmer climate evaporation of surface moisture is also accelerated.

The primary mechanism by which aerosols reduce global mean rainfall is by reducing the solar radiation reaching the surface of the earth (supporting online material). Direct observations revealed that the presence of absorbing black carbon magnified by a factor of 2 to 3, the reduction of solar radiation due to sulfates and other scattering aerosols. A reduction in the surface solar radiation can slow down the atmospheric hydrological cycle and reduce rainfall because roughly 75% of the globally averaged radiative heating of the surface is balanced by evaporation.

4.3. *Regional Climate, Monsoon Rainfall and Water Budget Effects*

In addition to their effect on global mean rainfall, ABCs alter the spatial gradient in surface and atmosphere solar heating (see Fig. 10) which leads to significant regional impacts, particularly on the S. Asian monsoon impacting over 2 billion Asians (Ramanathan *et al*. 2005, hereafter referred to as R2005). The link between aerosols and rainfall has been

suggested for selected regions, including the well-known Sahelian drought (Rotstayn and Lohmann, 2002), the north-south shift in the East Asian monsoon (Menon *et al.*, 2002), and the South Asian region (Chung *et al*,. 2002; R2005). I will focus on the impact of ABCs on the S. Asian monsoon since it has not only been examined extensively by us (Chung *et al.*, 2002; Chung and Ramanathan, 2003; R2005; Chung and Ramanathan 2006) but also impacts over a billion S. Asians. ABCs impact the monsoon circulation and climate through several distinctively different processes.

1) The dimming of the N. Indian Ocean and S. Asia decreases surface evaporation from the ocean (see Fig. 2 of R2005) and since the primary fuel for the monsoon rainfall is evaporation of moisture from the N. Indian Ocean (including the Arabian Sea and Bay of Bengal), this has a direct influence in reducing monsoon circulation and rainfall.

2) The dimming by ABCs is accompanied by a large increase in atmospheric solar heating (by 50% to 100%) which tends to stabilize the surface-atmosphere system by cooling the surface and warming the atmosphere. This redistribution of sunlight causes warmer air to overlie the colder surfaces. This increase in atmospheric stability, which has been observed using microwave satellite data (see Fig. 4 of R2005), tends to suppress convective cloud formation and rainfall. The increased atmospheric stability during the dry season can also decrease the ventilation of the pollution away from the boundary layer leading to increased occurrences of pollution events and fog formation.

3) Since ABCs and their dimming effects are predominantly concentrated in the N. Indian Ocean (NIO), Arabian Sea and Bay of Bengal (due to their proximity to pollution sources in S and SE Asia) with minimal effects in the S. Indian Ocean (see Fig. 10), ABC induced dimming suppresses the sea surface temperatures (SSTs) north of the equator while allowing the S. Indian Ocean (SIO) to warm in response to the greenhouse warming (R2005). Observed sea surface temperatures reveal that this differential masking of the N. Indian Ocean SSTs is indeed happening (Chung and Ramanathan 2006). For example, 50 year trends (from 1950 to 2000) in observed SSTs reveal tropical SIO warming by as much as 0.7 K while the trend is much smaller in the NIO with an almost zero warming trend in the N. Arabian Sea and N. Bay of Bengal. Since deep convection and tropical rain bands gravitate toward warmer oceans, climate model simulations suggest that the differential warming trend has led to a slowing down of the monsoon circulation resulting in decreased rainfall over land regions (S. Asia) and increase in oceanic rainfall over the southern tropical Indian Ocean.

The monsoon rainfall over India has decreased since the 1950s and our studies suggest that ABCs may be the primary driver of this negative trend. Clearly more studies of this important influence of ABCs are required to confirm these findings. Impact on the monsoon circulation has potentially a large impact on S. Asian agriculture, since monsoon rainfall correlates linearly with rice production (Webster *et al.*, 2002). The impact of the haze on monsoon rainfall provides another strong rationale for reducing air pollution in the developing nations.

SUMMARY

Manmade particles in ABCs lead to about 1.6 million deaths annually, decrease photosynthetically active solar radiation, have direct agricultural impacts, lead to a large dimming of the planet, decrease the precipitating efficiency of clouds, stabilize the surface-atmosphere system during dry seasons, alter sea surface temperature gradients, alter regional rainfall patterns, and lead to global cooling and drying. In addition, the ASIAN monsoon system affecting the lives of over 2 billion people is vulnerable to ABCs. The only positive thing that can be said about ABCs is that their lifetime is only about a few weeks long which implies that actions to regulate their emissions will result in immediate benefits.

REFERENCES

Andreae, M.O., Rosenfeld, D., Artaxo, P., Costa, A.A., Frank, G.P., Longo, K.M., and Silva-Dias, M.A.F., Smoking rain clouds over the Amazon, *Science*, 303, 1337-1342, 2004.

Andreae, M.O., Jones, C.D., and Cox, P.M., Strong present-day aerosol cooling implies a hot future, *Nature*, 435, 1187-1190, 2005.

Andreae, M.O., U.E.C. Confalonieri, A.J. McMichael, D.E. Bloom, D.L. Heymann, P. Martens, W. Steffen, and M.E. Wilson, Global Environmental Change and Human Health, *In this issue* (pp. 374-393), 2006.

Chameides, W.L., *et al.*, Case study of the effects of atmospheric aerosols and regional haze on agriculture: An opportunity to enhance crop yields in China through emission controls?, *Proceedings of the National Academy of Sciences of the United States of America*, 96, 13,626-13,633, 1999.

Chung, C.E., V. Ramanathan, and J.T. Kiehl, Effects of the South Asian Absorbing Haze on the Northeast Monsoon and Surface-Air heat exchange, *J. Climate*, 15, 2462-2476, 2002.

Chung, C.E., and V. Ramanathan, South Asian Haze Forcing: Remote Impacts with Implications to ENSO and AO, *J. Climate*, 2003.

Chung, C.E., V. Ramanathan, D. Kim, and I.A. Podgorny, Global Anthropogenic Aerosol Direct Forcing Derived from Satellite and Ground-Based Observations, *J. Geophys. Res.*, 110, D24207, doi: 10.1029/2005JD006356, 2005.

Chung C.E., and V. Ramanathan, Weakening of N. Indian SST gradients and the monsoon rainfall in India and the Sahel, *J. Climate*, In Press, 2006.

Crutzen, P.J., and V. Ramanathan, The parasol effect in climate, *Science*, 302, 1679-1680, 2003.

Garland, C. and F. Garland, Do sunlight and vitamin D reduce the likelihood of colon cancer?, *Int. J. Epidemiology*, 9, 227-31, 1980.

Gorham, E., C. Garland, and F. Garland, Acid haze air pollution and breast and colon cancer in 20 Canadian cities, *Canadian J. Public Health*, 80, 96-100, 1989.

Hulme, M., T.J. Osborn, and T.C. John, Precipitation sensitivity to global warming: Comparison of observations with HadCM2 simulations, *Geophys. Res. Lett.*, 25, 3379-3382, 1998.

Kaufman Y.J., Remote sensing of direct and indirect aerosol forcing, in: *Aerosol Forcing of Climate*, R.J. Charlson and J. Heintzenberg, John Wiley & Sons, 297-332, 1995.

Kaufman, Y.J., D. Tanre, and O. Boucher, A satellite view of aerosols in the climate system, *Nature*, 419, 215-223, 2002.

Krishnan, R., and V. Ramanathan, Evidence of surface cooling from absorbing aerosols, *Geophys. Res. Lett.*, 29, 10.1029/2002GL014687, 2001.

Lelieveld, J *et al.*, The Indian Ocean Experiment: Widespread Air Pollution from South and Southeast Asia, *Science*, 2001, 291, 1031-1036, 2001.

Mayol-Bracero, O.L., Gabriel, R., Andreae, M.O., Kirchstetter, T.W., Novakov, T., Ogren, J., Sheridan, P., and Streets, D.G., Carbonaceous aerosols over the Indian Ocean during the Indian Ocean Experiment (INDOEX): Chemical characterization, optical properties and probable sources, *Journal of Geophysical Research*, 107, 8030, doi:10.1029/2000JD000039, 2002.

Menon, S., J. Hansen, L. Nazarenko, and Y. Luo, Climate Effects of Black Carbon Aerosols in China and India, *Science*, 297, 2250-2253, 2002.

Meywerk, J., and V. Ramanathan, 2002: Influence of Anthropogenic Aerosols on the Total and Spectral Irradiance on the Sea Surface During INDOEX, *J. Geophys. Res. Atmospheres*, 107, D19, 8018, doi: 10.1029/2000JD000022, 2002. 17-1 to 17-14, 1999.

Molina, L.T. and M. Molina, *Air Quality in the Mexico Mega City*, Kluwer Academic Publishers, pp. 379, 2002.

Novakov, T., V. Ramanathan, J.E. Hansen, T.W. Kirchstetter, Mki. Sato, J.E. Sinton, and J.A. Satahye, Large historical changes of fossil-fuel black carbon aerosols, *Geophys. Res. Lett.*, 30, doi: 10.1029/2002GL 016345, 2003.

Qian, Y., D.P. Kaiser, L.R. Leung, and M. Xu, More frequent cloud-free sky and less surface solar radiation in China from 1955 to 2000, *Geophys. Res. Lett.*, 33, L01812, doi: 10.1029/2005GL024586, 2006.

Ramanathan, V., The Greenhouse Theory of Climate Change: A Test by an Inadvertent Global Experiment, *Science*, 240, 293-299, 1998.

Ramanathan, V., *et al.*, The Indian Ocean Experiment: An integrated analysis of the climate forcing and effects of the Great Indo-Asian Haze, *J. Geophys. Res.*, 106, 28371-28398, 2001a.

Ramanathan, V., Crutzen, P.J., Kiehl, J.T. and Rosenfeld, D., Aerosols, climate, and the hydrological cycle, *Science*, 294, 2119-2124, 2001b.

Ramanathan, V. and M.V. Ramana, Atmospheric Brown Clouds: Long Range Transport and Climate Impacts, *EM*, December, 28-33, 2003.

Ramanathan, V., C. Chung, D. Kim, T. Bettge, L. Buja, J.T. Kiehl, W.M. Washington, Q. Fu, D. R. Sikka and M. Wild, Atmospheric brown clouds: Impacts on South Asian climate and hydrological cycle, *Proceedings of the National Academy of Sciences of the USA*, 10.1073/pnas.0500656102, 2005.

Roberts, G, G. Mauger, O. Hadley, V. Ramanathan, North American and Asian aerosols over the Eastern Pacific Ocean and their role in regulating cloud condensation nuclei, *J. Geophys Res.*, In Press, 2006.

Rosenfeld, D., Suppression of Rain and Snow by Urban and Industrial Air Pollution, *Science*, 287, 1793-1796, 2000.

Rotstayn, L.D. and U. Lohmann, Tropical Rainfall Trends and the Indirect Aerosol Effect, *J. Climate*, 15, 2103-2116, 2002.

Satheesh, S.K., and V. Ramanathan, Large differences in tropical aerosol forcing at the top of the atmosphere and Earth's surface, *Nature*, 405, 60-63, 2001.

Seinfeld, J.H. and S. Pandis, *Atmospheric Chemistry and Physics*, John Wiley and Sons, pp. 1309, 1998.

Smith, K.R., National burden of disease in India from indoor air pollution, *Proceedings of the National Academy of Sciences* 97(24):13286-93, 2000.

Smith, K.R., In Praise of Petroleum, Editorial, *Science*, Vol. 298, Dec. 6, pp. 1847, 2002.

Stanhill, G., Cohen, S. Global dimming: a review of the evidence for a widespread and significant reduction in global radiation with discussion of its probable causes and possible agricultural consequences, *Agric. For. Meteorol.*, 107:255-278, 2001.

UNEP and C4, *The Asian Brown Cloud: Climate and Other Environmental Impacts*, UNEP, Nairobi, 2002.

Webster, P.J., V.O. Magana, T.N. Palmer, J. Shukla, R.A. Tomas, M. Yanai and T. Yasunari, Monsoons: Processes, predictability, and the prospects for prediction, *J. Geophys. Res.*, 103, 14451-14510, 1998.

Wild, M., *et al.*, From dimming to brightening: decadal changes in solar radiation at Earth's surface, *Science*, 308, 847-850, 2005.

GLOBAL CHANGES IN AQUATIC SYSTEMS AND THEIR INTERRELATIONS WITH HUMAN HEALTH

MICHEL H. MEYBECK

Aquatic systems are commonly considered to be essential determinants of human health, for example through multiple water-related diseases that affect the lives of some hundred million people or through their impact on food supplies, water-related safety, personal hygiene, etc. (Unesco, 1992; McMichael *et al.*, 1996; Gleick, 1993; McMichael, 2003). Continental aquatic systems (CAS) include streams, rivers, wetlands, lakes, ground waters, coastal wetlands and estuaries in their broader sense (e.g., deltas, tidal estuaries). These systems are regarded by economists and water sector industrialists as essential water sources for most human needs and activities. For Earth System scientists, CAS are also looked at in terms of fluxes, reservoirs and cycles of materials, such as water, carbon, nutrients, ions, metals, sediments, shaping the surface of continents and feeding the coastal zone (Garrels *et al.*, 1973; Berner and Berner, 1987; Steffen *et al.*, 2004). These cycles are controlled by processes such as water balance, atmospheric inputs, soil leaching and erosion, chemical weathering, biological uptake, flow routing, food web cycling, and particulates retention (Berner and Berner, 1987; Schlesinger, 1997; Meybeck and Vörösmarty, 2005).

Continental aquatic systems have gradually changed since the very early development of humanity, with the first irrigated fields and their related dams and reservoirs, and with the first agricultural drainage. In addition to water use in agriculture, most other human activities use continental waters and in return are impacting them at the local to regional scale in quantitative ways (e.g., water balance, river discharge) or in qualitative ways (e.g., pollution, habitat change, aquatic biota modification). Pristine CAS are now seldom found, as less than 17% of the present-day continental surface can be considered to be without a direct human foot-

print (Sanderson *et al.*, 2002). Most of this pristine land is actually found in desert regions, where CAS are actually very limited, as in the Sahara, Central Australia, Central Asia and the Kalahari. Pristine river basins are limited to some boreal regions of Siberia and North America, and to the Amazon and Congo basins.

These local to regional impacts are now combined with global environmental changes, such as climate change and sea level rise, which both occur at the global level and which have begun to influence all aquatic systems on land (Kabat *et al.*, 2004) and those at the land/sea interface (Crossland *et al.*, 2005). Although the present extent of climate change impact on river basins is difficult to establish and to differentiate from direct human impacts such as damming (Dynesius and Nilsson, 1994) or irrigation (Gleick *et al.*, 2001), most scientists are convinced that the next 20 to 50 years will see major changes of the water balance at local to global scales due to climate change, and that sea level will gradually rise (Kabat *et al.*, 2004; Steffen *et al.*, 2004).

In addition to global environmental changes, the last one or two hundred years have also been characterized by accelerated human changes across the planet such as population development, economic changes, technical innovations, and social and political changes (Steffen *et al.*, 2004). In the human health sector, other major global changes should also be considered, such as human behavior, health care, public infrastructure, global circulation of humans, animals, food, and their related diseases. All these changes are now considered as part of global change, which has been accelerating over the last fifty years. Recently Paul Crutzen (2002) defined this present-day Earth System as the Anthropocene, the new geologic era following the Holocene (the last 10,000 years), in which human control on Earth Systems dynamics, particularly climate, is now equal to or exceeds the natural forcing, e.g., solar radiation and the internal heat of the Earth.

Many aspects of climate change and health have been extensively treated within the IPCC and by UN agencies such as WHO, WMO and UNEP (Unesco, 1992; McMichael *et al.*, 1996; McMichael, 2003), and by Gleick *et al.* (2001) and will be summed up in Section 6 of this paper after a short presentation (Sections 1-3) of health issues related to aquatic systems. The core of this paper is devoted to the direct influence of humans on aquatic systems, which is far more important and which proceeds faster than climate change impacts (Meybeck and Vörösmarty, 2005; Vörösmarty and Meybeck, 2004). The quality of water and of aquatic sys-

tems is specifically addressed (Sections 4 to 6) although its global assessment is difficult due to lack of relevant data. Finally, in Section 7, I am proposing some possible scenarios for the future evolution of aquatic systems in the Anthropocene, particularly for water quality, in relation to various human responses to changes.

1. Health Issues in Natural Aquatic Systems

Under natural conditions, aquatic systems do not always facilitate human development and good health. Three major types of health issues are identified here (Table 1, Figure 1): (i) problems related to chemical composition of water resources in natural conditions; (ii) problems related to the occurrence of illness vectors such as insects, snails, bacteria, viruses and other microorganisms; (iii) problems related to drought and flood risks, which will be addressed in the next section together with droughts and floods generated by climate change.

The natural water chemistry of continental waters can be very variable (Meybeck, 1998, 2003a). The total dissolved solids (TDS sum of major ions) may range from 0.1 to 10 g/L in streams and rivers and reach up to 400 g/L in saline lakes such as the Dead Sea. However, most rivers and open lakes have a TDS content much less than 3 g/L, which is fit for human consumption. Exceptions are noted for springs and small river basins (Figure 1, b) underlain by rare rocks types such as pyritic shales, gypsum and rock salt; for these waters the dominating Ca^{2+} and HCO_3^- ions may be replaced by Na^+-Cl^- or Mg^{2+}-SO_4^{2-} ionic associations, which are much less appropriate for drinking. In semi-arid and arid regions, the surface waters are gradually evaporated (Figure 1, c). This results in an increase of TDS and enrichment of Na^+, Mg^{2+}, Cl^-, and SO_4^{2-} which sometimes exceeds the WHO water quality criteria. Some extreme water bodies do not allow most water uses, including use as drinking water, and yet have a very high conservation and biodiversity value. Unique waters, with very high dissolved organic carbon (DOC) contents (peat bogs), very low pH (pH 1 for Lake Kawah Idjen, Indonesia), or very high pH (pH 12 in Lake Bogoria, Kenya) or hypersalinity (Dead Sea, Kara Bogaz), may host very resistant and generally endemic species.

Groundwaters are generally more mineralized than surface waters, consequently water quality criteria are more often exceeded. In some groundwaters, fluoride or arsenic-containing rocks may release these ele-

ments when pumped at the surface. These water quality issues are found at the regional level and may affect hundred of thousands to a million people as in Tanzania, Senegal, and Rajasthan for fluorosis, and in Chile and Bangladesh for arsenic poisoning (Chilton, 1989).

In regions far from the inputs of marine aerosols that naturally provide iodine, deficiencies of this element in water may put hundreds of millions of people at risk of goiter, as in Central China (Meybeck *et al.*, 1989). In contrast, the marine intrusion of sea salt into coastal aquifers, particularly in deltas, is a natural limitation of most water uses, including drinking (Figure 1, d).

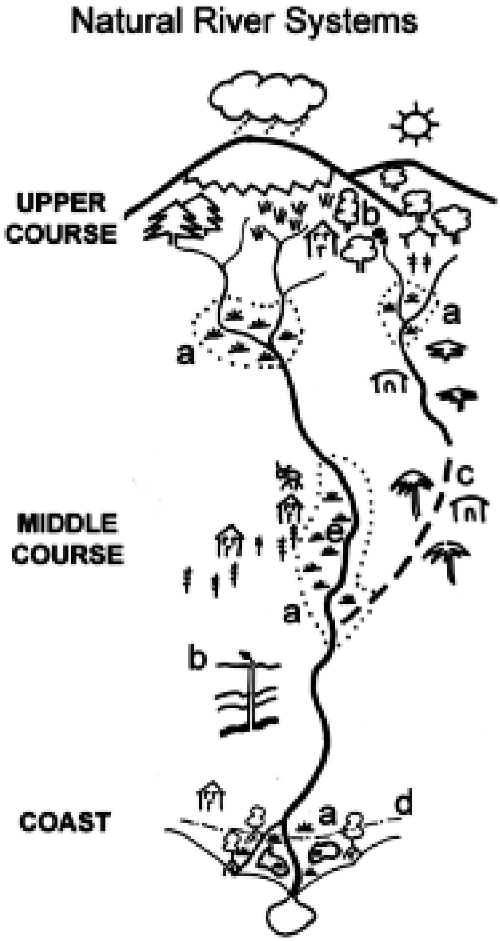


Figure 1. Schematic landscape view of continental aquatic systems and related health issues in natural conditions (# a to X, see Table 1).

TABLE 1. HEALTH IMPACTS INDUCED BY AQUATIC SYSTEMS IN NATURAL CONDITIONS (Coding refers to Figure 1 and Table 1).

Aquatic System (Coding)	Description	Vectors and/or issues	Health impact and Population affected[(1)]
Wetlands and lakes occurrence (a)	lowland humid regions	• Mosquito occurrence	Filariasis: 117 million infected per year
			Yellow fever: epidemics in tropics
			Malaria: 2.4 billion at risk
			Dengue: 50 million infected per year
		• Snail occurrence	Schistosomiasis: 200 million
		• Cyclops (crustacean) occurrence	Dracunculiasis: 100 000 infected per year
Running waters	Tropical regions (Africa)	• Simulium (Black flies)	Onchocerciasis: 18 million infected per year
Evaporated surface and groundwaters (c)	arid regions	• saline waters (Na^+, Cl^-, SO_4^{2-}, Mg^{2+})	dozens of millions worldwide
Specific surface and groundwaters (b)	occurrence of F^-, Arsenic	• fluorosis (Senegal, Tanzania, Radjasthan)	
	high levels of dissolved salts	• saline waters (arid belt)	
Inland waters	lack of iodine	• iodine deficiency (China)	Goitre: millions in China
All waters	tropical regions mostly	• infective bacterial agents (*Salmonella, Vibrio cholerae, Leptospira*)	
		• Viral agent (hepatitis A, poliomyelitis)	
		• Parasites (*Amoeba, Giardia*)	
		• Enteric diseases (diarrhoeas)	
Coastal groundwaters (d)	marine intrusion (salt wedge)	• Saline waters	

(1) 1990s statistics (Unesco, 1992 with update according to McMichael *et al.*, 1996)

Most vectors of water-related diseases are found in humid tropical regions and, less often in temperate regions (McMichael *et al.*, 1996). Mosquitoes which are the vectors of filiariasis, malaria, yellow fever and dengue (Table 1) can be found in all types of wetlands, either located in the upper course of a river basin, the middle course or in the coastal zone (Figure 1, a). Freshwater snails that host schistosomiasis-transmitting trematodes are commonly found in tropical ponds and irrigated areas. Black flies that propagate onchocerciasis were common in West African middle sized rivers until a major WHO sponsored program limited their development. Cyclops, a tiny crustacean propagating dracunculiasis, is common in Africa. There is also now growing evidence that cholera outbreaks are linked to the development of multiple *Vibrio cholera* hosts in some aquatic systems such as blue-green algae and coastal copepod zooplankton (see next section on meteorological extremes). Further information can be found in the reports by Wilson (2001) and by McMichael (2003).

2. Health Issues Induced by Climate Change and Sea Level Rise

Climate change and health issues have been extensively reviewed in the last decade particularly within the Intergovernmental Panel on Climate Change (IPCC) and at WHO (McMichael *et al.*, 1996; McMichael, 2003). I focus here more on (i) the gradual changes of the water balance and its related impact on land cover and (ii) on extreme meteorological and hydrological impacts.

The longest records for river discharge reach back 100 years. While direct impacts from water use are evident for this period (see next sections) (Vörösmarty and Meybeck, 2004), it is difficult to draw a global picture of hydrological changes due to climate change. For instance, the analysis of very diverse rivers such as the Athabaska (N.E. Canada), N. Dvina (European Russia), Lena (Siberia), Niagara (N. America), Parana (S. America), Congo and Amazon does not show definite trends for the 1900-1990 period. In dry and semi-arid regions, cyclic natural variations are observed, as for the Niger River discharge (W. Africa) (Laraque *et al.*, 2001), Lake Chad levels (Central Africa) (Lemoalle, 2004), and Moroccan Rivers discharging SE to the Sahara. The Central Asia regions from the Gobi desert to the Caspian also show very variable runoff with relatively limited climate variability over the last hundred years: the Gobi was drained by the Kerulen River in Mongolia, which was connected to the Amur River basin

in the 1900s, and the Amu Darya river had been connecting the Aral and Caspian Seas through the Uzboi channel some 3000 years ago (Aladin *et al.*, 2004). Both river basins had very different and more humid waterscapes from what was observed in the 20th century.

The expected hydrological changes linked to climate change and their impact on human health and on Earth System functions (e.g., carbon balance, fluvial morphology, and aquatic biodiversity) are presented in Table 2 and in Figure 2, using a river basin structure identical to the one of

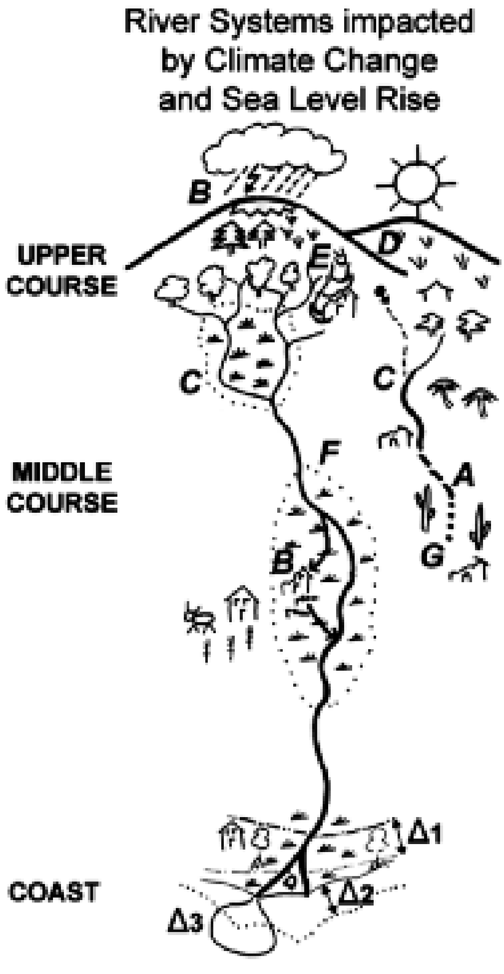


Figure 2. Schematic landscape view of continental aquatic systems and related health issues with Climate Change (# A to G) and Sea Level Rise (Δ1 and Δ2) impacts (see Table 2).

TABLE 2. MAJOR HYDROLOGICAL CHANGES DUE TO CLIMATE CHANGE AND SEA LEVEL RISE, AND THEIR RELATED ISSUES

Environmental changes	Local to regional changes	Global Impacts						
		A	B	C	D	E	F	G
Climate variability and Climate Change	A Development of non-perennial rivers	•	•	•	•	•	•	•
	B Development of extreme flow events	•	•			•	•	•
	C Changes in wetland distribution	•	•	•	•		•	•
	D Changes in chemical weathering				•			•
	E Changes in soil erosion	•			•	•		•
	F Changes in flow regimes	•	•			•	•	•
	G_1 Salinization through evaporation	•	•	•			•	
Sea Level Rise	Δ_1 Salt water intrusion	•		•			•	
	Δ_2 Coastal erosion					•	•	

A: human health, B: water availability, C: water quality, D: carbon balance, E: fluvial morphology, F: aquatic biodiversity, G: coastal zone impact. Only the major links between issues and impacts are listed here (Meybeck *et al.*, 2004; adapted from Meybeck, 1998) (Codes refers to Figure 2).

Figure 1. Increased droughts will lead from seasonal to permanent river dryness – river and lake desiccation – (A, Figure 2, Table 2), which affects all river functions within the Earth System and leads to severe health issues. These aspects are developed further with other extreme events such as floods. The major impact of climate change is probably the shift in wetland distribution (C). Some dry regions will be exposed to more humid climate and new wetlands will develop, while in other regions they will decrease (see Steffen and Lambin, 2006).

In subarctic regions, permafrost melting due to global warming will leave millions of hectares of new wetlands, although their direct impact on human health may be more limited than the occurrence of new wetlands in the tropics. Changes in chemical weathering (D) and in soil erosion (E) due to warming, land cover change and water runoff change will impact the Earth System functions more than human health except for the development of extreme storm events that increase landslide occurrence, particularly in coastal regions. The gradual evaporation resulting from drier climates will lead to the extension of regional salinization (G1).

The specific issues in the coastal zone include the intrusion of sea salt into aquifers and coastal erosion, both related to sea level rise (Figure 2, Table 2).

In addition to these permanent changes, climate extremes will generate specific hydrological events that have an extreme impact on societies, particularly on human health (Hales *et al.*, 2003) (Table 3). They can be expressed at various temporal scales from very short (hourly and daily rainfall) to seasonal and decadal. During El Niño years, the hydrological balance of the Earth System is very much affected from the local to the global scale but changes of climate extremes may also be related to local and regional events as hurricanes and typhoons (Kabat *et al.*, 2004). Many studies have found some correlations with extreme events generated by the El Niño Southern Oscillations (ENSO events) or by the Southern Oscillations Index (SOI) with the expansion of malaria and dengue epidemics, however in most cases these correlations are very local and not yet fully explained (Hales *et al.*, 2003).

During the extreme events, surface hydrology may be greatly modified and social and economic infrastructures can be completely altered leading to catastrophic events, essentially defined by their socio-economic impacts: the same hydrological extreme may be mitigated very differently in different locations, as is regularly observed for Caribbean Hurricanes. WHO has defined a catastrophic flood or drought as an event that is (i) affecting more than 200 people, or (ii) killing more than 10 people, or (iii) requires assistance from a central or provincial government.

Both floods and droughts can be associated with the development of enteric diseases. In the tropics, diarrhoeal diseases typically peak during the rainy season. Extreme floods could help propagate the pathogens, while drought conditions lead to an increase of hygiene-related diseases (Hales *et al.*, 2003).

During heavy rainfall, surface water supplies and even some karstic-groundwater supplies can be more turbid, even after treatment to meet drinking water quality, and contain clay-sized particles with cryptosporidium, giardia, shigella, typhoid, and viruses that cause diarrhoea. *One important and long-term impact of extreme events is the destruction of sanitation infrastructures during floods or landslides, and the contamination or destruction of drinking water systems and subsequent fecal contamination.*

The fine scale mapping of the climate change impacts on human health, such as the geographic distribution of malaria, schistosomiasis, dengue and other vector-borne diseases, will be very difficult due to the multiple factors that have to be predicted at this scale, such as temperature, seasonal pattern of rainfall, and occurrence of new wetlands, which are still very much debated among modelers (Martens *et al.*, 1995). For instance, there is growing evidence of correlations between heavy rainfall and inundations during El Niño years and cholera outbreaks, as in the Ganga-Brahmaputra delta and in the

Amazon floodplain. However, the links between the multiple controlling factors such as river pH and phytoplankton blooms are not yet fully understood (Pascual *et al.*, 2002; Colwell, 1996; McMichael *et al.*, 2003). Another major difficulty in assessing the impact of climate change is taking into account the current dynamics of aquatic systems exposed to direct human impacts.

TABLE 3. HEALTH IMPACTS INDUCED BY AQUATIC SYSTEMS MODIFICATIONS DURING EXTREME EVENTS (modified from Kovats R., 1999; and Hales *et al.*, 2003; McMichael *et al.*, 1996; McMichael, 2003) (Coding refers to figure 2 and table 2).

EVENT (Coding)	TYPE	DESCRIPTION	POTENTIAL HEALTH IMPACT
Heavy precipitation (B)	Meteorological	Extreme rain	Increased leaching of soil microorganism (*Cryptosporidium, Giardia*) (W_B)
	Geomorphological	Wetlands formation	Increased water-related vectors (mosquitoes…) (W_B)
		Landslide; mud slide	Break of drinking water and sewage collection systems; pathogens contaminants (W_B)
Flood (B, E)	Hydrological	Over bank flooding; temporal wetlands	Water-related vectors (mosquito abundance; cholera hosts) (V_B)
	Social	Property damage	Contamination of water supply with faecal matter and rat urine (W_B)
	'Catastrophic event'		multiple drowning; respiratory infection; diarrhoeal disease; population displacements (W_B)(1)
			Crop losses and famine (PP)
Drought (B, C)	Meteorological	Riverbed dried up	Development of some disease vectors (V_B)
	Social	Reduction of water supply; reduction of sewage dilution	Water quality degradation (W_B)
	'Catastrophic event'		Water-washed diseases spreading (PP)(2)
			Population displacement; crop losses and famine (PP)

WB: water-borne diseases; VB: vector-borne diseases; PP: Person to person diseases; (1) e.g. outbreaks of hepatitis A, leptospirosis, typhoid (McMichael *et al.*, 1996); (2) e.g. outbreaks of scabies, conjunctivitis.

3. Health Issues Induced by Direct Human Impacts

Direct human impacts on aquatic systems have been, so far, much more important and faster than the gradual climate change impacts we are experiencing now. In a few millennia, the land cover change due to agriculture and global human settlement has reached about 80% of the Earth's surface (Steffen and Lambin, 2006). In the last 50 to 100 years, the river hydrological network has been completely fragmented and regulated by dams, dikes and levels, reservoirs, water diversion and irrigation practices (Dynesius and Nilsson, 1994; Gleick, 1993; Gleick *et al.*, 2001; Vörösmarty and Meybeck, 2004). The changes generated by water uses have now reached a level similar to those induced by slow climate variations that occurred over the last 20,000 years since the Last Glacial Maximum (Meybeck and Vörösmarty, 2005).

Most human activities, for example mining, smelting, industries, urbanization, and intensive agriculture, have generated an enormous amount of wastes, which are dumped, leached or eroded into aquatic systems and slowly carried by river networks to the coastal ocean. Other human activities such as transportation and hydropower generate a profound modification of river course morphology and aquatic habitat. These processes can be regarded as an acceleration of transfers at the Earth's surface for organic carbon, nutrients, metals, sediments, and some hydrocarbons. Water quality surveys also reveal the occurrence in aquatic systems of new materials that do not exist in natural conditions, such as pesticides, polychlorinated-biphenyls, solvents, and drugs, which are termed *xenobiotics* and are harmful for animals and humans.

These impacts are generally made at the local to sub-regional scales, but they are occurring now on all continents and can be regarded as a global scale issue (Cole *et al.*, 1993; Seitzinger *et al.*, 2002; Meybeck, 2003b; Steffen *et al.*, 2004; Vörösmarty and Meybeck, 2004). They are presented in Figure 3, using the same schematic river network, from headwater to coast, as in Figures 1 and 2. The alteration of Earth System functions together with the health issues in association to these human pressures are presented in Tables 4 and 5.

Most human pressures on aquatic systems have some potential health impacts, and produce a combination of several types of Earth System alteration. This results in a growing complexity of the interrelation between humans and aquatic systems. It is important to note that some physical alterations of natural systems are targeted to facilitate human settlements

and agriculture, and/or to safeguard crops and properties. They can also lead to important Earth System dysfunctioning, which in turn may have an impact on human health, for instance through the modification of aquatic habitats and their related biodiversity. Intensive irrigation and reservoir flooding are often associated with new wetlands that favor disease vectors such as mosquitoes and freshwater snails. In semi-arid regions, irrigation may result in a marked increase of the dissolved salt content in irrigation returns and in groundwaters that can exceed WHO criteria for drinking water. Groundwater pumping in Bengal and Bangladesh has modified the chemical equilibrium of arsenic species that are naturally present in this aquifer leading to massive As poisoning in this region.

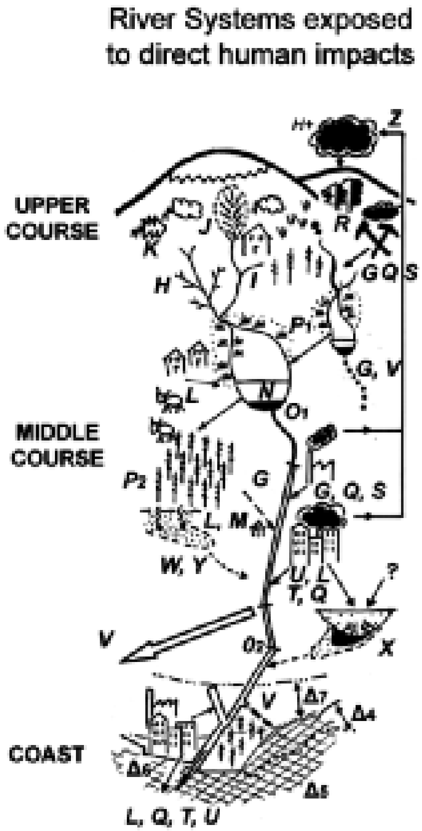


Figure 3. Schematic landscape view of continental aquatic systems and related health issues and degradation of Earth Systems functions under direct Human pressures (# H to Z, see table 3; Δ3 to Δ7, see Table 5).

TABLE 4. MAJOR LOCAL AND REGIONAL PRESSURES ON CONTINENTAL AQUATIC SYSTEMS AND RELATED ISSUES. A: human health, B: water availability, C: water quality, D: carbon balance, E: fluvial morphology, F: aquatic biodiversity. Only the major links between issues and impacts are listed here (Meybeck *et al.*, 2004; adapted from Meybeck, 1998) (Codes refers to Figure 1, right).

Pressures	Local to regional changes of environmental states	Global Impacts					
		A	B	C	D	E	F
Land use change	H Wetland filling or draining	•	•	•	•		•
	I Changes in water pathways and water balance		•	•			
	J Change in sediment transport				•	•	
	G Salinization	•		•			•
	K Alteration of first order streams					•	•
	L Nitrate and phosphate inputs	•		•	•		
	M Pesticide occurence	•		•			
	W Groundwater contamination (NO_3^-, pesticide)	•		•			
River damming and channelisation	N Nutrient, carbon and particulates retention	•			•	•	
	O Loss of longitudinal (O_1) and lateral (O_2) connectivity						•
	P Creation of new wetlands	•		•	•		•
Industrialisation and mining	Q Inputs of heavy metals and POPs	•		•			
	R Acidification of surface waters	•		•			•
	G Salinization	•		•			•
	S Inputs of particulates			•		•	•
Urban wastes	L Nitrate and phosphate inputs	•		•	•		•
	T Faecal contamination	•		•			
	U Carbonaceous	•		•			•
	Q Inputs of heavy metals and POPs	•		•			
Irrigation/water transfer	V Partial to complete decrease of river inputs	•	•		•	•	•
	G Salinization (evaporation and percolation)	•	•	•			
Groundwater changes	W Infiltration of contaminated waters	•		•			
	X Wastes dumps leakage	•		•			
	Y Overpumping groundwaters	•	•	•			
Atmospheric pollution	Z Emission of atmospheric pollutants	•		•			

The coastal zone (Table 5) is very sensitive to Global Change (Crossland *et al.*, 2005). The effects of this zone encompass (i) upper river course and middle course impacts through river flow modification, (ii) direct specific impacts such as groundwater pumping and oil extraction in coastal alluvial aquifers (Δ7, Figure 3), coastline artificialisation (Δ6) (e.g., dredging navigation channels; digging canals in deltas that modifies the water dynamics) and (iii) sea level rise impacts (Δ1 and Δ2, Figure 2 and Table 4). Wetland

TABLE 5. MAJOR LOCAL PRESSURES AND GLOBAL CHANGES IN COASTAL SYSTEMS AND RELATED ISSUES. A: human health, B: water availability, C: water quality, D: carbon balance, E: fluvial morphology, F: aquatic biodiversity. Only the major links between issues and impacts are listed here (Issues linked to Sea Level Rise: see Table 2).

Pressures	Local to regional changes of environmental states	Global Impacts					
		A	B	C	D	E	F
Coastal Issues	V Partial to complete decrease of river inputs		•		•		•
	_3 Enhanced sediment input				•	•	•
	_4 Sediment starving and erosion					•	•
	_5 Coastal eutrophy, Harmful algal blooms	•			•		•
	_6 Coastline artificialisation and wetland filling				•	•	•
	_7 groundwater overpumping, oil extraction, salt intrusion, subsidence		•	•			
	L, Q, T, U Coastal contamination	•		•	•		•

drainage and filling (Δ6) is another important pressure in coastal regions, sometimes carried out for mosquito control, but most of the time for agricultural development and urbanization.

Where nutrient inputs to the coast increase either through direct release from cities and agriculture, or through river inputs (L), coastal eutrophication may occur, resulting for some deltas in the destruction of the oxygen balance and severe modification of the food-web (dystrophy) (Rabalais and Turner, 2001) or in development of harmful algal blooms that can be associated with high levels of toxins in filtering mollusks (oysters, clams, mussels) (Chorus and Bartram, 1999; Anderson *et al.*, 2002). After a drought in February, 1996, all 126 patients in a haemodialysis unit in Caruaru, north-east Brazil, developed signs and symptoms of acute neurotoxicity and subacute hepatotoxicity following the use of water from a lake with massive growth of cyanobacteria (blue-green algae) (Pouria *et al.*, 1998). Other specific impacts on the coastal zone concern the sediment inputs that can be markedly increased in smaller river basins (Δ3) after land use change such as deforestation and farming, particularly in the humid tropics, which can 'blanket' the coral reefs with fine mud (Syvitski *et al.*, 2005). Conversely, the damming and reservoir construction that has been exponentially increasing at the global scale since the 1900s, and the water diversions that are common in some regions, result in a decrease of all river inputs (V) to the coasts particularly for water and for sediments (Δ3), i.e., the sediment 'starving' of the coast. These impacts are modifying

the Earth System balance (e.g., inverse sediment balance in some deltas with the dominance of coastal erosion), yet their relation with human health requires more study, particularly in the long term (50-100 yrs). In addition to these quantitative human impacts, from headwaters to estuaries, on the continental aquatic systems, the water quality is impacted by anthropogenic sources of contaminants (see Column C, Table 4).

The transfer of these contaminants in aquatic systems is not straightforward and depends on

(i) the water residence time in the different water bodies of the river basin;

(ii) the reactivity of contaminants in these water bodies;

(iii) the trapping of particulates in the system.

The transit time of surface waters in big rivers ranges from a few weeks to nearly a year for the longest ones. River aging due to reservoir construction can increase these figures by an order of magnitude (Vörösmarty *et al.*, 1997; Vörösmarty and Meybeck, 2004). In addition, it must be considered that surficial aquifers which contribute to river base flow during dry periods have a much larger residence time, from years to decades as is the case for many large lakes. Once these water bodies are impacted, their restoration will take 2 to 3 times longer than the residence time, due to multiple environmental inertia in soils, sediments and to their non-piston flow renewal.

The reactivity of water borne material and the trapping of particulates within aquatic systems is also very variable. Fluvial filters (Meybeck and Vörösmarty, 2004), which include mountain slopes and piedmonts, headwater wetlands, floodplains, lakes, and estuarine systems control the fluxes of particulates and their attached contaminants, nutrients and pathogens, as well as many fluxes of dissolved and/or reactive nutrients and contaminants. Each river system can be described by its specific assemblage of fluvial filters, which is now changing fast as a result of water engineering and land use change. *It is estimated that more than 90% of suspended matter derived from erosion is naturally retained in large systems, and up to 99.9% is retained by large reservoirs, while only 18% of total nitrogen inputs to river basins to river basins (natural and anthropogenic) are exported to oceans* (Vörösmarty *et al.*, 2003; Green *et al.*, 2004).

Determining the direct impact of human activities on aquatic systems with regard to health issues is complex and involves water quality issues, the positive impacts of flood and drought regulation, settling and processing of particulate contaminants, and attached pathogens in reservoirs. Although the global picture shows an overall degradation of water quality from natural conditions, there are striking differences in time and space for each type of issue as presented in the next sections.

4. Defining and Assessing Water Quality at the Global Scale

The perception of water quality through its color, turbidity, taste, or effects on man and animals is as old as water use: water quality management rules existed in most ancient hydraulic civilizations from Mesopotamia to Egypt. The first chemical analyses of water were performed following the development of analytical chemistry some 200 years ago. Since that period, water quality perception and definition has constantly evolved with societal development. Water quality is a rapidly evolving field with multiple metering approaches (Meybeck, 2005). Unlike many other Global Change impacts, it is very site specific.

4.1. *Water Quality: a Fast Evolving Field*

The first major water quality surveys were performed on the Thames and Seine rivers following cholera outbreaks in the mid 1800s with only a few descriptors such as resistivity, dissolved oxygen, ammonia, chloride and fecal contamination indicators (fecal coliforms). Throughout the 20th century, water quality studies and monitoring grew exponentially in step with water demand, the occurrence of problems (eutrophication since the 1960s, acidification in the 1970s, endocrine disruptors more recently, radionuclides since the 1950s, pesticides since the 1980s) and the development of analytical chemistry.

'Water quality', initially defined by sanitary engineers and hydrologists using a few chemical descriptors in one sample or at one station, has now shifted to an overall appreciation of the 'aquatic environment quality' based on chemical, physical, and biological descriptors (Chapman, 1996). Water quality monitoring is getting very complex (Chapman, 1996; Mc Cutcheon *et al.*, 1993) and the total number of potential water quality descriptors probably now exceeds several hundred, while in the 1900s they were just one or two dozen (Figure 4, Trajectory A) (Meybeck, 2005). However, due to financial and technical constraints, the best-equipped monitoring stations consider routinely one hundred descriptors at best (Trajectory B), while in the Least Developed Countries, monitoring stations, when they exist, can still barely measure a dozen descriptors (Trajectory C).

In contrast to the situation in atmospheric chemistry, the aquatic environment cannot be simply described by one or two emblematic descriptors such as CO_2, which is continuously measured at the Mauna Loa observatory in Hawaii. The increase in global CO_2 measured at this station triggered

the climate change concern in the 1960s (see Steffen *et al.*, 2004). Global warming can be also tracked on the basis of one simple indicator, average air temperature, commonly measured for 200 years and now widespread at tens of thousands of meteorological stations. Sea level rise is also based on one indicator. Yet evaluating water quality involves dozens of descriptors and their evolution is station specific. Vörösmarty (2002) observed that hydrologists were lacking a 'Mauna Loa'-like curve as a reference for global water balance. It is even worse for water quality, for which no 'Mauna Loa' curves can be established.

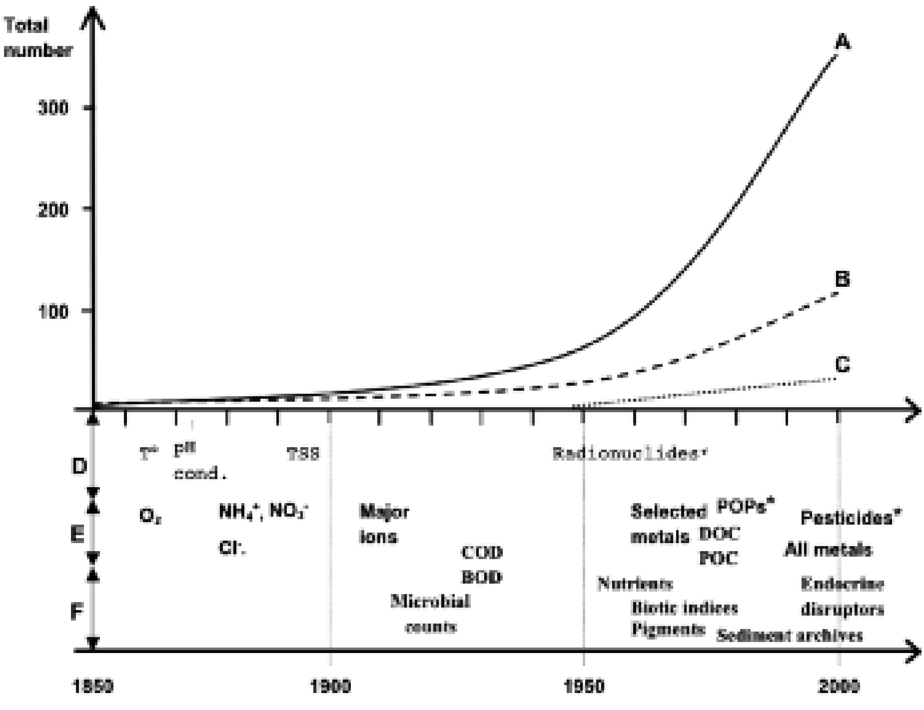


Figure 4. Exponential growth of water quality descriptors since 1850 and occurrence of their first analysis in regular surveys. Trajectory A = total maximum number of variables that should be considered if all regulations were implemented. B = number of variables actually routinely monitored in the first grade surveys. C = monitoring capacities of Least Developed regions. D = Physical descriptiors. E = Chemical descriptors. F = Biological and Ecological descriptors (*: non natural products) (adapted from Meybeck, 2004).

4.2. *Assessing Water Quality is Complex*

Each water user and each hydroscientist is interested in different aspects of water quality; these are very rarely collected and synthesized (Meybeck, 2005) and may lead to multiple definitions of water quality (Boon and Howell, 1997). Two main streams of water quality assessment co-exist: one referring to a *hypothetical natural background*, which is a common vision among Earth system scientists, and the other referring to *potential water uses*, which is the vision of all water users; particularly for drinking.

Establishing water quality metrics for users is an important part of water management. The drinking water criteria as established by WHO are probably the only universal criteria accepted by countries, which generally transcribe these criteria into legal and regulatory thresholds such as for chloride, nitrate, lead or arsenic. These criteria are not the only ones. In most cases, the water quality metrics used in management result from political decision balancing: (i) socio-economic activities responsible for pressures; (ii) socio-economic activities impacted by water-quality degradation; (iii) perception of water-related issues by societies through the media; (iv) dissemination of technical and scientific knowledge. *Multiple water quality metrics are needed and must be agreed upon by stakeholders and regularly revised, particularly when sharing water bodies.* These scales may evolve: even the most widely used reference, the WHO drinking water standard, is periodically revised on the basis of new scientific knowledge and, probably, on new levels of risk acceptance.

4.3. *Water Quality Issues Depend on Water Bodies*

The occurrence and extent of major water quality issues depend on the nature of the water bodies (Meybeck *et al.*, 1989; Chapman, 1996). These issues are listed in Table 6 and their location in continental aquatic systems is schematically indicated on Figure 3. Their relevance to direct health impacts is evaluated at three levels. Pathogens and vector-borne diseases are associated with communicable diseases (C, Table 6). However, most water quality issues correspond to non-communicable diseases (NC). Some issues such as eutrophication and suspended solids occurrence only occur in surface waters, others are primarily observed in groundwaters, such as salinization and high nitrate levels. The occurrence of fecal pathogens is observed more often in running waters than in other water bodies. The range of water residence time in lakes, reservoirs and groundwater is from weeks to hundreds of years while in streams and small rivers, it is days, thus permitting a higher spread of fecal pathogens.

TABLE 6. WATER QUALITY ISSUES AND HUMAN HEALTH

Issue		Health relevance		Water body			
	Coding[1]	Importance[2]		Rivers	Lakes	Reservoirs	Groundwaters
Faecal pathogens	T	+++	C	♦♦♦	♦	♦	♦
Suspended solids (as host of pathogens)	S, U	+	C	♦♦	na	♦	na
Decomposable organic matter	U	+	C	♦♦♦	♦	♦♦	♦
Eutrophication	L	+	NC	♦	♦♦	♦♦♦	na
Nitrate	L	+	NC	♦			♦♦♦
Salinisation/salt contents F⁻	G	++	NC	♦	♦	♦	♦♦♦
Trace metallic elements + Arsenic	Q	+++	NC	♦♦	♦♦	♦♦	♦♦
Organic micropollutants	Q	+++	NC	♦♦♦	♦♦	♦♦	♦♦♦
Acidification	R	++	NC	♦	♦♦	♦♦	?

[1] See Figure 4 and Tables 4 and 5; (2): + (low) to +++ (high); Importance: C = communicable, NC = non communicable; Occurrence: ♦ (low) to ♦♦♦ (high); na: non applicable.

4.4. *Water Quality is Site Specific*

River water quality measured at one station is actually a spatial integration of the multiple sources, sinks and controls occurring in the intercepted drainage area (Figures 1, 2 and 3 combined). In Earth System Science, global scale scientists have often only taken into consideration the riverine fluxes of material to the oceans based on a dozen of well-documented major rivers (Meybeck, 1982; Seitzinger *et al.*, 2002; Caraco, 1994; Ludwig *et al.*, 1996). Yet the users' demand for water quality information is of course very different from global geo-chemistry and requires much finer resolution.

Water quality cannot be detected by remote sensing apart from color, temperature, suspended solids and pigments. Therefore we must rely on *spatially discrete information performed at stations*. Usually, stations are located where water is most used but the risks associated with water quality must be assessed everywhere.

Spatial integration and interpolation rules must be applied from stations to reaches, subbasins, basins, and depend on station density. In developed countries, the density of water quality monitoring stations is

similar to that of meteorological stations (circa 1 station for 250 km^2 and 25,000 people in France), but it is between one and two orders of magnitude lower in the least developed countries. Spatial representativity also depends on the mixing state of the water body: a few stations may be adequate for a large lake, while a large aquifer may need hundred of stations. Conversely, the survey frequency should be high for rivers, medium in lakes and reservoirs, but can be low (yearly or less) for groundwaters.

Our appreciation of water quality closely reflects the complex relations between humans and water, at a given place, a given period, and for a given society. It is now based on dozens of indicators. Many of them cannot be afforded by the least developed countries and are still barely documented in some developed countries. In addition, water quality and its trends are often site-specific. The next section will address the diversity of human responses to water quality degradation and to its socio-economic and health impacts.

5. Societal Responses to Water Quality Issues

The response of societies to environmental changes depends on many factors such as the identification of an issue, the recognition of its links with human pressures, the consensus that can be built to define adequate measures, and the availability of financial, technological, technical or regulatory means. The study of water quality issues provides good examples of the combined inertia of societies and of water bodies, which generally extends over decades.

5.1. *Timing of Societal Responses: Example of a Restoration Cycle*

The full restoration and stabilization cycle of a water quality issue presents a good example of societal responses, depending on intensity of human impact, time constants and varying societal conditions (Meybeck, 2003b, 2002). The start of human impact is set at time T_0. Then, the following stages can be distinguished (Figure 5):

(i) *hydrosystem reaction to contamination* (T_0-T_1), depending on system size and contaminant pathways (e.g., dissolved vs. particulate transfer): this process depends on water and particulates residence time in hydrosystems;

(ii) *impact detection* (T_1-T_2) of hydrosystem changes by water users, scientists, specific citizen groups ('sentinels');

(iii) development of *societal awareness* (T_2-T_3): time for the development of general knowledge and understanding of the issue, sometimes delayed by lobbying from various social or economic groups;

(iv) *policy lag* (T_3-T_4): time for authorities or politicians to decide on the appropriate action; such decisions can be reached through environmental awareness of all stakeholders (bottom-up consensus) or obtained and imposed by political decision (top-down);

(v) *financial and technical lags* (T_4-T_5): time to fully implement and enforce the decisions;

(vi) *hydrosystem reaction to restoration and remediation measures* to limit (T_5-T_7) or decrease the environmental and societal impacts (T_{6A}-T_{6B}).

Depending on the timing of impact detection, impact duration and remediation effectiveness, various threshold levels can be reached. If the environmental control is not delayed and is sufficiently effective, a limited

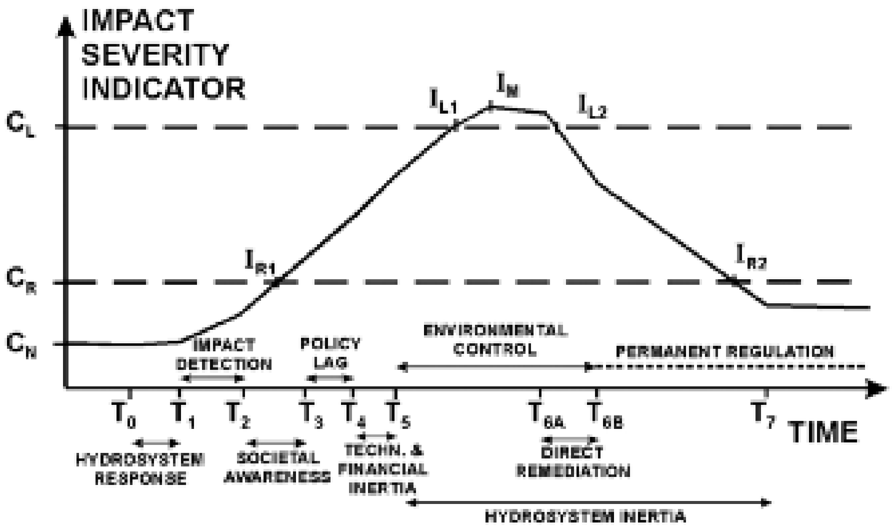


Figure 5. Successful restoration of water quality in an aquatic environment illustrated by a bell-shape trend in water quality. C_N, C_R, C_L: natural, recommended and limit concentrations of water quality indicator. T_0: start of environmental pressure, T_1: first change of water quality, T_2: detection of change, T_3: established societal concern on issue, T_4: political decision concerning the issue, T_5: start of implementation of environmental measures, T_M: time of maximum impact, $T_{6A\text{-}6B}$: direct remediation measures, T_7: new steady state. Time scale (T_0-T_7) varies according to issues and basin sizes (generally expressed in decades) (Meybeck, 2002).

level of maximum degradation is reached, followed by an improvement phase. If not, it can exceed the critical level C_L. In many examples of successful CAS restoration it has been necessary to perform a direct remediation of the aquatic system (dredging the contaminated sediments, inactivation of sediments below a layer of new sediments, direct chemical treatment of water or contaminated soils etc.). *The restoration cycle of a small to mid size catchment (1,000 to 100,000 km²) is generally a few decades*.

5.2. *Recent Trends of Water Quality in Impacted Rivers*

This bell-shaped successful evolution of a water quality issue (Figure 5) is not often actually observed in rivers. Multiple trend patterns are documented (Anderson *et al.*, 1996; Foster and Charlesworth, 1996; Meybeck, 2002). As concentration measurements may not always fully represent the evolution of river systems, fluxes of riverine materials are also often considered as an alternative metric. They are the product of concentration and water discharges. In the great majority of documented cases, flux trends are linked primarily to changes in concentrations, few are linked to river flow changes only. To allow for their inter-comparison and typology, they are here normalized to the beginning of impacts (time T_1), (Figure 6) (Meybeck 2002).

Flux trends are contradictory: many fluxes increase due to rising concentrations (types B, D1, D2, D3, F, H, I, J, K, Figure 6) but some of them actually decrease (types C, E, G) owing to a decrease of water discharge due to water use, water diversion or to the biogeo-chemical and physical retention in an impoundment. Three types of flux decrease can be defined:

(i) the hydrological changes caused by water diversion or use, mostly for irrigation, result in a *gradual decrease* of all water-borne fluxes (type G, Figure 6). This is the case for many impounded basins in arid and semi-arid zones;

(ii) *complete retention* caused by the settling of all particulate matter including attached pathogens in reservoirs (type E), which exceeds 90% when the water residence time exceeds two months and might be responsible for trapping at least 30% of river particulates (Vörösmarty *et al.*, 2003);

(iii) *partial retention* resulting from degradation of organic matter including its pathogens and from the uptake of nutrients (type C) in reservoirs. In eutrophied and/or impounded rivers the Si/N ratio may decrease markedly and cause severe degradation of coastal-zone food webs and development of harmful algal blooms (Turner *et al.*, 2003).

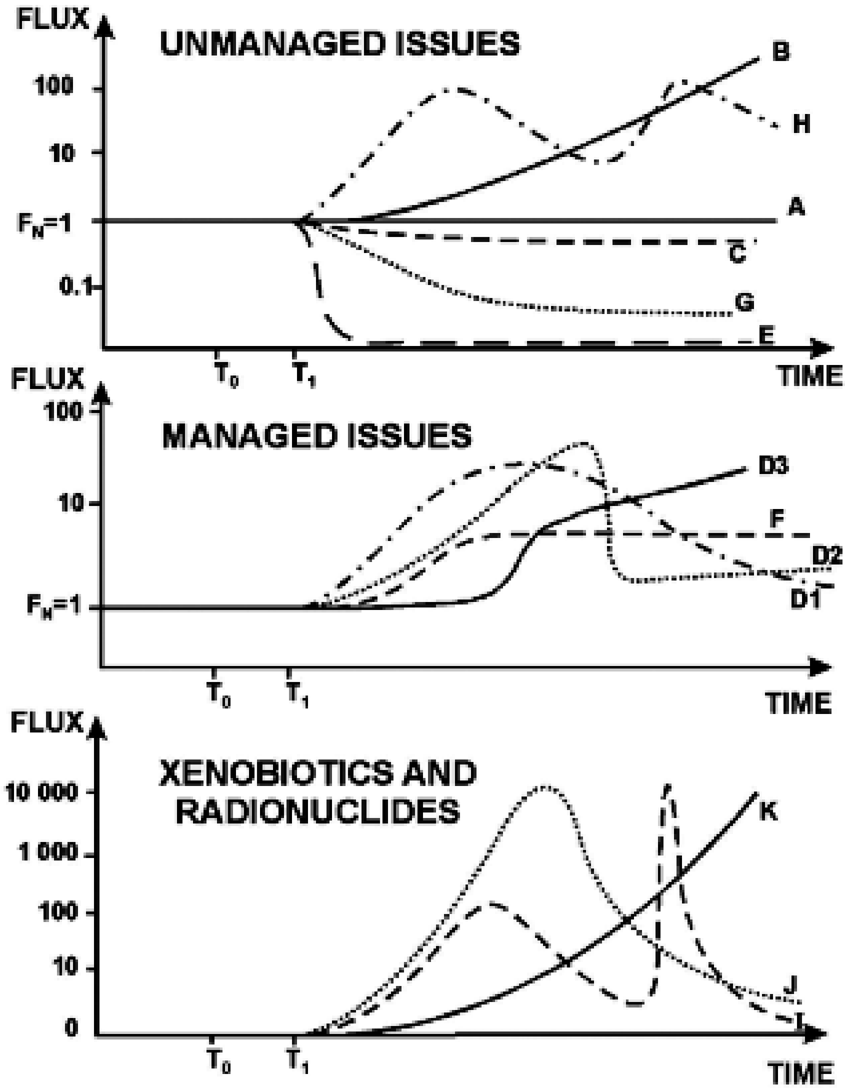


Figure 6. Types of river flux trends normalized to pristine fluxes (F_N) since the beginning of impacts (T_1) related to human pressures (T_0) (Meybeck, 2002).

– Unmanaged issues: A: stable evolution, B: gradual increase, C: partial retention, E: complete retention, G: gradual retention, H: multiple cycles (e.g. BOD_5)

– Managed issues: D1: bell-shaped control, D2: stepwise improvement, D3: stepwise degradation, F: stabilized contamination.

– Xenobiotics products: I: multiple cycles of some radionucleids, J: total ban (as for DDT), K: gradual xenobiotic contamination.

Very few chemical elements are barely affected by human activities and present a stable evolution (type A): most of them are not very related with health such as Ca^{2+}, Mg^{2+}, HCO_3^- or particulate Al, Fe, or Si. The *gradual increase* (type B) of many water-quality indicators corresponds to the development of pressures, for example, Na^+, Cl^-, K^+, NO_3^-, SO_4^{2-}.

Bell-shaped control (type D1) characterizes a successful and gradual control.

Stepwise improvement (D2) is characteristic of a sudden decrease of contaminant concentrations (dissolved and/or particulate) in river systems. These trends are essentially caused by a drastic reduction in contaminant point sources such as that caused by the construction of urban or industrial sewage treatment plants, or by the reduction or closure of economic activities during economic crises.

Stepwise degradation (D3) is the symmetric evolution corresponding to the installation of major industries, collection of urban sewage without subsequent treatment or to sudden change in land use.

Multiple cycles (type H) of contamination/improvement are often observed in very long series, as in sediment archives of metal contamination (e.g., Bronze Age, Roman and Renaissance).

Stabilized contamination (type F), i.e., very limited change over decades despite pressures, may result from long-term water quality protection, e.g., from international treaties for shared water bodies.

Xenobiotic pollutants have specific trends. The DDT evolution in the Northern Hemisphere CAS presents a gradual increase from flux zero, a marked peak, then a decrease after its ban in the early 1970s, but can still be detected in trace amounts in some rivers due to its great environmental persistence (*total ban*, type J). Herbicide contents, such as that of atrazine, increase generally (*gradual xenobiotic contamination*, type K) until they are finally regulated or the product is banned. For these less persistent products the decline may be rapid (a few years). Artificial radionuclides are often characterized by multiple cycles, as for the artificial radiocaesium in the Northern Hemisphere (*radioactive contamination*, type I) which peaked in 1962-1963, following trends of inputs into the atmosphere from nuclear tests, then again in 1986 after the Chernobyl accident.

Trends of river fluxes are very variable with patterns of both increase and decrease. A similar typology can also be used to describe concentration trends in rivers, lakes and reservoirs, estuaries, and, apart from trend types E and C, to describe groundwater contamination.

5.3. *Water Quality Management: Indicators of Societal Responses to Environmental Issues*

Riverine trends illustrate very different types of water quality management. They are a good example of complex and various societal responses to environmental change (Meybeck, 2002) from the absence of management to full management.

– *Unnecessary management*: water quality is not affected by human pressures; environmental or economic impacts are minimal and the rates of change are slow and predictable.

– *Unplanned improvement*: unexpected and/or unplanned decrease of contamination linked to the reduction of human pressures: closure of mines and industries, changes in technologies, economic crises.

– *Unperceived issue*: deterioration of water quality and/or its link to human pressure is not detected or perceived. Water-quality improvements are unplanned and result from the balance between pressure and natural river basin response. The scientific and technical progress of analytical chemistry has been a major regulator of the detection of water quality problems (Meybeck, 2005). The endocrine disruptors originating from drug residues in domestic wastes, hospital wastes, and veterinary wastes are now beginning to be detected in specific river surveys (Trajectory A, Figure 4), they will probably be regarded soon as an important issue.

– *Natural pressure endurance and suffering*: in some rarely-found geological and climatic conditions (see Section 1). Depending on the availability of alternative water resources, the uses of such resources may lead to limited or to severe health and/or economic impacts.

– *Precaution management*: environmental, health and economic impacts are kept to the minimum acceptable level. If action is taken too late, the level may first exceed the management target then is reduced, such cases are found in highly developed and environmentally aware countries. This type is still rarely found on the planet (Gilbertson, 2001).

– *Maximum impact management*: targeted at the maximum acceptable limit, commonly chosen in international treaties (e.g., salinity in the Rhine and Colorado).

– *Total ban*: ban on manufacturing and/or use of products; usually targeting xenobiotics only after severe problems have been detected, demonstrated and recognized by all stakeholders.

– *Delayed pollution regulation*: established after a period of lack of management and subsequent severe impacts; targeted levels are usually the maximum acceptable ones for economic reasons.

– *Laissez-faire*: although its severity is now well established and even studied, the situation has not yet been adequately tackled for multiple reasons: lack of environmental awareness or societal consensus for the level of severity, shortage of financial means, lack of environmental regulations or of political will to enforce them.

– *Natural pressure remediation*: direct treatment of unsuitable natural water resources (desalinization, defluorization, removal of arsenic).

– *Remediation of ancient contamination*: in most cases there is no present-day economic or administrative entity directly linked to the contamination ('orphan pollution'), which often occurred at times of 'unperceived issues' or of 'laissez-faire'. The corresponding restoration measures are very costly and rarely set up since they require environmental knowledge, societal consensus and financial means (Hines *et al.*, 2001).

– *Cyclic management*: over the very long term (50 to 100 yrs) water-quality presents multiple cycles of deterioration and improvements resulting from the complex interactions of human pressures, environmental impacts and human responses.

Water quality trends result from a combination of human pressures, hydrosystem responses to human pressures, development of social and societal awareness, advances in environmental science and in analytical techniques, political decision processes, financial, technical or policy means and finally, from the hydrosystem response to environmental control. Most of the documented trends concern the last 30 yrs only.

6. Global Assessment of Water Quality Issues

The global assessment of water quality is regularly required by international health programs as well as by Global Change programs (Vörösmarty *et al.*, 2005), although it remains very limited due to the type of information available. There is only one program, launched by UNEP and WHO in 1978, devoted to monitoring harmonization, analytical quality control, data collection and assessment at the global scale: the GEMS-Water program (www.gemswater.org; Robarts *et al.*, 2002). However, despite continuous efforts, the GEMS-Water database is insufficient for global analysis of many issues. Expert judgment based on dozens of country and regional reports and hundreds of publications must be used. However, there are structural limits to our knowledge of water quality: (i) the density of water quality stations is one to two orders of magnitude inferior to hydrological

stations (river gauging and groundwater levels); (ii) the quality of the information is extremely variable ranging from a few basic parameters of limited relevance for human health to several dozens of chemicals including trace contaminants (see Section 4); (iii) survey frequency can limit the assessment of water quality.

The extent of water quality issues may be quite variable, even within one type of contaminant, as demonstrated below for the heavy metals. A tentative global analysis of issues is then proposed.

6.1. *Metal Pollution in River Particulates, an Example of Global Contamination Ranking*

Metals are identified as one of the most dangerous substances found in the environment for their toxic properties and their sensitivity to human pressures. When metals contaminate aquatic systems, they can affect humans through drinking water, aquatic biota and food. Cadmium intoxication (Itai-Itai disease) through contaminated rice, and mercury intoxication (Minamata disease) through contaminated coastal fish are some of the worst environmental issues ever reported and had a great impact on the creation of UNEP in 1972.

A global survey of metal contamination in river basins remains to be established (Salomons *et al.*, 1995). Even at the regional scale, as for Europe, the relevant data to assess the status of contamination are still very limited (Stanner and Bourdeau, 1995). Some synthetic assessments are made for some regions as in the USA (Rice, 1999) or for some specific elements such as cadmium (Cd), mercury (Hg) or arsenic (As). A global vision for other metals, copper (Cu), nickel (Ni), lead (Pb), zinc (Zn), chromium (Cr) and metalloids such as antimony (Sb) and tin (Sn) is difficult (Foster and Charlesworth, 1996; Salomons *et al.*, 1995). Finally, it must be noted that the analysis of metals in dissolved form is very difficult and the analyses from unfiltered waters, often used in the water quality industry, have little environmental meaning (Horowitz, 1995; Meybeck, 2005).

As a first global estimate, I am using here a set of analyses made on river basins from 10,000 km^2 to more than one million km^2 (Meybeck, in preparation). These analyses have been essentially published since 1980 and multi-elemental analytical techniques are used after a complete digestion of the river material. Several sets of stations have been used: (i) natural background stations (BGR) for which there is no or very limited human impact (they also include analysis of pre-industrial river deposits,

from sediment archives); (ii) impacted stations (IMP) for which there is a known human pressure and (iii) undetermined stations (UND). Their combination constitutes the general set (GEN) which holds from 100 to 500 different analyses for a given element. In addition to the most toxic elements, I am also using silver (Ag), lithium (Li), beryllium (Be), barium (Ba), titanium (Ti), strontium (Sr), molybdenum (Mo) and phosphorus (P) for comparison. I also include a set of 10 to 15 multi-elemental urban sewage particulate analyses from all continents (SEW).

Using the medians (GEN_{50}, BGR_{50}) and upper deciles (IMP_{90}, BGR_{90}), I construct three indicators of global scale contamination (i) $I_A = GEN_{50}/BGR_{50}$ for the global sensitivity to contamination; (ii) $I_B = SEW_{50}/BGR_{90}$ for the global occurrence of a metal source in urban sewage and (iii) $I_C = IMP_{90}/BGR_{90}$ for the local occurrence of marked contamination. The three indicators are very convergent although they also express specificities for some elements (Table 7).

Many elements such as Ag, As, Ba, Be, Co, Li, Mo, Sr, Ti, and V are not globally affected by human activities. Human impact on Cr, Ni, P (particulate) and Sb is still very limited. This does not mean there is no impact at all at the local scale: the I_C indicator shows that the upper arsenic and phosphorus deciles of the general distribution are 3 to 5 times higher than high background values (upper decile of BGR set).

The elements that show the most evidence of environmental contamination are mercury, cadmium, possibly tin (to be confirmed on a larger data set), then copper, lead and zinc. These elements are also found at very high levels in at least 10% of documented stations, exceeding ten times the extreme background values ($I_C > 10$). Silver and copper are also often found locally at very high levels.

The analysis of sewage particulate matter (sludge) provides some clue to the origin of elements. Urban sources are very likely for silver, mercury, cadmium, zinc, copper, lead and phosphorus. Urban sludge is not a major source of contamination in rivers since it is actually diluted for As, Ba, Co, Li, Mo, Ni, Sr and V with regard to background levels in river particulates, probably due to the presence of large amounts of sewage organic material.

Other major sources of metals at the global scale include ore extraction and processing in mines and smelters (e.g., Pb, Zn, Ti, Cd) and plating (Cd, Cr, Hg) (see Figure 3 and Table 4). The assessment of the inorganic contamination should now be refined at the regional scale: the contamination orders may be different from those presented on Table 7 depending on the

TABLE 7. GLOBAL SCALE CONTAMINATIONS OF RIVER PARTICULATE MATTER FOR 19 ELEMENTS BASED ON THREE INDICATORS (IA, B, C). Statistics based on 100 to 500 river stations (Meybeck, in preparation).

A – Global sensitivity to contamination ($I_A = GEN_{50}/BGR_{50}$)

I_A	< 1.1	1.1 – 1.25	1.25 – 1.5	1.5 – 1.75	> 1.75
	no effect	some effect	low contamination	medium contamination	high contamination
	Ag	Cr	Cu	Cd	Hg
	As	Ni	Pb	(Sn)	
	Ba	P	Zn		
	Be	Sb			
	Co				
	Li				
	Mo				
	Sr				
	Ti				
	V				

B – Occurrence of metal source in urban sewage ($I_B = SEW_{50}/BGR_{90}$)

I_B	< 1	1 – 2	2 – 5	5 – 10	10 - 20	> 20
	no source	limited source	some source	important source	high source	essential source
	As	Cr	(Sn)	Cd	Hg	Ag
	Ba	Sb		Cu		
	Be			P		
	Co			Pb		
	Li			Zn		
	Mo					
	Ni					
	Sr					
	V					

C – Local occurrence of marked contamination ($I_C = IMP_{90}/BGR_{90}$)

I_C	< 1.5	1.5 – 3	3 – 5	5 – 10	10 - 20	20 - 50	> 50
	not likely	very limited	limited	common	very common	common hotspots	very common hotspots
	Ba	Co	As	Cr	Cu	Hg	Ag
	Be	Ni	P	Sb	Pb	Zn	Cd
	Li	Ti					Sn
	Mo	V					
	Sr						

Note: Sn on 60 rivers only; As contamination is better assessed on dissolved As.

human pressures ratio mining:industrial:urban. This can only be done on the basis of systematic sampling of river particulates such as has been done for the past twenty years for the conterminous USA (Rice, 1999; Horowitz *et al.*, 2001).

6.2. *Global Ranking of Water Quality Issues Based on Regional Assessment*

The first global assessment of water quality (Meybeck *et al.*, 1989) already pointed out our fragmented information on water quality at the global scale. A second attempt has been made with the Dublin International Conference on Water and the Environment (ICWE) (Meybeck *et al.*, 1991). It has been recently updated for the Millennium Assessment (Vörösmarty *et al.*, 2005).

Eleven variables are considered and ranked: fecal pathogenic agents, organic matter (oxygen-consuming, also termed carbonaceous pollution), salinization, nitrate (as a contaminant), fluoride (mostly from natural sources), eutrophication (and/or nutrient levels), pesticides, industrial organics (PAH, PCB, petroleum products, etc.), heavy metals, suspended sediment (as limiting water uses), and acidification (may occur only if the natural buffering capacity of soils is low). The scoring ultimately reflects the aggregate impact of human pressures, natural rates of self-purification and pollution control measures (Figure 7).

Updated results show that *pathogens and organic matter pollution are still the two most pressing global issues* (Figure 7), reflecting the widespread lack of waste treatment. As water is often used and reused in a drainage basin context, a suite of attendant public health problems arise, thus directly affecting human well-being. At the other extreme, acidification is ranked #10 and fluoride pollution #11 on the global scale.

At the regional scale, any issue can be important or severe, e.g., acidification in Northern Europe and Northeast North America, salinization for the Arabian peninsula, fluoride in the Sahel or the African Great Lakes (see maximum scores reached on Figure 7). Fluoride and salinization issues are mostly due to natural conditions (rock types and climate), but mining-related salinization can also be found (e.g., W. Europe), and salinization can be enhanced by irrigation returns to CAS as in the Aral Sea basin (Aladin *et al.*, 2004). Other issues are directly caused by human impacts. It is important to note that in many regions of the world still under limited human pressures, many of these issues have been judged as negligible.

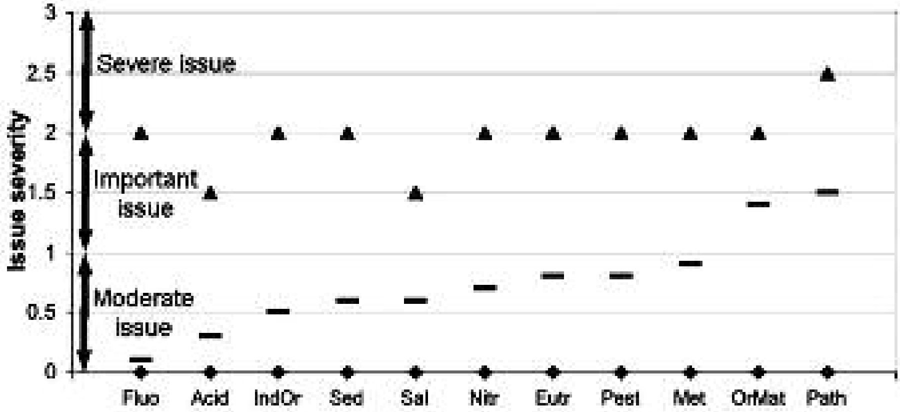


Figure 7. A tentative ranking of globally significant water quality issues based on expert judgement for ▬ regional assessment (typically 108 km²). (The information base upon which to quantify the degree to which water supplies are compromised is currently insufficient). (Fl = Fluoride; Acid = acidification; IndOr = industrial organics; Sed = suspended sediment; Sal = salinisation; Nitr = nitrate as a contaminant; Eutr = eutrophication and/or nutrient levels; Pest = pesticides; Met = heavy metals; OrMat = organic matter; Path = pathogenic agents). (Issue severity rank: 0= No problem or irrelevant, 1= Some pollution: water can be used if appropriate measures are taken, 2= Major pollution: impacts on human health and/or economic use, or aquatic biota is important, 3= Severe pollution: impacts are very high, losses concern human health and/or economy and/or biological integrity, ND= Could not be assessed (mostly for industrial pollutants). ▲ Maximum issue severity reached at the regional level, ◆ Average issue level at global scale, ▬ Minimum issue severity observed (modified from Revenga *et al*., 2005).

6.3. *Human Impacts on Continental Aquatic Systems within an Earth System Analysis; the River Syndromes*

Changes occurring in continental aquatic systems will also generate indirect effects on human health through their participation in global environmental change, i.e., river fluxes and concentration of carbon, nutrients and contaminants in aquatic systems (Tables 2, 4 and 5, columns B to G). Human activities are generating fast impacts, which can be organized into a set of global river syndromes.

The concept of global syndromes has been developed by the German Advisory Council on Global Change (GACGC, 2000) and defined as 'typical patterns of problematic people-environment interactions which can

be found worldwide and can be identified as regional profiles of damage to human society and ecosystems'. This concept has been extended to 10 river syndromes (Meybeck, 2003a,b): flow regulation, fragmentation of river course, riverbed silting, desiccation, chemical contamination, acidification, eutrophication and microbial contamination, and to land subsidence and groundwater over-pumping in deltas. Other syndromes, such as thermal regime alteration, radio-nuclide contamination, and biological invasion, are likely to occur, but will not be discussed here. Each syndrome is defined by a set of symptoms and causes and can be illustrated from well-studied river basins. The desiccation syndrome (originally termed neoarheism by Meybeck, 2003b) corresponds to the drastic reduction of river flow and/or lake area due to water diversion and water use, particularly for irrigation as observed in the Amu Darya basin (Kayunov, 2004). It is here understood as a flow or area reduction of at least 50% with regards to previous average.

The modification of river systems, either natural or anthropogenic, can be analyzed from both Earth System's and water resources' perspectives, including health aspects. River syndromes affect:

(i) sediment balance, which controls fluvial and coastal morphology and generates alluvial aquifers and flood plain habitat;

(ii) the hydrological balance of large continental water-bodies and regional seas in particular, which may also influence coastal nutrient dynamics as from up-welling, and deep ocean water formation;

(iii) carbon balance, such as organic carbon transfer and burial, CO_2 uptake during silicate rock weathering (a major control of atmospheric CO_2 at the geological time scale), and CO_2 release by wetlands and large rivers;

(iv) the nutrient balance of nitrogen, phosphorus and silica species which control level and type of aquatic primary production (e.g., diatoms vs. cyanobacteria);

(v) emission of green house gases; and

(vi) the aquatic biodiversity and trophic balance of continental and coastal systems.

The ecological responses of continental aquatic systems to these syndromes are not developed here, except for eutrophication, although their extension and importance is now more and more established (Revenga *et al.*, 1998; WCMC, 1998; Rabalais and Turner, 2001).

The syndromes are examined in Table 8, where example rivers and the relative alteration of Earth System functions are given. They generally occur at medium (10-50 years) to long-term time scales (> 50 years) (with

reference to human time scales) after the beginning of riverine change and at local (10^2-10^4 km^2), regional (10^4-10^6 km^2) continental and global (10^6-10^8 km^2) scales. They can develop far away from their primary causes (teleconnections over 1,000 km). For instance, the impacts of large dams rapidly and profoundly modify the sediment routing of fine suspended particles and of sand, but the related coastal zone erosion and shoreline regression in response to this 'sediment starving' may be maximum with a 50 to 100 years time-lag after construction of the reservoir and last as long as the reservoir, i.e., hundreds of years. The response of river bio-coenoses and of its biodiversity, therefore of diseases vectors, to changes may be rapid (i.e., damming effect on migratory species) or slow (e.g., species invasion through interconnection of basins by navigation canals). The global loss of aquatic biodiversity is certainly a major change in the Earth System although its long-term impact has not yet been assessed.

Desiccation is one of the most spectacular syndromes (more than 90% flow reduction for the Colorado, Nile, and Amu Darya, 80 % reduction for the Indus, seasonal desiccation of the Huang He, etc.). It is caused by consumptive water use, especially in large-scale irrigation in arid and semi-arid areas, estimated to be *ca.* 4,000 $km^3.y^{-1}$ (Gleick *et al.*, 2001), i.e., *ca.* 10% of the natural river water flux to oceans, and should be considered in Global Climate Models.

Both positive and negative impacts are noted for human health (Table 8). The positive impacts essentially concern water quantity: water storage for drought protection, reduction of extreme flows and increased flow regularity have been permanent targets for civil engineers for millennia. Some water-related health hazards have also been reduced by land use changes such as wetland reclamation and pesticide use against malaria or onchocercosis (Holland and Peterson, 1995). The trapping of contaminated particulate matter in river systems can be also regarded as positive if permanent. The negative impacts of river syndromes on aquatic resources mostly concern water quality (see Table 4).

The direct impacts of human activities on aquatic systems with regards to health issues are very diverse, mixing negative impacts, mostly a degradation of water quality and a loss of river dilution power, with positive impacts such as flood and drought regulation, settling and processing of particulate contaminants, and attached pathogens in reservoirs. Impacts on Earth System functions are multiple and occur sometimes at very broad temporal and spatial scales. The present global distribution of these river syndromes will have to be established.

TABLE 8. MAJOR SYNDROMES OF CHANGES IN AQUATIC SYSTEMS AND RELATED HEALTH ISSUES (modified from Meybeck, 2003b).

Syndrome[1]	Coding [2]	Examples	Health Issues	Global Health impacts
Acidification[(*)]	R (Z)	Scandinavia, Kola P., E. Ontario, Quebec, Pennsylvania	• increased Al^{3+} and dissolved heavy metals	*
Faecal contamination	T	Most W. Europe rivers in mid 1900's; Brazil (Piracicaba), India (e.g. Yamuna); populated China, etc.	• Waterborne pathogens	***
Chemical contamination[(***)]	U	as for faecal contaminants	• Anoxic waters; ammonia; H_2S	***
	Q	Most W. Europe rivers (1950-1980), Kola peninsula rivers, Don	• Increased metals contents	0 to ***
	L	W.Europe, China, India; many aquifers	• Nitrate contamination	***
	M, Q	W. Europe rivers, Mississippi	• Xenobiotics occurrence	** to ***
	Q, X	Idrija R., Rio Tinto, Coeur d'Alene L., Love Canal	• Persistent pollutants; leaks from Historical pollution	*
Salinization[(**)]	G	Amu Darya, Syr Daria, Colorado, Murray Rivers	• Increased salt contents	**
	G (Q)	Rhine, Weser, mining districts		
Eutrophication[(**)]	L, Δ5	W. Europe rivers (Rhine, Seine, Loire), Volga; Mississippi, Danube deltas; North Sea, Brittany coastal zones	• Harmful algal blooms	*
Flow regulation[(**)]	C	Most European and US rivers, most dammed rivers (Moscowa, Nile, Indus, Parana, Murray Rivers)	• Flood plain area reduction: loss of vector-borne wetlands	++
	O_1-O_2	Colorado, Rio Grande, Columbia, Missouri, Volga, Dniepr, Murray, Bay James, Orange, Sao Francisco Rivers	• Reduction of flood/droughts hazards	+++
Damming and fragmentation [(***)]	H	Colorado, Rio Grande, Columbia, Missouri, Volga, Dniepr, Murray, Bay James, Orange, Sao Francisco Rivers	• Trapping of microorganisms attached to particulates	+++
	H	Colorado, Rio Grande, Columbia, Missouri, Volga, Dniepr, Murray, Bay James, Orange, Sao Francisco Rivers	• Trapping of particulate contaminants	+++
	P_1	Colorado, Rio Grande, Columbia, Missouri, Volga, Dniepr, Murray, Bay James, Orange, Sao Francisco Rivers	• Creation of new wetlands	*
Change of sediment balance[(**)]	J	Huang He, Kotri (Nepal); Madagascar rivers; most small tropical island rivers; Queensland rivers; New Guinea rivers	• Accelerated erosion and transfer of microorganisms	*
	Δ3	many tropical islands; Queensland	• Coastal silting and coral die out	?
	B	Huang He	• River course shifting: flooding hazards	**
Desiccation[(**)]	A	Huang He; Amu Darya; Syr Darya	• Shift from permanent flow to seasonal drought; major reduction of annual flow: loss of dilution power	**
	V	Colorado, Rio Grande, Nile, Indus, Huang He, Amu Darya, Syr Darya, Shattal Arab, Ebro, Orange	• Marked to total reduction of material fluxes at river mouth: trapping of contaminants on land	++
	P_2	Lake Nasser; L. Kariba; L. Volta	• Creation of wetlands in irrigated area	*
Delta subsidence[(*)]	Δ7	Mississippi, Rhone, Indus	• Salinization	*
Groundwater abstraction[(*)]	Y	Bengal aquifer	• Remobilisation of arsenic	**

Health issues degradation: * locally important, ** regionally important, *** globally important; Health issues improvement: + (some improvement) to +++ (major improvement) (1)Global impact on Earth System functions: * locally important, ** regionally important, *** globally important (2)Coding refers to Figure 1 and Tables 1, 2, 3.

7. Future Evolution of Aquatic Systems

The global picture of aquatic systems in the next 50 or 100 years is still very fuzzy. There are now growing efforts to model some issues or syndromes at the global scale as is done for nutrients and carbon river fluxes. These models are originally based on multi-regression analysis linking human pressures and the resulting state of river quality. Mixed models now integrate pressures, river basin filters and water routing (e.g., Green *et al.*, 2004). A new generation of process-based models is now developed at the basin scale for nutrients (Billen *et al.*, 2001) but their application at the global scale will be difficult for lack of basic data at the appropriate resolution. As for the Global Climate Models (GCM), these models will be our tools to explore the future of aquatic systems. They should be validated first on the present situation. Yet the available data on aquatic systems is often relatively short-term, particularly concerning water quality. Considering the scales of responses of aquatic systems to climate variations during the last thousand years and to human activities, long-term evolutions (> 100 y) are also needed to validate the river basin models. They will have to be established using a combination of methods now developed by the paleohydrology community (PAGES-LUCIFS, 2000). Once these models are validated they will be used to explore the future, combining GCM scenarios, scenarios of water use, and scenarios of human responses to changes. The validity of the prediction will greatly depend on the model resolution (2° for most GCM, 0.5° for most river flux models).

7.1. *General Evolution of Human Pressures and Responses*

The timing of global human pressures, environmental impacts and societal responses is schematically depicted in Figure 8 (Meybeck, 2003b and Meybeck *et al.*, 2004).

Major human pressures only are considered here, and it is postulated how an increasing fraction of the Earth's surface has been exposed to these. River engineering here includes damming, channelization, diversion and irrigation canals. The evolution of proportions of global area or affected global population is still speculative owing to the lack of databases, but there are growing efforts in reconstruction of historical land use and population density. The progression towards a global scale impact can take two pathways. With the first, impacts are locally displayed, but because of the pandemic distribution of a particular class of

change, the consequences are global. A good example is the widespread conversion of land to agriculture and forestry.

Global scale impacts also arise from teleconnections operating over the planetary domain. An example is the long-range atmospheric transport of pollutants such as NO_x and SO_2, responsible for the acidification and/or eutrophication of surface waters, sometimes hundreds of kilometers away from emission sources. These statements should not imply that all riverine impacts are now globally significant (see previous section). In fact, most well documented impacts on aquatic systems are local to regional. Since the majority of human induced sources of pressure on the CAS have had an exponential rate of increase over the last two hundred years, the spatial distribution of these combined forces has now moved on to the planetary scale. The continuing and fast rate of change thus necessitates the accelerated time scale adjustment on Figure 8.

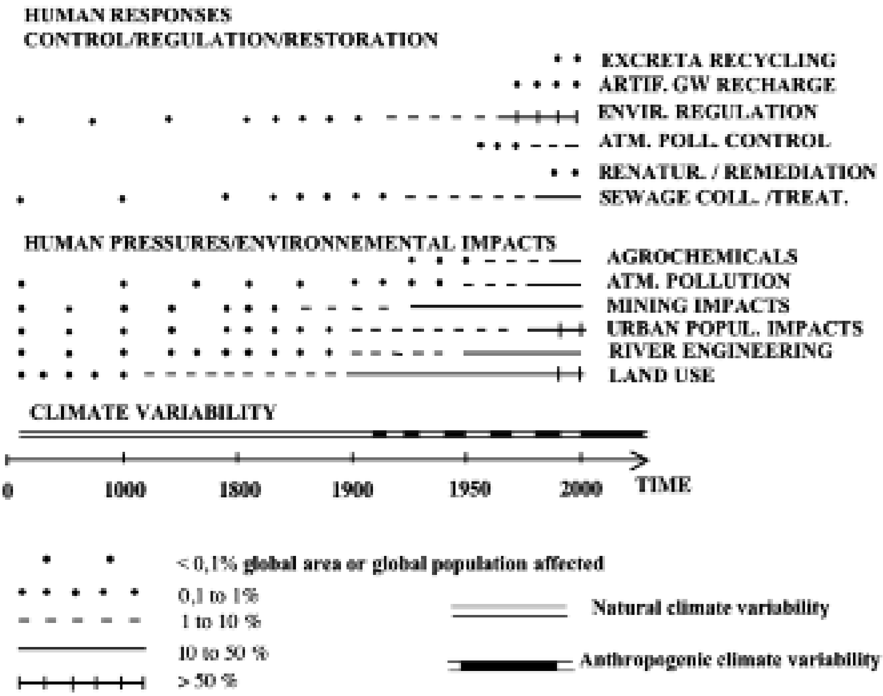


Figure 8. Working hypotheses on the occurrence of some major pressures on continental aquatic systems at the global scale and related environmental remediation responses (accelerated time scale) (Meybeck, 2003b).

The key-control of river impacts is the relative timing of human pressures and societal responses, such as policy and regulation, emission control, and restoration, as has been developed for chemical contamination. As seen before, these responses have generally been developed with a considerable lag related to pressures.

7.2. *Contrasted Historical Evolution of Continental Aquatic Systems*

The past evolution of CAS may be very different from one region to another. The reconstruction of trajectories of these environmental changes is essential to avoid present-day mismanagement, and to consider environmental issues on a long-term basis, i.e., more than 50 years (Harremoes *et al.*, 2001). River basin evolution has been rarely addressed so far (Schwartz *et al.*, 1990; Messerli *et al.*, 2000; Vörösmarty and Meybeck, 2004), although this field is now covered by the IGBP-PAGES programme (PAGES-LUCIFS, 2000).

Reconstruction of historical interactions between human and aquatic systems is based on four types of information:

(i) *Sedimentary archives (10^2 to 10^3 years)*: they can be deciphered to reconstruct the past riverine concentrations and/or fluxes on alluvium, in lakes, deltas and coastal sediments (Valette-Silver, 1993; Foster & Charlesworth, 1996) as performed at the global scale by the IGPP-PAGES LUCIFS project (Meybeck *et al.*, 2004). More recent archives (10-100 yrs) can be obtained from reservoirs;

(ii) *Archaeological and historical archives* give valuable information on river systems and their uses and on societal responses to river basin changes (Guillerme, 1983; Schwartz *et al.*, 1990). The longest and most promising historical records of man and river interactions are probably found in China (Elvin, 1993; Elvin and Liu, 1998) and in Egypt;

(iii) *Direct observations*: date back to the early 1800s, and the earliest regular river surveys started before the 1900s;

(iv) *Back-casting* of river basin quality combines present-day validated biogeochemical or ecological models and historical information on human pressures, such as land use and water use (Billen *et al.*, 2001).

Two examples of working hypotheses for past river evolution are presented here (Figure 9) for Western Europe, an example of very ancient impacts, and South America for recent ones (Meybeck, 2003b). Four river quality indicators specifically related to health issues are proposed here, using an accelerated time scale, reflecting the evolution of some human

impacts: organic and fecal contamination (Figure 9, 1), heavy metals (2), nitrate (3), and pesticides (4). As in Section 5, a simplified issue severity scale in three steps is used here, where C_N is the natural or pristine concentration, C_R a first threshold above which environmental impact, health issues, cultural or economic loss are occurring, and C_L a second threshold above which severe impacts are occurring.

In Western Europe (Figure 9), the earliest changes in river chemistry and assumed severe impacts have been recorded for metals in mining districts (4) as early as the Bronze age period (~4500 yrs B.P.) as in the Rio Odiel, Spain (Leblanc *et al.*, 2000), with maximum levels of Hg, Pb and Zn equivalent to those found presently in some highly contaminated European rivers. Such mining impact is likely to have been very localized: larger basins were probably much less contaminated. Other examples of metal contamination from mining are documented for Roman times (Wales rivers and Humber catchment, England; Macklin *et al.*, 1997), the Middle Age in Central Germany (Goslar), and in the 1700s in Brittany. Modern contamination peaks have been observed in the mid-19th century then in the 20th century as in the Rhine and Meuse rivers (Middlekoop, 1997). The most recent metal contamination generally peaked in most Western European rivers between 1950 and 1980.

Organic and fecal contaminations (1) can be multi-cyclic as is well documented for the last 150 years in the lower Thames River (Schwartz *et al.*, 1990): it mostly depends on the relative production, collection and treatment of urban wastes, i.e., on the ratio of collected population/sanitation, as is also well documented for the Seine Basin (Barles, 2002). In many European rivers, the maximum general contamination was noted in the 1950s and 1960s when the sewage collection rate increased, yet without appropriate waste water treatment which was generalized in the 1970s and 1980s. This evolution is well documented through oxygen demand, ammonia and fecal coliforms, which peaked during the 1950-1970 period.

Nitrate contamination has gradually developed after World War II following the general use of fertilizers in intensive agriculture (Cole *et al.*, 1993). In Western Europe, it is now approaching the severe impact level (50 mg NO_3^-/L) set by WHO at which the water should not be used for drinking. But the severe level for coastal phytoplankton development, set at a much lower river concentration, had already been exceeded in the 1960s to 1970s. As a consequence, severe coastal algal blooms have followed in the North Sea and in Brittany. In the Rhine River, nitrates have been slowly decreasing since 1990 (ICPR, 2001); in other rivers, they are nearly stabilized (Seine) or still increasing (Southern Europe).

Contamination by pesticides (4) has been rapidly growing since the 1970s and, in a given medium-sized basin such as the Seine's, over 100 different active molecules might be used (Chevreuil *et al.*, 1998). The use of such xenobiotic substances is now more and more regulated in Western Europe and North America.

In South America, the expected evolution of river chemistry is somewhat different (Figure 9, lower part). Riverine quality is not likely to have changed much prior to the arrival of European settlers, except for limited

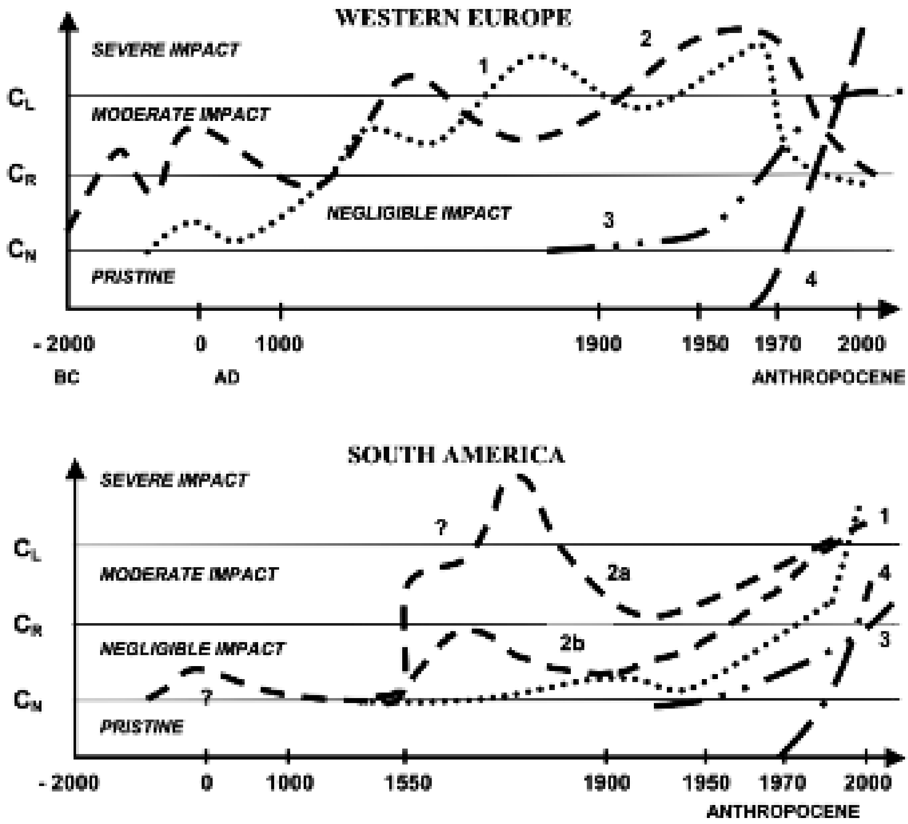


Figure 9. Working hypotheses on the evolution of some chemical contamination indicators in medium-sized Western Europe and in South American river basins. (1): organic and faecal contaminations. (2): metal contamination. (3): nitrate contamination. (4): pesticides. Accelerated time scale. C_N, C_R, C_L: natural, recommended and limit levels for related uses. South America: 2a = evolution of basins affected by Colonial American mining, 2b = other South American rivers (Meybeck, 2003b).

agricultural land-use impacts from Pre-Columbian civilizations for example on river sediment fluxes. A very slight atmospheric lead contamination during the Roman times is theoretically possible since long-range human impact of Pb and Ag mining and smelting has been documented in the Northern Hemisphere (Shotyk *et al.*, 1998 ; Renberg *et al.*, 2000), although less likely to have impacted the Southern Hemisphere. The most striking possible feature of human impacts on South American rivers can be found in Peru and Bolivia (2a): the gold and silver mining and the mercury amalgamation performed by the Spanish settlers since the mid-1500s (Brading and Cross, 1972) have probably generated an enormous direct and indirect mercury contamination via atmospheric pathways (Lacerda *et al.*, 1999), which remains to be validated using sediment archives.

In other South American regions, riverine changes are likely to have occurred mostly over the last 50 yrs but will probably be much faster than in Europe. Organic pollution, toxic metals, and xenobiotics contamination are probably now reaching their maximum levels, due to the growing imbalance between pressures and environmental regulation. Examples are the Piracicaba River in Sao Paulo state (Mariely *et al.*, 2002) and other Brazilian rivers (Knoppers *et al.*, 1999).

7.3. *Possible Scenarios for the Future Contamination of Rivers*

The evolution of rivers over the past 2,000 years can help us to foresee some possible future scenarios of water quality for the next 50 years. Although the precise evolution of rivers will be basin-specific, a schematic trend is proposed here as a working hypothesis and illustrated for chemical contamination (Figure 10). It is probably valid for other river syndromes.

Human pressures have started at a very local scale. There is an inverse relationship between the impact severity and the spatial scale of contamination (Meybeck *et al.*, 1989, 2004). Chemical contamination was still very limited some 2,000 years ago and likely occurred at the local scale only ($< 10^4$ km^2). Two hundred years ago, most chemical contamination symptoms were moderately developed at the regional scale. In the mid-19th century, chemical contamination reached a moderate to severe level in some regions of Western Europe and in parts of the Eastern USA, but was still negligible on many continents, while the transfer of atmospheric pollutants over long distance was already limiting the occurrence of truly pristine basins. In the present Anthropocene period, river chemical contamination is now widespread and the occurrence of very severe contamination levels at the local

scale is well documented as has been demonstrated for some metals (Table 7) (mega-cities, historical pollutions, mining and smelting districts, etc).

Three main future scenarios (2000-2050) are envisaged here (Figure 10):

(i) *Business as usual and Laissez-Faire* (Figure 10, Curve A): although regulation/restoration responses may be developing on all continents, human pressure is still increasing rapidly. The global contamination and the artificialization of continental aquatic systems accelerate, leading to a generalized degradation of aquatic habitat and an expected response of aquatic biota, particularly in the coastal zone. From the analysis of recent river evolution, it can be assumed that such a policy has been applied until the end of the 1980s in Eastern Europe and in the Former Soviet Union (Kimstach *et al.*, 1998), and in most fast-developing countries such as in China (Wang *et al.*, 2000), Brazil, and India (Meybeck *et al.*, 1991);

(ii) *Priority reduction of river impact hot spots* (Figure 10, Curve B). Such a scenario applies mostly to the water quality issues. Environmental management is here targeted to the most severe pollution issues, either contemporary or historical (remediation of polluted sites), according to a cost/benefit analysis. This policy has been applied in the past in most Western European countries and the USA in the 1960s to 1980s. In such a scenario, the biggest point sources of pollution and the most contaminated sites are cleaned up first, but there is a gradual shrinking of the remaining sub-pristine river basins and a homogenization of river conditions towards a mediocre quality;

(iii) *Precaution management* (Figure 10, Curve C): in addition to the previous management rule, human impacts, either direct or indirect, are generally limited to the lowest acceptable impact. This type of policy is now being developed by the European Union in its new Water Framework Directive. It has been favored for two to three decades by some countries such as the Scandinavian countries, Switzerland, and Canada. However, in such a scenario, some moderate and even severe impacts are likely to remain at the local level due to structural factors (e.g., a mega-city located on a small watershed with limited dilution power). This policy requires a combination of citizen awareness, water literacy, scientific and technical knowledge, political will and financial means, which is unlikely to be found everywhere.

These scenarios should now be combined with water runoff and river flow scenarios resulting from Global Climate Models (GCM), and with water use and water engineering scenarios. The occurrence and future development of dams, water diversions and irrigation will greatly influence the dis-

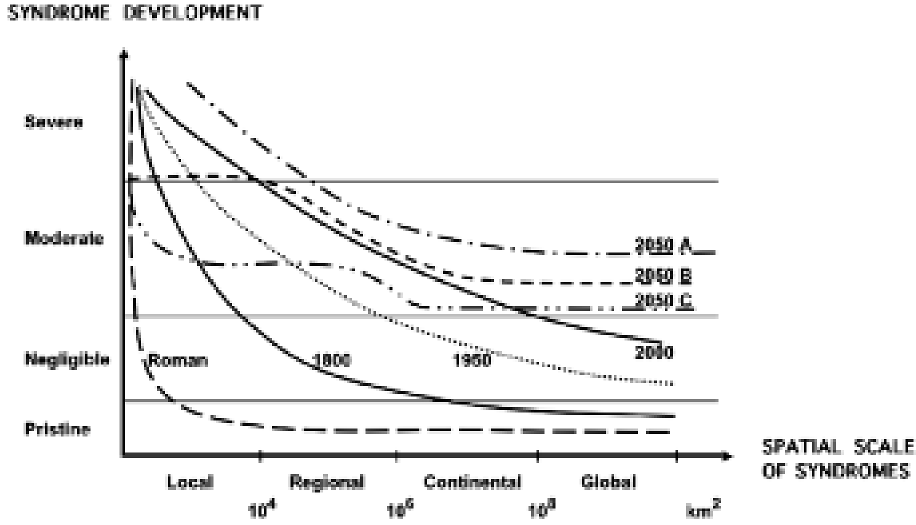


Figure 10. Schematic evolution of the chemical contamination from point sources at different space scales from Roman period to contemporary. Scenarios for year 2050 correspond to (A) business as usual, (B) priority reduction of the most polluted sites, (C) a general application of precaution principle (Meybeck, 2003b).

tribution types of water bodies, their ecological functions (Petts, 1984) and quality, as well as river fluxes. Some fluxes of materials (nitrogen, phosphorus, mercury, cadmium) across river systems and to oceans have increased already much beyond the natural Holocene variations. Others, such as sediments, are probably stable (Walling and Fang, 2003) despite an acceleration of sources, due to a simultaneous acceleration of retention in riverine filters principally reservoirs. A global decrease of silica concentration levels and fluxes to the coastal zone is now likely. These changes are resulting in marked impacts on coastal biogeochemistry (Rabouille *et al.*, 2001) and food webs (Rabalais and Turner, 2001). These contrasted Earth System modifications will have long-term impacts on both water resources and human health although these cannot yet be assessed.

The continental aquatic systems are now shared by the natural and human components of the Earth System (Crutzen, 2002; Steffen *et al.*, 2004; Meybeck, 2002, 2003b). Their analysis can be made using the OECD Driver-Pressures-State-Impact-Response (DPSIR) already used in some

environmental studies such as coastal zone management (Turner and Bower, 1999; von Bodungen and Turner, 2000; Crossland *et al.*, 2005).

Natural Earth System drivers are climate change and sea level rise, natural climate variability and tectonic forcings (Figure 11, right part). Human drivers such as population increase, economic and technical developments, lead to the increased exploitation of natural resources generating multiple pressures such as water use, waste release, land use, biomass use, increasing development of xenobiotics, and river flow regulation. These pressures modify the state of the Earth System and, in turn, have multiple impacts on water resources. The combined changes of aquatic systems are part of the general modification of the Earth System through alteration of water fluxes, both vertical and lateral, green house gas emissions and river borne fluxes. They occur at various time scales from years or shorter (vertical fluxes to the atmosphere) to hundred of years and more (lateral fluxes of particulates). The coastal response to modified river fluxes may also range from decades to a hundred years and more.

On the human side (Figure 11, left part), these modifications of Continental Aquatic Systems are generating both negative impacts (Figure 11, lower left) and positive impacts (water supply security and drought control, secured communications through waterways, decrease of water-borne and water-related diseases, flood control). The assessment of the evolution of continental aquatic systems (CAS) under Global Change combines multiple degrees of complexity, which are just beginning to be addressed (GWSP, 2005). The health issues related to the evolution of aquatic systems must be addressed at sub-regional to local scales due to their spatial heterogeneity and to the variety of human responses, which can take decades in most cases. The first global scale models of water quality now established for nitrogen and phosphorus at 30'x30' resolution (i.e., 50x50 km at the equator) already show an enormous heterogeneity. This scale is sufficient for global assessment but not for local management. The analysis of water quality trends and the construction of hypotheses about their historical variation also reveal the complex interactions of humans and water bodies, which vary from one river basin to another although regional patterns are likely.

The future of human development will greatly depend on how we will balance these negative and positive aspects. This will require new approaches (e.g., the Earth System analysis with its teleconnection and delayed impacts should now be considered in the Integrated River Basin Management) and new concepts such as hydrosolidarity and water liter-

acy will have to be considered (Falkenmark, 1997; Lundqvist and Falkenmark, 2000; Falkenmark and Lundqvist, 1998). This multidisciplinary field is now wide open.

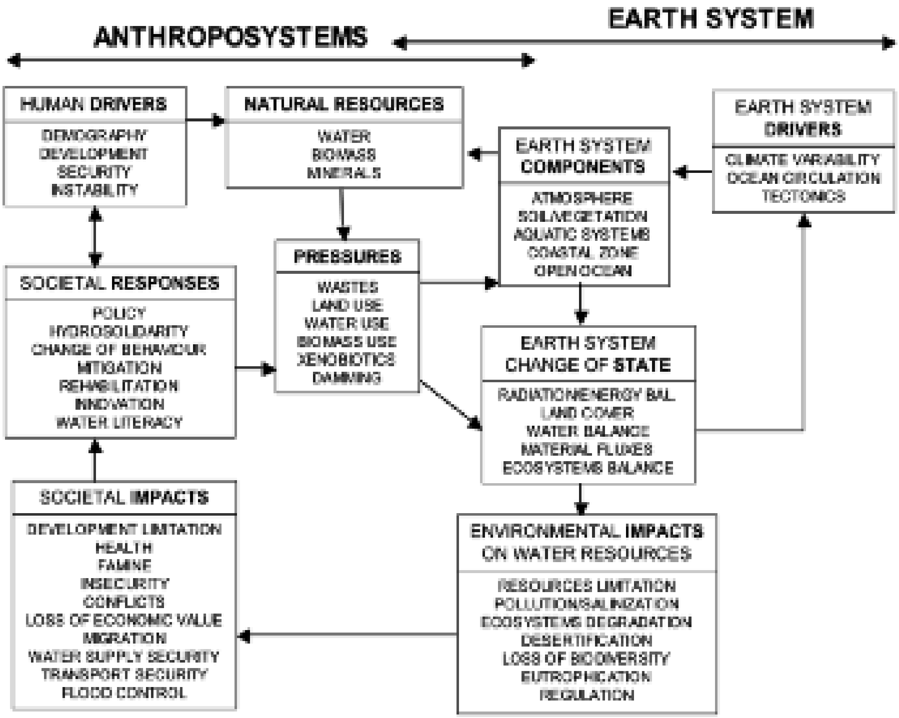


Figure 11. The Driver-Pressure-State-Impact-Response cycle on continental aquatic systems (modified from Salomons *et al.*, 1999 and Falkenmark *et al.*, 1999).

ACKNOWLEDGEMENTS

I am particularly indebted to Séverine Roussennac for her technical support. Many ideas developed in this paper have been raised through interactions with numerous colleagues in IGBP projects, BAHC, LOICZ and GWSP, particularly with Charles Vörösmarty, Wim Salomons and Christian Levêque, I thank them all for our fruitful discussions.

REFERENCES

Aladin, N.V., Plotnikov, I.S., Létolle, R. (2004), Hydrobiology of the Aral Sea. In: J.C.J., Nihoul., P. Zavialovand and P.P. Micklin, *Dying and Dead Seas*, Kluwer, NATO Series, pp. 125-157.

Anderson D.H., Glibert P.M., Burkholder J.M. (2002), Harmful algal blooms and eutrophication: nutrients sources, composition and consequences, *Estuaries*, 25, 704-726.

Anderson M.G., Peters N.E., Walling D. (1996), Trends in water quality, *Hydrol. Processes* (Special Issue), 10, 127-356.

Barles S. (2002), L'invention des eaux usées: l'assainissement de Paris 1780-1930. In: Bernhardt C., Massard-Guilband G. (eds.), *The Modern Demon. Pollution in Urban and Industrial European Societies*, Presses de l'Univ. Blaise Pascal, Clermont-Ferrand, pp. 129-156.

Berner E.K., Berner R.A. (1987), *The global water cycle, geochemistry and environment*, Englewood Cliffs, NJ: Prentice-Hall.

Billen G., Garnier J., Ficht A., Cun C. (2001), Modelling the response of water quality in the Seine River Estuary to human activity in its watershed over the last 50 years, *Estuaries*, 24, 977-993.

Boon P.J., Howell D.L. (eds.) (1997), *Freshwater quality: defining the undefinable?*, The Stationary Office, Edinburgh, pp. 552.

Brading D.A., Cross H.E. (1972), Silver mines in colonial America, *Hispanic Am. Hist. Rev.*, 52, 547-549.

Chapman, D. (ed.) (1996), *Water Quality Assessments: A Guide to the Use of Biota, Sediments and Water in Environmental Monitoring*, Second edition, Chapman and Hall, London, pp. 450.

Chevreuil M., Thevenot D., Scribe P., Blanchard M., Duclos Y., Garmouna M., Gonzalez-Calderon A., Trace S., Teil M.J. (1998), Micropollutants organiques: une boîte de Pandore. In: *La Seine en son bassin*, M. Meybeck, G. de Marsily, E. Fustec (eds.), pp. 439-481, Paris: Elsevier.

Chilton, J. (1989), Salts in surface and groundwaters. In: M. Meybeck, D. Chapman and R. Helmer, *Global freshwater quality, a first assessment*, Blackwell, pp. 139-158.

Chorus I., Bartram J. (eds.) (1999), *Toxic Cyanobacteria in Water. A guide to their Public Health Consequences, Monitoring and Management*, Spon Press, pp. 416.

Cole J.J., Peierls B.L., Caraco N.F., Pace M.L. (1993), Human influence on river nitrogen. In: M. McDonell and S. Pickett, *Humans as components of ecosystems: the ecology of subtle human effects and populated areas*, Berlin, Springer, pp. 141-157.

Colwell, R.R. (1996), Global climate and infections disease: the cholera paradigm, *Science*, 74, 1025-1031.

Crossland, C.J., Kremer, H.H., Lindeboom, H.J., Marshall, J.I. and M.D.A. Le Tissier, (2005), *Coastal fluxes in the Anthropocene*, Springer, pp. 231.

Crutzen P. (2002), Geology of mankind: the Anthropocene, *Nature*, 415, 23.

Elvin M. (1993), Three thousand years of unsustainable growth: China's environment from Archaic times to the Present, *East Asian Hist.*, 6.

Elvin M., Liu T.J. (eds.) (1998), *Sediments of time. Environmental society in Chinese history*, New York, Cambridge University Press.

Falkenmark M. (1997), Society's interaction with the water cycle: a conceptual framework for a more holistic approach, *Hydrolog. Sci. J.*, 42, 451-466

Foster I.D.L., Charlesworth S.M. (1996), Heavy metals and the hydrological cycle: trends and explanation, *Hydrol. Process*, 10, 227-261.

Froebrich, J. and O. Kayumov (2004), Water management aspects of Amu Darya. In: J.C.J. Nihoul, P. Zavialov and P.P. Micklin, *Dying and Dead Seas*, Kluwer, NATO Series, pp. 49-76.

GACGC (2000), *World in transition. Strategies for managing global environmental risks*, German Advisory Council on Global Change, Annual Report 1998, Berlin: Springer.

Garrels R.M., Mackenzie F.T., Hunt C. (1973), *Chemical cycles and the global environment*, Los Altos, CA: William Kaufman Inc.

Gems Water: www.gemswater.org

Gilbertson M. (2001), The precautionary principle and early warnings of chemical contamination of the Great Lakes. In: *Late lessons from early warnings: the precautionary principle, 1896-2000*, Environmental issue report 22, pp. 126-134, Copenhagen: European Environment Agency.

Gleick P.H. (1993), *Water in crisis. A guide to the world's freshwater resources*, London: Oxford University Press.

Gleick P.H., Singh A., Shi H. (2001), *Threats to the world's freshwater resources*, Oakland, CA: Pacific Institute for Studies in Development, Environment, and Security and UNEP.

Green P., Vörösmarty C.J., Meybeck M., Galloway J., Peterson P., Boyer C. (2004), Preindustrial and contemporary fluxes of nitrogen through rivers: a global assessment based on typology, *Biogeochemistry*, 68, 71-105.

Guillerme A. (1983), *Les temps de l'eau: la cité, l'eau et les techniques*, Seyssel, France: Champ Vallon.

GWSP (2005), Global Water System Programme (www.gwsp.org).

Hales S. (2003), Impacts on health of climate extremes. In: McMichael, *Climate change and human health. Risk and responses*, pp. 79-102.

Harremoes, P. (ed.) (2001), *Late Lessons from Early Warnings: the precautionary principle, 1896-2000*, Environmental Issue Rpt 22, European Envir. Agency.

Hines M.E., Horvat M., Faganelli J. (eds.) (2001), *Mercury in the Idrija Region and the Northern Adriatic, Portoroz workshop*, RMZ-Materials and Geoenvironment (Special Issue), 48, 1-259.

Holland H.D., Peterson U. (1995), *Living dangerously. The Earth, its resources and the environment*, Princeton University Press.

Horowitz, A. (1995) The use of suspended sediment and associate trace element in water quality studies, *Int. Ass. Hydrol. Sci. Special Publ.*, 4, pp. 58.

Horowitz, A.J., Elrick, K.A. and Smith, J.J. (2001), Annual suspended sediment and trace element fluxes in the Mississippi, Columbia, Colorado and Rio Grande drainage basins, *Hydrological Processes*, 15, 1169-1207.

ICPR (2001), *International Commission for the Protection of the Rhine*, Annual Reports, Coblentz.

Kabat, P., Claussen, M., Dirmeyer, P.A., Gash, J.H.C., Bravo de Guenni, L., Meybeck, M., Pielke, R.A., Vörösmarty, C.J., Hutjes, R.W.A. and Lütkemeier, S. (eds.) (2004), *Vegetation, Water, Humans and the Climate*, Springer.

Kimstach V., Meybeck M., Baroudy E. (eds.) (1998), *A Water Quality Assessment of the Former Soviet Union*, E&FN Spon, London, pp. 611.

Knoppers B., Ekau W., Figueiredo A.G. (1999), The coast and shelf of east and northeast Brazil and material transport, *Geo-Marine Letters*, 19, 171-178.

Lacerda L.D., Ribeiro M.G., Cardeiro R.C., Sifedine A., Turcq B. (1999), Atmospheric mercury deposition over Brazil during the past 30 000 years, *Ciencia e Cultura*, 51, 363-371.

Laraque A, Mahé G, Orange D, Marieu B (2001), Spatiotemporal variations in hydrological regimes within Central Africa during the XXth century, *J. Hydrol.*, 245,104-117

LeBlanc M., Morales J. A., Borrego J., Elbaz-Poulichet F. (2000), 4,500-year-old mining pollution in southwestern Spain: Long-term implications for modern mining pollution, *Economic Geology*, 95, 655-662.

Lemoalle, J. (2004), Lake Chad: a changing environment. In: J.C.J. Nihoul, P. Zavialov and P.P. Micklin, *Dying and Dead Seas*, Kluwer, NATO Series, pp. 321-340.

Ludwig W., J.L. Probst and S. Kempe (1996), Predicting the oceanic input of organic carbon by continental erosion, *Global Biogeochem. Cycles*, 10, 24-41.

Lundqvist J., Falkenmark M. (eds.) (2000), Towards hydrosolidarity. Focus on the upstream-downstream conflicts of interests, *Water Int.*, 25, 168-319.

Macklin M.G., Hudson-Edwards K.A., Dawson E.J. (1997), The significance of pollution from historic metal mining in the Pennine ore fields on river sediment contamination fluxes to the North Sea, *Sci. Total Environ.*, 194/195, 391-397.

Mariely D.H.B., Montebello A.A., Bernards M.C., Ometto J.P.H., De Camargo P.B., Krusche A.V., Ballester M.V., Victoria R., Martinelli L.A. (2002), Effects of urban sewage on dissolved oxygen, dissolved inorganic and organic carbon, and electrical conductivity of small streams along a gradient of urbanization in the Piracicaba River basin, *Water Air Soil Pollut.*, 136, 189-206.

Martens W.J.M. (1995), Climate change and vector-borne diseases: a global modeling perspective, *Global Env. Change*, 5, 195-209

Mc Cutcheon S.C., Martin J.L., Barnwell T.O. (1993), Water quality. In: *Handbook of Hydrology*, Maidment D.R. (ed.), Mac Graw Hill, 11.1-11.73.

McMichael, A.J. (ed.) (2003), *Climate change and human health. Risk and responses,* WHO, Geneva, pp. 37.

McMichael, A.J., A. Haines, Slooff, R., Kovats, S. (1996), *Climate change and human health,* WHO, Geneva, pp. 297.

Messerli B., Grosjean M., Hofer T., Nuñez, L., Pfister C. (2000), From nature-dominated to human-dominated environmental changes. In: *Past global changes and their significance for the future*, K.D. Alverson, F. Oldfield and R.S. Bradley (eds.), pp. 459-479, *Quat. Sci. Rev.*, 19, 1-5.

Meybeck, M., Carbon, nitrogen, and phosphorus transport by world rivers, *American Journal of Science*, 282, 401-450, 1982.

Meybeck M. (1998), The IGBP Water Group: a response to a growing global concern, *Global Change Newsletters*, 36, 8-13.

Meybeck M. (2002), Riverine quality at the Anthropocene: Propositions for global space and time analysis, illustrated by the Seine River, *Aquatic Sciences*, 64, 376-393.

Meybeck M. (2003a), Global occurrence of major elements in rivers. In: Holland HD and Turekian K.K. (eds.), *Treatise on Geochemistry, Volume 5, Surface and Ground Water, Weathering and Soils*, J.I. Drever (ed.), Elsevier, Pergamon, 207-224.

Meybeck M. (2003b), Global analysis of river systems: from earth system controls to Anthropocene controls, *Philosophical Transactions of the Royal Academy B*, 358, 1440, 1935-1955.

Meybeck M. (2005), Looking for water quality, *Hydrological Processes*, 19, 1, 331-338.

Meybeck M., Chapman D. and Helmer R. (eds.) (1989), *Global Fresh Water Quality: A First Assessment*, Basil Blackwell, Oxford, pp. 307.

Meybeck M., Helmer R. (1989), The quality of rivers: from pristine stage to global pollution, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* (Global and Planetary Change), 1, 283-309.

Meybeck M., Helmer R., Dray M., El Ghobary H., Demayo A., Ramadan S., Khan K., Peterson P.J., Jackson J. (1991), Water quality, progress in the implementation of the Mar del Plata Action Plan, WHO/UNEP, pp. 79 (*UN Conference on Water*, Dublin, 1992).

Meybeck M., Vörösmarty C.J. (2005), Fluvial filtering of land-to-ocean fluxes: from natural Holocene variations to Anthropocene, C.R. *Geosciences*, 337, 107-123.

Meybeck M., Vörösmarty C.J., Schultze R., Becker A. (2004), Scaling relative responses of terrestrial aquatic systems to global changes. In: P. Kabat *et al*. (eds.), *Vegetation, Water, humans and the Climate: a New Perspective on an Interactive System*, Springer, Berlin, 455-464.

Middelkoop H. (1997), Embanked floodplains in the Nederlands, *Nederlandsee Geografische Studies*, 224, Utrecht, pp. 342.

Pascual, M., Bouma, M.J. and Dobson, A.P. (2002), Cholera and climate: revisiting the quantitative evidence, *Microbes and Infection*, 4, 237-245.

Petts G. E. (1984), *Impounded rivers, perspectives for ecological management*, Chichester, UK: John Wiley.

Pouria, S., de Andrade, A., Barbosa, J., Cavalcanti, R.L., Barreto, V.T.S., Ward, C.J., Preiser, W., Poon, G.K., Neild, G.H., Codd, G.A. (1998), Fatal microcystis intoxication in a haemodialysis unit in Caruaru, Brazil, *Lancet*, 352, 21-26.

Rabalais N.N., Turner R.E. (2001), Coastal hypoxia, *Coastal Estuary Studies*, 52.

Rabouille C., Mackenzie F.T., Ver L.M. (2001), Influence of the human perturbation on carbon, nitrogen and oxygen biogeochemical cycles in the global coastal ocean, *Geochem. Cosmochim. Acta*, 65, 3613-3641.

Renberg I., Brännvall M.L., Brindler R., Emteryd O. (2000), Atmospheric lead pollution history during four millennia (2000 BC to AD 2000) in Sweden, *Ambio*, 29, 150-156.

Revenga C., Murray S., Abramovitz J., Hammond A. (1998), *Watersheds of the world: ecological value and vulnerability*, Washington, DC: World Resources Institute and Worldwatch Institute.

Rice, K. (1999), Trace element concentrations in streambed sediment across the conterminous United States, *Envir. Sci. Technol.*, 33, 2499-2504.

Robarts R.D., Fraser A.S., Hodgson K.M., Paquette J.M. (2002), Monitoring and assessing global water quality: the GEMS-WATER experience, *Ecohydrol. Hydrobiol.*, 2, 19-27.

Salomons, W., Förstner, U., Mader, P. (eds.) (1995), *Heavy metals. Problems and solutions*, Springer.

Sanderson E.W., Jaiteh M., Levy M.A., Redford K.H., Wannelo A.V., Woolmer G. (2002), The human footprint and the last of the wild, *Bioscience*, 52, 891-904.

Schlesinger W.H. (1997), *Biogeochemistry: an analysis of Global Change*, 2nd ed. Academic Press, San Diego.

Schwartz H.E., Emel J., Dickens W.J., Rogers P., Thompson J. (1990), Water Quality and flows. In: B. L. Turner *et al*. (ed.), The Earth as Transformed by human Action. Global and Regional Changes. In: *The Biosphere Over the Past 300 years*, Cambridge University Press, pp. 253-270.

Seitzinger S.P., Kroeze C., Bouwman A.F., Caraco N., Dentener F., Styles R.V. (2002), Global patterns of dissolved inorganic and particulate nitrogen inputs in coastal ecosystems: recent conditions and future projections, *Estuaries*, 25, 640-655.

Shotyk W., Weiss D., Appleby P., Cheburkin A., Frei R., Glor M., Kramers J., Reese S., Van de Knaap W. (1998), History of atmospheric lead deposition since 12370 14C yr BP from a peat bog, Jura Mountains, Switzerland, *Science*, 281, 1635-1940.

Steffen, W., and E. Lambin, Earth system functioning in the Anthropocene: Human impacts on the global environment. In: *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 112-144, Pontifical Academy of Sciences, Vatican City, 2006.

Steffen W., Sanderson A., Tyson P.D., Jäger J., Matson A., Moore B., Oldfield F., Richardson K., Schellnhuber H.J., Turner B.L., Wasson R.J. (eds.) (2004), *Global change and the Earth system: a planet under pressure*, Springer, Berlin.

Syvitski, J.P. (2005), Dynamics of the coastal zone. In: J.C.J. Crossland, *Coastal fluxes in the Anthropocene*, Springer, pp. 39-94.

Turner R.E., Rabalais N.N., Justic D., Dortch Q. (2003), Global patterns of dissolved N, P and Si in large rivers, *Biogeochem.*, 64, 297-317.

Turner R.K., Bateman I.J., Adger W.N. (eds.) (2001), *Economics of coastal and water resources: valuing environmental functions*, Kluwer, Dordrecht.

Turner R.K., Bower B.T. (1999), Principles and benefits of integrated coastal zone management (ICZM). In: Salomons W., Turner R.K.,

Lacerda L.D. and Ramachandra S. (eds.), *Perspectives on coastal zone management*, Springer, Berlin, 1-34.

Unesco (1992), *Water and Health*, Int. Hydrol. Progr., Tropics Programme Series 3, Paris, Unesco, pp. 47.

Valette-Silver, N.N.J. (1993), The use of sediment cores to reconstruct historical trends in contamination of estuarine and coastal sediments, *Estuaries*, 16, 577-588.

Van der Weijden C.H., Middleburg J.J. (1989), Hydrogeochemistry of the River Rhine: long term and seasonal variability, elemental budgets, base levels and pollution, *Water Research*, 23, 1247-1266.

von Bodungen B., Turner K. (eds.) (2000), *Science and Integrated Basin Management*, Dahlem workshop series, Wiley.

Vörösmarty C.J. (2002), Global Change, the water cycle, and our search for Mauna Loa, *Hydrological Processes*, 16, 135-139.

Vörösmarty C.J., Meybeck M. (2004), Responses of continental aquatic systems at the global scale: new paradigms, new methods. Chapter D4. In: Kabat *et al.*, *Vegetation, Water, humans and the Climate: a New Perspective on an Interactive System*, Springer Verlag IGBP Synthesis Series, 375-413.

Vörösmarty C.J., Meybeck M., Fekete B., Sharma K., Green P., Syvitski J. (2003), Anthropogenic sediment retention: major global-scale impact from the population of registered impoundments, *Global Planetary Changes*, 39, 169-190.

Vörösmarty C.J., Sharma K., Fekete B., Copeland A.H., Holden J., Marble J., Lough J.A., (1997), The storage and aging of continental runoff in large reservoir systems of the world, *Ambio*, 26, 210-219.

Vörösmarty C.J., *et al*. (2005), Fresh Water. In: *Ecosystems and human well-being. Volume 1: Current state and trends*, Hassan, R., Scholes, R., and Ash, N. (eds.), Island Press, Washington, D.C., pp. 165-207.

Walling, D.E., Fang, D. (2003), Recent trends in the suspended sediment loads of the world's rivers, *Global Planet. Change*, 39, 111-126.

Wang, Y., Ren, M.E. and Ouyang, Z. (2000), *China Water Vision. Foreign Language Book* no. 120, Beijing, China Meteorological Press.

WCMC (1998), *Freshwater biodiversity: a preliminary global assessment. Biodiversity series no. 8*, World Conservation Monitoring Centre, See www.unep-wcmc.org/information-services/publication/freshwater/1.htm.

Wilson M.L. (2001), Ecology and infections disease. In: *Ecosystem change and public health, a global perspective*, J.L. Aron & J.A. Patz (eds.), pp. 283-324, Baltimore, MD: Johns Hopkins University Press.

EARTH SYSTEM FUNCTIONING IN THE ANTHROPOCENE: HUMAN IMPACTS ON THE GLOBAL ENVIRONMENT

WILL STEFFEN & ERIC LAMBIN

Hurt not the earth, neither the sea, nor the trees.
Revelation 7:3, the Holy Bible

Most Gracious is Allah, Who reveals Himself
In the Qur'an, in man's Intelligence
And in the nature around man.
Balance and Justice, Goodness and Care,
Are the Laws of His Worlds...
Summary from *Surah* 55, the Holy Qur'an

Without the willow, how to know the beauty of the wind.
Lao She, Buddhist monk

We're only here for a short amount of time to do what we've been put here to do, which is to look after the country. We're only a tool in the cycle of things. ... (we) go out into the world and help keep the balance of nature. It's a big cycle of living with the land, and then eventually going back to it ...
Vilma Webb, Noongar People, Australian Aborigines, from:
Elders: Wisdom from Australia's Indigenous Leaders

1. *The Earth's Environment as a System*

Through the ages, humans have recognized two important features of the planet that we inhabit. First, the Earth is a single system that provides a hospitable environment for humans. Second, humans are an integral part of Earth itself, and not an outside force perturbing an otherwise pristine, natural world. Words and phrases such as 'balance', 'cycles', 'care ', 'hurt not the earth' and 'looking after the country' recognize the nature of the Earth System and important aspects of the human-environment relationship.

Modern science has added much detail and a quantitative underpinning to the notion that the Earth operates as a single, interlinked system. In the context of contemporary environmental change the term *Earth System* has come to mean the suite of interacting physical, chemical and biological global-scale cycles and energy fluxes that provide the conditions necessary for life on the planet (Oldfield and Steffen, 2004). Several features of this definition of the Earth System are important. First, the forcings and feedbacks *within* the Earth System are as important as the external drivers of change, such as variability in solar energy input. Second, biological/ecological processes are an integral part of the functioning of the Earth System and not merely the recipient of changes in physico-chemical systems. Finally, the functioning of the Earth System exhibits many modes of natural variability; human-driven changes interact with this natural variability in complex and sometimes mutually reinforcing ways.

The most well-known evidence for the behavior of the Earth as a single, interlinked system comes from the Antarctic ice core data. The Vostok ice core record (Petit *et al.*, 1999) provides convincing evidence for the systemic nature of the planetary environment, with properties and behavior that are characteristic of the system as a whole. In particular:

– The variation of climate, as represented by a proxy for local temperature (the oxygen isotope ratio, $\alpha^{18}O$), and of atmospheric composition, as represented by the concentration of the trace gases carbon dioxide (CO_2) and methane (CH_4) trapped in air bubbles in the ice, are closely coupled throughout the record (Figure 1, see page 401).

– The main maxima and minima of temperature and trace gas concentrations, which mark the alternation between glacial and interglacial conditions, follow a regular, cyclical pattern through time, each cycle spanning approximately 100,000 years. The smooth changes in the eccentricity of the Earth's orbit act as the pacemaker for the observed periodicity but internal mechanisms within the Earth System drive the changes in climate and atmospheric composition.

– The range, over which isotopically inferred temperature and trace gas concentrations vary, is tightly limited. Throughout all four cycles, each interglacial gives rise to similar peak values; each glacial culminates in comparable minima. This points to a high degree of self-regulation within the Earth System over the whole of the time interval recorded in the Vostok ice core.

The recent publication of the ice core data from the Dome Concordia site (EPICA community members, 2004) extends the Antarctic record

back to 740,000 years. The data from the later part of the core confirm the Vostok record, while the earlier data show some intriguing differences. For example, the interglacial periods were not as warm then, but the Earth spent more time in the warm mode. Also, the main mode of periodicity was not so clear in the earlier part of the record, having features of both the 40k yr periodicity of the pre-900k Quaternary period and the pronounced100k yr periodicity of the Vostok record. In terms of the tight coupling between greenhouse gas concentration and climate, and the clear upper and lower bounds on temperature and gas concentrations, the Vostok and Dome C data are in good agreement.

The amplification and modulation of the external forcing of solar variability by the complex array of feedbacks and forcings within the Earth System is crucial for maintaining Earth's life support system. For example, without the greenhouse gases that naturally occur in the lower atmosphere, the Earth's surface temperature would be about 30°C lower than it is now, and much more harmful ultraviolet radiation would penetrate to the Earth's surface without the layer of ozone in the stratosphere. However, there is now strong evidence that the growth in the numbers and activities of people – the burgeoning 'human enterprise' – has created a geophysical force of global scale that is beginning to interfere with the internal forcings and feedbacks within the Earth System (Andreae *et al.*, 2004).

2. *The Human Enterprise from an Earth System Perspective*

Human-driven environmental change has operated largely at the local scale for nearly all of human history. This situation changed with the Industrial Revolution in the late eighteenth century. The introduction of fossil fuel-based energy systems transformed economic systems, the structure of human societies and the human capacity to affect the planet. The new energy system increased our capacity to extract, consume, and produce (Grübler, 1998), triggering a sharp and sustained rise in the global population, from just under one billion people in 1800 to a projected nine billion by 2050 (UN Population Division 2000 and Lutz, this volume). In addition, changes in lifestyle and consumptions patterns in most societies around the world have led to a global escalation in both total and per capita demands for Earth's resources, including fish stocks, inert materials, freshwater, and prime agricultural soils (Dicken, 1992; Tolba and El-Kohly, 1992). The resulting impacts on the Earth System are both *cumulative* – changes that occur on a local scale but are so ubiquitous

around the planet that when aggregated have a global-scale effect (e.g., Turner *et al.*, 1990) – and *systemic*, such as changes in atmospheric composition and climate. Historically, cumulative changes have had more impact on the environment, but systemic changes are expected to accelerate through this century.

Figures 2 and 3 (Steffen *et al.*, 2004b) capture graphically the profound transformation of the human enterprise and its growing impact on the Earth System. The change in the human enterprise over the past few hundred years is shown in Figure 2. The temporal scale begins before the start of the Industrial Revolution and the spatial scale is global. Although this masks important regional differences in trends, from the perspective of the Earth System, global-scale indicators are appropriate. Figure 3 shows the impacts of the changing human enterprise on the structure and functioning of the Earth System.

These two sets of plots together represent what is often called *global change*. Figure 2 shows the human component of global change, while Figure 3 depicts the environmental component; the latter is often called *global environmental change*. In reality, these two components are strongly coupled and indeed are parts of the same system – the Earth System. A further important distinction is between environmental change at the global scale and regional and local changes to the environment. This distinction will be made clearer via specific examples later in the paper.

Several features of Figures 2 and 3 are noteworthy. First, nearly all of the indicators in Figure 2 either began to change around 1950, or changed their rate significantly at that time. Clearly the human enterprise underwent a profound reorganization and acceleration after the end of the Second World War. Many of these socio-economic changes are associated with the phenomenon of *globalization*. Second, the human imprint on the Earth System is discernable in every component of the planet – oceans, coastal zone, atmosphere, land. Third, the implications of this human imprint go far beyond the well-known effects on climate to impact on a wide range of Earth System functions, including the hydrological cycle, the cycles of important chemical elements and the dynamics of the marine and terrestrial biospheres. Global change is more than climate change.

Figures 2 and 3 suggest that the *rates* of the human-driven changes to the Earth System are as important as their *magnitudes*. Figure 4, showing the rise in atmospheric CO_2 concentration, is a good example of this comparison. The Vostok ice core data (Petit *et al.*, 1999) show that the CO_2 concentration has varied naturally in the past and has shown rapid

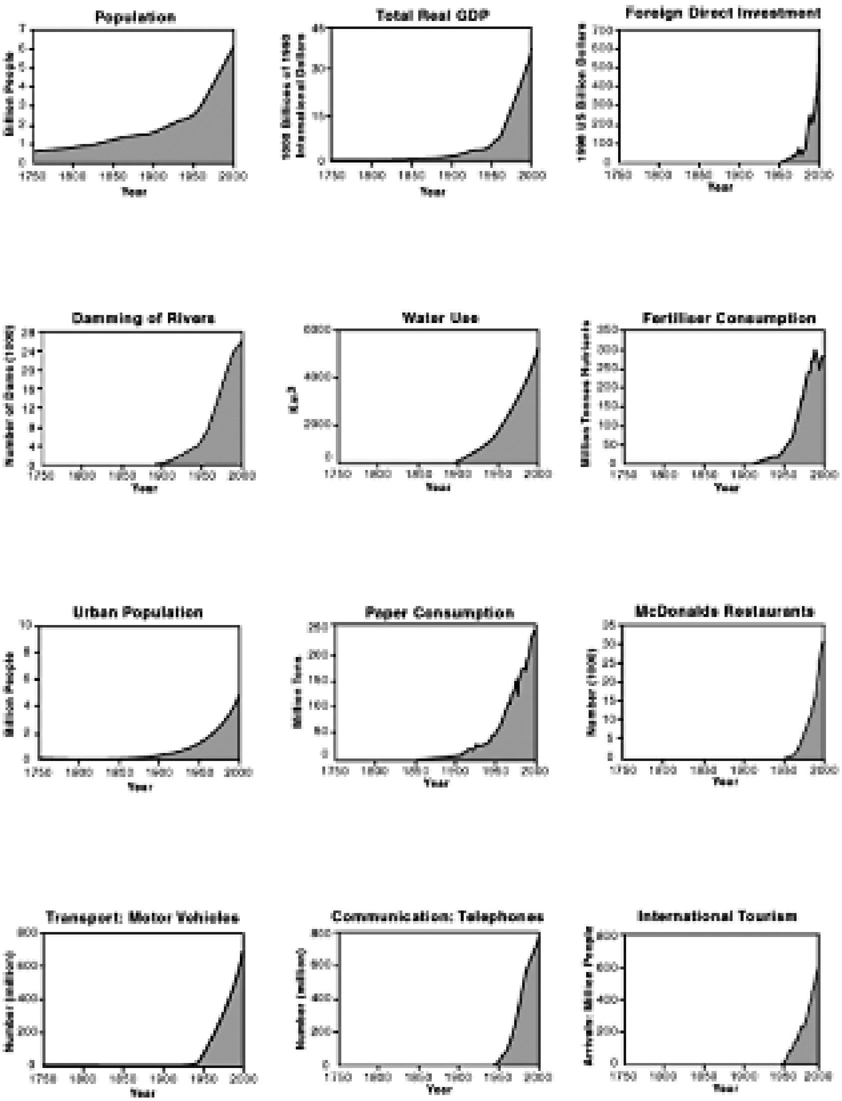


Fig. 2. The increasing rates of change in human activity since the beginning of the Industrial Revolution (Steffen *et al.*, 2004b). Significant increases in the rates of change occur around the 1950s in each case and illustrate how the past 50 years have been a period of dramatic and unprecedented change in human history. (US Bureau of the Census, 2000; Nordhaus, 1997; World Bank, 2002; World Commission on Dams, 2000; Shiklomanov, 1990; International Fertilizer Industry Association, 2002; UN Centre for Human Settlements, 2002; Pulp and Paper International, 1993; MacDonalds, 2002; UNEP, 2000; Canning, 2001; World Tourism Organization, 2002).

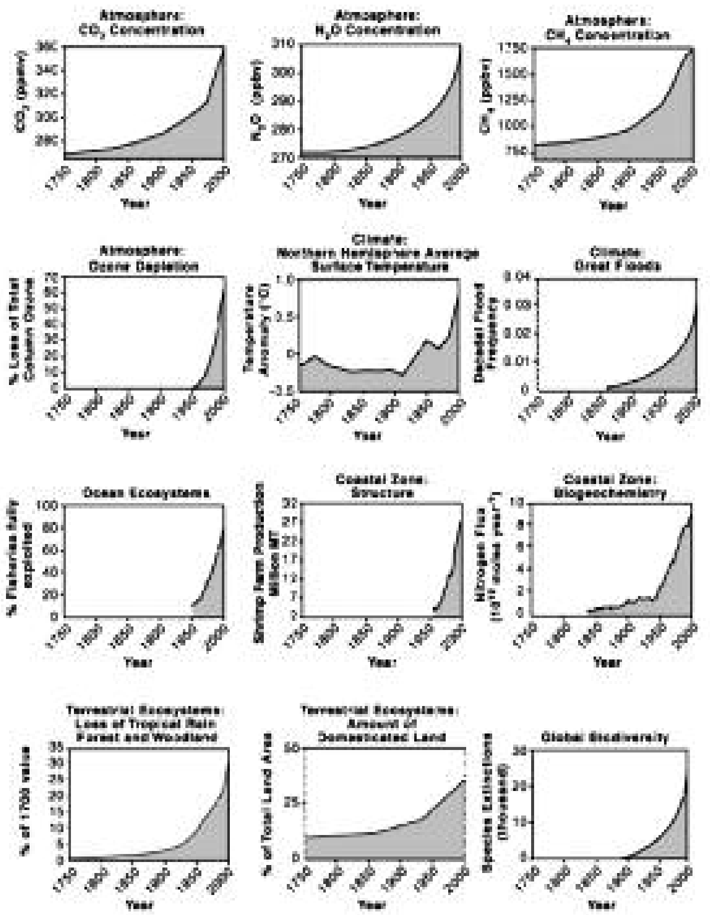


Fig. 3. Global-scale changes in the Earth System as a result of the dramatic increase in human activity (Steffen *et al.*, 2004b): (a) atmospheric CO_2 concentration (Etheridge *et al.*, 1996); (b) atmospheric N_2O concentration (Machida *et al.*, 1995); (c) atmospheric CH_4 concentration (Blunier *et al.*, 1993); (d) percentage total column ozone loss over Antarctica, using the average annual total column ozone, 330, as a base (Image: J.D. Shanklin, British Antarctic Survey); (e) northern hemisphere average surface temperature anomalies (Mann *et al.*, 1999); (f) decadal frequency of great floods (one-in-100-year events) after 1860 for basins larger than 200,000 km^2 with observations that span at least 30 years (Milly *et al.*, 2002); (g) percentage of global fisheries either fully exploited, overfished or collapsed (FAOSTAT, 2002); (h) annual shrimp production as a proxy for coastal zone alteration (WRI, 2003; FAOSTAT, 2002); (i) model-calculated partitioning of the human-induced nitrogen perturbation fluxes in the global coastal margin for the period since 1850 (Mackenzie *et al.*, 2002); (j) loss of tropical rainforest and woodland, as estimated for tropical Africa, Latin America and South and Southeast Asia (Richards, 1990; WRI, 1990); (k) amount of land converted to pasture and cropland (Klein Goldewijk and Battjes, 1997); and (l) mathematically calculated rate of extinction (based on Wilson, 1992).

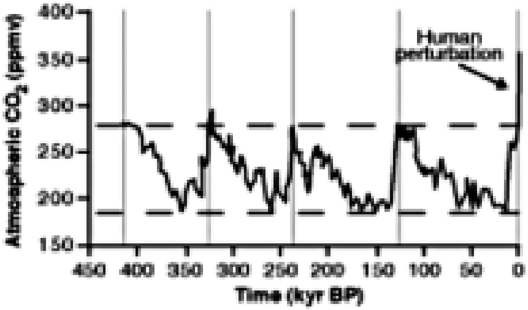


Fig. 4. Atmospheric CO_2 concentration over past 420,000 years from the Vostok ice core with the recent human perturbation superimposed (Petit *et al.*, 1999; Keeling and Whorf 2000).

changes in concentration on the geological timescale. However, within the limits of resolution of the ice-core records, the current concentration of 380 ppm appears to have been reached at a rate at least 10 and possibly 100 times faster than increases of CO_2 concentration at any other time during the previous 420,000 years (Falkowski *et al.*, 2000). Many other human-driven changes to the Earth System show similarly rapid changes compared to the pace of natural variability.

The remainder of this paper focuses on the extent of human impacts on the global environment over the past few centuries. We focus first on the transformation of the land surface; this is critically important for human well-being, as it is our primary source of food and provides the immediate environment in which we, as terrestrial creatures, must live. We then summarize the human impacts on the Earth System's two great fluids – the atmosphere and the ocean – before considering human-driven changes to climate and the hydrological cycle. We also consider the possibility of abrupt changes in major aspects of Earth System functioning, changes that are rapid on both geological and human timeframes and that could have catastrophic impacts on human societies.

3. *Transformation of the Land Surface*

Ever since humans have controlled fire and domesticated plants and animals, they have cleared forests to wring higher value from the land. About half of the ice-free land surface has been converted or substantial-

ly modified by human activities over the last 10,000 years. Undisturbed (or wilderness) areas represent 46% of the earth's land surface (Mittermeier *et al.*, 2003). Forests covered about 50% of the earth's land area 8000 years ago, as opposed to 30% today (Ball, 2001).

Concerns about transformations of the land surface emerged in the research agenda on global environmental change several decades ago with the realisation that land surface processes influence climate. Land-cover change modifies surface albedo and thus surface-atmosphere energy exchanges (Otterman, 1974; Charney and Stone, 1975; Sagan *et al.*, 1979); terrestrial ecosystems are sources and sinks of carbon and thus affect the global climate via the carbon cycle (Woodwell *et al.*, 1983; Houghton *et al.*, 1985; Ruddiman and Carmichael, this volume); and land cover controls the contribution of local evapotranspiration to the water cycle (Eltahir and Bras, 1996). A much broader range of impacts of land-use/cover change on ecosystem goods and services have now been identified. Of primary concern are impacts on biotic diversity worldwide (Sala *et al.*, 2000), soil degradation (Trimble and Crosson, 2000), and the ability of biological systems to support human needs (Vitousek *et al.*, 1997). Land-use/cover changes also determine, in part, the vulnerability of places and people to climatic, economic, or sociopolitical perturbations (Kasperson *et al.*, 1995). When aggregated globally, land-use/cover changes significantly affect central aspects of Earth System functioning, and thus are truly components of global environmental change, in addition to their local and regional impacts. All impacts are not negative, however, as many forms of land-use/cover changes are associated with continuing increases in food and fiber production, in resource use efficiency, and in wealth and well-being.

The area of cropland has increased globally from an estimated 300-400 million ha in 1700 to 1500-1800 million ha in 1990 (Ramankutty and Foley, 1999; Goldewijik, 2001), a nearly five-fold increase in three centuries and a 50% net increase just in the 20th century. The area under pasture, the estimates for which are more uncertain than for cropland, increased from around 500 million ha in 1700 to around 3100 million ha in 1990 (Goldewijk and Ramankutty, 2003). Forest area decreased from 5000-6200 million ha in 1700 to 4300-5300 million ha in 1990. Steppes, savannas, and grasslands also experienced a rapid decline, from around 3200 million ha in 1700 to 1800-2700 million ha in 1990 (Ramankutty and Foley, 1999; Goldewijk, 2001).

The Global Forest Resources Assessment 2000 (FAO, 2001b) estimated that the world's natural forests decreased by 16.1 million hectares per year

on average during the 1990s; that is a loss of 4.2% of the natural forest that existed in 1990. However, some natural forests were converted to forest plantations. Gains in forest cover arose from afforestation on land previously under nonforest land use (1.6 million hectares per year globally) and the expansion of natural forests in areas previously under agriculture, mostly in western Europe and eastern North America (3.6 million hectares per year globally). The net global decrease in forest area was therefore 9.4 million hectares per year from 1990 to 2000 (FAO, 2001b). The total net forest change for the temperate regions was positive, but it was negative for the tropical regions. Between 1990 and 1997, 5.8 ± 1.4 million hectares of humid tropical forest were lost each year (Achard *et al.*, 2002). Forest regrowth accounted for 1.0 ± 0.32 million hectares. The annual rate of net cover change in humid tropical forest was 0.43% during that period. A further 2.3 ± 0.7 million hectares of forest were visibly degraded. This figure does not include forests affected by selective logging.

The driving forces for tropical deforestation vary between the regions (Figure 5, see page 402). In Latin America, large-scale forest conversion and colonisation for livestock-based agriculture is prevalent, whereas cropland expansion by smallholders dominates in Africa. In Asia, intensified shifting agriculture, including migration into new areas, gradual change of existing areas toward more permanent agriculture, and logging explain most of the deforestation (FAO, 2001b; Achard *et al.*, 2002; Geist and Lambin, 2002). Within these regions, deforestation is largely confined to a few areas undergoing rapid change, with annual rates of deforestation from 2% to 5% (Figure 6, see page 402).

Historically, humans have increased agricultural output mainly by bringing more land into production. However, the amount of suitable land now remaining for crops is very limited in most developing countries (Young, 1999; Döös, 2002), where most of the growing food demand originates. Where there is a large surplus of cultivable land, land is often under rain forest or in marginal areas. The period after 1960 has witnessed a decoupling between food production increase and cropland expansion. The 1.97-fold increase in world food production from 1961 to 1996 was associated with only a 10% increase of land under cultivation, but also with a 1.68-fold increase in the amount of irrigated cropland, and a 6.87- and 3.48-fold increase in the global annual rate of nitrogen and phosphorus fertilisation (Tilman, 1999). In 2000, 271 million ha were irrigated (FAO, 2001a). Globally, the cropland area per capita decreased by more than half in the 20th century, from around 0.75 ha per person in

1900 to only 0.35 hectare per person in 1990 (Ramankutty *et al.*, 2002). The mix of cropland expansion and agricultural intensification has varied geographically (FAO 2001a).

In 2000, towns and cities sheltered more than 2.9 billion people, nearly half of the world population (UN Population Division, 2002). Urban population has been growing more rapidly than rural population worldwide, particularly in developing countries. Urban form and function have also changed rapidly. Built-up or paved-over areas are roughly estimated to occupy from 2% to 3% of the Earth's land surface (Young, 1999; Grübler, 1994). Urbanisation affects land in rural areas through the *ecological footprint* of cities, that is, consumption of prime agricultural land in peri-urban areas for residential, infrastructure, and amenity uses.

Other forms of rapid land-cover change that are thought to be widespread are still poorly documented at the global scale. Local- to national-scale studies, however, demonstrate their importance and ecological significance. Prominent among these are changes in the (sub)tropical dry forests; forest-cover changes caused by selective logging, fires, and insect damage; drainage or other forms of alteration of wetlands; soil degradation in croplands; changes in the extent and productive capacity of pastoral lands; and dryland degradation, also referred to as desertification.

Land-use change is always caused by multiple interacting factors originating from different levels of organisation of the coupled human-environment systems. The mix of driving forces of land-use change varies in time and space, according to specific human-environment conditions (Lambin *et al.*, 2001). Land-use change is driven by a combination of the following fundamental high-level causes (Lambin *et al.*, 2003): (i) resource scarcity leading to an increase in the pressure of production on resources; (ii) changing opportunities created by markets; (iii) outside policy intervention; (iv) loss of adaptive capacity and increased vulnerability; and (v) changes in social organisation, in resource access, and in attitudes. These changes are the product of multiple decisions resulting from interactions between diverse agents, who act under certain conditions, anticipate future outcomes of their decisions, and adapt their behaviours to changes in external (e.g., the market) and internal (e.g., their aspirations) conditions (Lambin *et al.*, 2003). Climate-driven land-cover modifications do interact with land-use changes. Slow and localized land-cover conversion takes place against a background of high temporal frequency regional-scale fluctuations in land-cover conditions caused by climatic variability, and it is often linked through positive feedback with land-cover modifications.

4. *Human Impacts on the Atmosphere and Ocean*

While the human imprint on the terrestrial surface of the planet is relatively easy to discern, the evidence for human impact on the two great fluids of Earth – the atmosphere and the ocean – is harder to visualize but nevertheless real and important. The human imprint on the atmosphere began millennia ago; there was clear evidence of enhanced atmospheric concentrations of lead, copper and other trace metals by the time of the Greek and Roman Empires, as shown by ice core data from Greenland (Hong *et al.*, 1994) and in lake sediments (Renberg *et al.*, 1994) and peats (Lee and Tallis, 1973; Livett *et al.*, 1979; Shotyk *et al.*, 1996, 1998; Martínez-Cortizas *et al.*, 1999) from Europe. Since then, the human imprint on the atmosphere has remained measurable and has grown strongly in the last two centuries.

The best known of the human impacts on atmospheric composition concerns the concentration of greenhouse gases (IPCC, 2001). Compared to their pre-industrial upper limits of about 280 ppm, as shown in the Vostok (Fig. 1) and Dome C ice cores (Petit *et al.*, 1999; EPICA TEAM, 2004), the concentration of CO_2 has now increased to nearly 380 ppm. This represents a doubling of the entire operating range of CO_2 between glacial and interglacial states. The increase in CH_4 concentration has been even more dramatic and rapid, with the current concentration of 1700 ppb nearly three times that of the pre-industrial era (670 ppb). A third major greenhouse gas, N_2O, has risen in concentration more modestly, from 285 to more than 310 ppb. For all three of these gases, the evidence is very strong that human activities are responsible for the observed increases (IPCC, 2001).

Less well known, but perhaps as important as the change in greenhouse gas concentrations, has been the sharp rise of aerosol particles in the atmosphere (Andreae *et al.*, 1995, 2005). Aerosols are defined as atmospheric mixtures containing liquid or solid particulates of various sizes and compositions suspended in carrier gases. The most studied of the aerosol particles is sulphate, which began to increase in the North Atlantic region from the start of the 20th century (Mayewski *et al.*, 1990; Legrand *et al.*, 1997). Since the 1960s, however, concern about local and regional air pollution has led to measures to curb sulphate emissions and atmospheric concentrations of sulphate in the region have steadily decreased from the 1980s. At the same time, the concentration of sulphate particles in the atmosphere in Southeast, South and Temperate

East Asia are increasing steadily (Fu *et al.*, 2002). Globally, anthropogenic sources of oxidized compounds of sulphur are more than double the natural sources, indicative of the large human influence on the atmosphere's aerosol particle burden.

Carbonaceous particles (soot) from the combustion of fossil fuels and from biomass burning are other sources of aerosol particles in the atmosphere (Kanakidou *et al.*, 2005). Although biomass burning is a natural part of the dynamics of some ecosystems, extensive deforestation and the use of fuel wood by humans has led to an estimated 30-50% increase over the last century in aerosol production from fires (Scholes *et al.*, 2003). Globally, the amount of soot in the atmosphere due to human activities has increased by about 10-fold over natural sources (Brasseur *et al.*, 2003). Mineral dust arising from arid regions is a third important type of aerosol particle. Like carbonaceous particles from burning, dust particles also arise to some extent from natural sources. However, due to human-driven land-cover change, the atmospheric loading of dust has increased over the last couple of centuries, perhaps by as much as 30-50% (Heintzenberg *et al.*, 2003). Global distributions of aerosol particle loading as characterized by the aerosol optical depth are shown in Figure 7 (see page 403).

Aerosols are a good example of an environmental problem that has grown in scale from local and sometimes regional to one that is now undeniably global. The rapid geographical expansion of anthropogenic sources of aerosols coupled with intercontinental atmospheric transport have generated global distributions of many important aerosol particles. This rapid expansion of aerosol particles has impacts on air quality in places far removed from their source and is also significant for climate, both directly through their radiative properties, and indirectly through their effects on cloud physics and their interaction with the hydrological cycle.

The amount of oxidizing gases, often called photo-oxidants, in the atmosphere has also increased due to human activities. The most important of these is ozone, which is harmful in the troposphere, in contrast to its role in the stratosphere (Crutzen, 1995). Tropospheric ozone is not produced directly by human activities, but rather is the result of chemical reactions involving precursors that are produced by human activities. The degree of increase in tropospheric ozone is difficult to determine, but measurements at mountain sites in Europe suggest that it has increased by a factor of four or so during the last century (WMO, 1999). As for aerosols, the increase in emissions, coupled with atmospheric transport,

has elevated oxidizing gases from a local air pollution problem to a global change issue.

One of the most prominent changes in atmospheric composition observed over the past decades has been the decrease in stratospheric ozone over the southern high latitudes owing to the emission of chlorine- and bromine-containing gases of industrial origin (primarily the so-called chlorofluorocarbons). Peak ozone loss has approached 60% at the South Pole – the so-called 'ozone hole'; ozone loss in the stratosphere over the mid-latitudes, although much less, is still measurable. Total column ozone during the period 1997-2001 was 3% less than the pre-1980 value for the northern mid-latitudes and 6% less for the southern mid-latitudes (UNEP/WMO, 2002).

The human imprint on the marine environment is most strongly expressed in the coastal zone, as more than 50% of the human population lives with 100 km of a coast (Kremer and Crossland, 2002) and even more use the coastal zone for recreational activities. Humans alter the coastal zone most directly via geomorphological changes through the construction of shoreline engineering structures, ports and urban developments. Other direct impacts include the conversion of natural ecosystems to managed systems designed for food production, for example, the conversion of mangrove forests to prawn farms. Globally, approximately 50% of mangrove systems have been converted to other uses as a result of human activities (Kelleher *et al.*, 1995; WRI, 1996, 2000; Naylor *et al.*, 2000). Coastal ecosystems have also been modified by the direct or inadvertent introduction of non-indigenous species, which can alter the structure and functioning of coastal ecosystems.

One of the most important roles of the coastal zone in terms of Earth System functioning is the transport and transformation of materials from the land to the ocean. These include the flow of water itself, sediments suspended in the water and a large range of chemical species dissolved in the water. Human activities have altered all of these functions.

Perhaps the most profound change has been the flow of water itself, described in more detail in the next section on changes in the hydrosphere (see also Meybeck, this volume). The flow of suspended material from upland areas is important for maintaining the structure of the coastal zone as sediments provide the building material for river deltas and for coastal geomorphology more generally. The overall impact of human activities on the delivery of sediments to the coastal zone is difficult to determine due to two opposing effects. In many regions, delivery

of sediment to the coastal zone has been decreased through sediment trapping within reservoirs and other pondages upstream. In other areas there have been regional increases in the delivery of sediments to the coastal zone through increased soil erosion driven by construction, mining, forestry and agriculture (Steffen *et al.*, 2004b).

Changes in nutrient flows through the coastal zone are more clear-cut (Andreae *et al.*, 2004). The sharp increase in the delivery of nitrogen to the world's oceans via the coastal zone is typical of the increased flow of nutrients more generally from land to ocean. A synthesis of the current understanding estimates that nitrogen delivery via rivers entering the North Atlantic has increased by a factor of between 3 and 20 (Howarth *et al.*, 1996). Although some of this increased nitrogen flux is due to direct human injection in the coastal zone from urban sewage treatment plants and from industrial activities, much of the nitrogen is derived from fertilizer application to agricultural areas upstream. Additional nitrogen is delivered via atmospheric deposition of NO_x originating from agricultural, industrial or transport activities (Jaworski *et al.*, 1997; Howarth *et al.*, 1996). Phosphorus delivery to the ocean has similarly increased in the last 50 years or so. In addition to nutrients, contaminants such as heavy metals, persistent organic pollutants, various other synthetic chemicals, radioactive materials, bacteria and slowly degrading solid waste like plastics are transported from land to the coastal regions.

The primary direct human impact on marine ecosystems is through fisheries. The total fish catch globally increased steadily through the 20th century but has leveled off during the last decade (Figure 8), probably representing the upper limit that can be attained. Among the major marine fish stocks or groups of stocks for which information is available, about 47-50% of stocks are fully exploited, 15-18% are overexploited and 9-10% have been depleted or are recovering from depletion (FAO, 2000). In terms of impacts of this fish harvest further down the food chain, one recent study estimates that humans ultimately harvest 8% of the primary production of the oceans, with much greater percentages for the upwelling and continental shelf areas (Pauly and Christensen, 1995). Human fisheries cause other changes to marine food webs. For example, commercial fisheries probably influence population characteristics of species incidentally caught in the fishery, as an equivalent of 25% of the annual production of marine fisheries is discarded as bycatch each year (FAO, 2000).

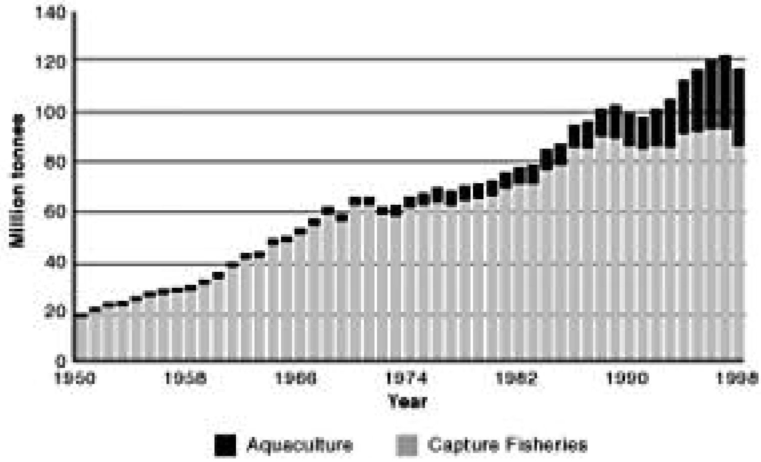


Fig. 8. World capture fisheries and aquaculture production (FAO, 2000).

5. *Changing Hydrosphere and Climate*

Given the central importance of water for human well-being, it is not surprising that the hydrological cycle has been heavily modified by human activities (Meybeck, this volume). The widespread rise of agriculture around 4,000 or 5,000 years ago triggered the start of major human influence on the hydrological cycle through the first development of water engineering works.

More recently, humans have come to dominate continental aquatic systems through a strong increase of activity during the past century. During that period, reservoirs and flow diversions have become both massive in size and widespread around the globe (Figure 9, see page 405). About 45,000 dams over 15 m high are registered in the world and many hundreds of thousands of smaller dams have been built on farms and on rivers. At present about 40% of the total global runoff to the oceans is intercepted by large dams, that is, about 16,000 km^3 yr^{-1} of discharge. Approximately 70% of this discharge flowing through large reservoirs experiences a sediment trapping of 50% or more (Vörösmarty *et al.*, 2003). In the northern hemisphere only 23% of the flow in 139 of the

largest rivers is unaffected by reservoirs in one way or another (Dynesius and Nilsson, 1994).

Now that much of the easily accessible surface water has been captured and diverted for human use, the extraction of groundwater is increasing sharply. In many cases, the current extraction of deep groundwater amounts to mining of the resource. For example, in the Chad Basin, Nigeria, the groundwater reservoir has not been fully recharged for thousands of years, with only shallow water recharge near the margins. A similar situation exists in arid areas of the Middle East, where natural recharge of the groundwater is also close to zero so that pumping of groundwater for irrigation of agriculture is equivalent to the mining of fossil water. Extraction of groundwater from the large Ogallala aquifer in the Great Plains area of the United States for agricultural use has led to decreasing water levels in the aquifer and to abandonment of agricultural land (Steffen *et al.*, 2004b).

Water quality has also changed significantly over the last few centuries, but the types of pollutants vary widely in the different regions of the world (Meybeck and Vörösmarty, 1999; Vörösmarty and Meybeck, 2003; Meybeck, this volume). For example, in the developed world agricultural runoff is probably the largest source of pollutants – sediments, nutrients, pesticides and coliform bacteria. According to the US Environmental Protection Agency (USEPA, 1989), 76% of pollution entering rivers and lakes in the United States is from non-point sources, of which agriculture contributes 64%. Most of the chemical pollutants are nitrogen and phosphorus compounds derived from fertilizers, and halo-organic compounds arising from pesticides. In the developing world, much of the water quality problem centers on faecal and organic pollutants from untreated human wastewater, a problem that was eliminated in developed countries a few decades ago. Loading of industrial and urban pollutants on freshwater resources continues to be a major problem in many parts of the world.

Climate change is the most well-known aspect of global change, and is often confused with global change itself. However, at the beginning of the 21st century, climate change probably still has less impact on human societies than the range of other global changes – direct changes to the land surface, atmosphere, coastal zone, and hydrological cycle – described herein. In addition, the recent Millennium Ecosystem Assessment has documented the extent to which human activities have modified terrestrial, coastal and marine ecosystems, and thus affected

their ability to provide ecosystem services to support human development and well-being (Reid *et al.*, 2005).

As this century progresses, climate change will almost surely become more important as the climate moves further from its 'operating range' of the last few millennia (IPCC, 2001) and the impacts on a wide range of natural and managed ecosystems become more apparent with increasing consequences of various types (Hare, 2005). Furthermore, the rate of land-cover change, particularly tropical deforestation in South America and Southeast Asia, is projected to slow through this century under most scenarios (Reid *et al.*, 2005). Thus, climate change will likely become relatively more important by mid-century compared to other global changes, and may dominate the other changes in the second half of the century.

The debate on whether climate change is 'real' or not is over; the vast majority of global change scientists accept that the climate is now moving beyond the bounds of natural variability and that human activities play a significant role in observed contemporary climate change (IPCC, 2001). The best-known feature of climate change is the rise in global mean temperature (Figure 10, see page 406), largely due to an increase in the nighttime minimum temperatures rather than an increase in maximum temperature. In terms of geographical variability, the high latitudes of the northern hemisphere and the interior of continents more generally are warming at a faster rate than the rest of the Earth's surface. The evidence for the general warming trend is also clearly discernible in the cryosphere, in both the retreat of most land glaciers around the world and the observed degradation of permafrost (Anisimov, 2004; ACIA, 2004). The response of the terrestrial biosphere to the warming trend is now also unmistakable (Parmesan and Yohe, 2003).

The magnitude and patterns of precipitation also appear to be changing beyond natural variability. For the mid and high latitudes in the northern hemisphere, for example, it is very likely that precipitation increased during the 20th century by 5-10% in many areas (IPCC, 2001). This was predicted, as increasing temperature leads to increasing evaporation from the land surface and a more active hydrological cycle in general. However, there is some debate about whether evaporation is actually increasing, as the preferential increase in nighttime minimum temperatures would not increase evaporation (Roderick and Farquhar, 2004), and in some regions of the world the cooling effect of increased aerosol loading in the atmosphere may dampen evaporation from the land surface (Ramanathan *et al.*, 2001, 2005). In addition, there are some count-

er-trends in precipitation change, with rainfall decreasing by about 3% on average over much of the sub-tropical land areas (IPCC, 2001).

Extreme events like floods and droughts are particularly important aspects of climate for human well-being. Although there is inconclusive evidence that the incidence of drought has increased globally, there is some evidence that climate change is increasing the number of severe floods around the world. The analysis by Milly *et al.*, (2002), which examines 100-year floods (river discharge that has a probability of 0.01 of being exceeded in any given year) for basins larger than 200,000 km^2 with observations that span at least 30 years, shows that the frequency of extreme flood events is increasing. Only half of the observational record of 2066 station-years was made after 1953, but 16 of the 21 extreme floods occurred after 1953. The observed increase in extreme floods is consistent with the projections of climate models, which suggests that the increase in frequency of extreme floods will continue into the future.

6. *Abrupt Changes in the Earth System*

The palaeo record shows that abrupt changes of various types are a common feature of the behavior of the Earth System, and closer examination of processes across many facets of the Earth System in the contemporary period shows that nonlinear behavior occurs frequently (Rial *et al.*, 2004). The term 'abrupt change', the most common term used to describe nonlinearities, usually refers to changes in major features of Earth System functioning that occur at an unexpectedly rapid rate (Steffen *et al.*, 2004a). The term 'unexpectedly rapid' depends on the perception and time scale considered, human or geological. Here we consider those changes that would be considered abrupt on a geological time scale, but would also be so rapid that they would be discernible within a human lifetime, thus would occur within a decade or, at most, a few decades.

The sudden, unexpected formation of the stratospheric ozone hole over Antarctica, mentioned briefly in Section 4, is an example of an abrupt change that has already occurred (Crutzen, 2004). The rapid and unexpected loss of ozone over the southern high latitudes was first recognized in the mid-1980s through measurements taken over decades by the British Antarctic Survey. The cause of the ozone hole was traced to man-made compounds called chlorofluorohydrocarbons (CFCs), produced as refrigerants and insulators and thought to be completely harmless. Under cold conditions in the stratosphere, the CFCs undergo a com-

plex series of reactions, eventually producing Cl_2, which is photolyzed by sunlight to create Cl radicals. These start a catalytic chain of reactions, leading to the destruction of ozone. The ozone hole is an example of a chemical instability in the Earth System, where a reaction sequence triggered by synthetic compounds caused an abrupt shift in ozone concentration in a large region of the stratosphere.

Instabilities in ocean circulation and regional climate are now attracting much attention as a type of abrupt change that could severely impact on modern civilizations. The focus is on the North Atlantic region; in this area the Gulf Stream transports much heat from the tropical Atlantic Ocean northwards, where the water cools and releases heat that is then delivered by the westerly winds to Northern Europe and Scandinavia. Much of the abrupt change seen in the Greenland ice core records (Bond *et al.*, 1999) is associated with reorganizations of this pattern of oceanic circulation (Rühlemann *et al.*, 1999). The abrupt climate shifts observed in the Greenland ice cores extend well beyond the North Atlantic region, as shown by a range of other synchronous environmental changes observed around the northern hemisphere, indicating that the abrupt changes are at least hemispheric and possibly global in extent (Alverson *et al.*, 2003). This evidence suggests a bipolarity in the states of the Earth System, where the North Atlantic ocean circulation appears to flicker or flip-flop between the two states, triggering extremely rapid and large temperature swings of up to 10°C in a decade (Alley *et al.*, 2001). Such changes, if they occurred now, would have severe consequences for the modern civilizations of the North Atlantic region. Studies show that a weakening of the circulation is possible, and abrupt and irreversible changes in ocean circulation cannot be excluded within the range of projected climate change over this century (Rahmstorf and Stocker, 2004).

The world's coral reefs are undergoing changes now that are the result of both local pressures and systemic global changes. There is a high probability that the combination of such forcings will cause most reefs to cross a threshold and convert them to algal beds sometime in the second half of this century, if not sooner. The global pressures are due to both changes in ocean chemistry and sea surface temperature. Increasing atmospheric CO_2 leads to increased dissolution of CO_2 in the upper ocean layers, changing the acidity of the ocean water and thus the ability of reef organisms to create their calcium carbonate shells (Orr *et al.*, 2005). A doubling of atmospheric CO_2 is estimated to lead to a 30% reduction in calcification rate of reef organisms (Kleypas *et al.*, 1999). The warming of

the ocean's surface waters adds another stress on the reefs; the exceptionally warm year of 1998 triggered widespread bleaching of reefs around the world (WRI, 2001). These global-scale forcings affect all reefs, whether they are pristine or under local pressure from human activities (e.g., fishing, tourism, nutrient loading from agriculture on adjacent lands), but where the local and global pressures act together on reefs, they are particularly susceptible to bleaching events and a conversion to algal beds. It is estimated that 58% of the world's coral reefs are currently at medium to high risk of degradation and possible conversion to algal beds as result of changing environmental conditions (Burke *et al.*, 1998), and this proportion will likely increase through this century.

Abrupt changes in human systems are rarely considered in global change studies, but they may be more common than many think and, indeed, could be triggered by gradual changes in the biophysical world. Examples of potential changes in human systems include instabilities caused by large and growing inequalities both within countries and between countries exacerbated by global change; instabilities in the human health system, including pandemics, re-emergence of old diseases and emergence of new diseases; and the possibility that gradual changes in the biophysical world could trigger abrupt changes in the globalizing economic system.

A summary of the potential biophysical part of the Earth System is given visually in the global map of 'switch and choke points' (Schellnhuber, 2002; Figure 11, see page 406). The map is an attempt to identify those areas where change – often abrupt change – at the regional level may lead to significant changes in the way that the Earth System as a whole operates.

7. *Human Health in a Changing Earth System*

The previous sections have described how human activities are affecting the functioning of the Earth System in many different ways. It is clear that the Earth is currently operating in a no-analogue state. In terms of key environmental parameters, the Earth System has recently moved well outside the range of natural variability exhibited over at least the last half million years. The *nature* of changes now occurring *simultaneously* in the Earth System, their *magnitudes* and *rates of change* are unprecedented (Steffen *et al.*, 2004b). The implications of the changing Earth System are complex and potentially serious.

Direct impacts of changing climate and atmospheric composition are probably the most obvious effects of global change on health. The heat

wave of 2003 in southern France that killed over 15,000 people is an example of such an impact. Less well-known is an equally severe heat wave in Australia in February 2004. Although Australian society is better equipped than French society to deal with high temperatures, the heat wave nevertheless caused numerous collapses due to heat stress in Adelaide and Sydney as well as a surge in ambulance call-outs in Brisbane, an event described by the Queensland commissioner of ambulance services as '... the most significant medical emergency in the south-east corner (of Queensland) on record'. (Steffen *et al*., 2005).

Deterioration of air quality due to emissions of a variety of deleterious compounds is another. Originally considered to be a local or regional problem, the long-range transport and transformation of chemical species has now become so widespread that pollution from North America affects Europe, while that from Europe affects Asia and so on. Air pollution has become a global change issue (Brasseur *et al*., 2003).

In addition to direct impacts, global change affects human health via ecological changes that can affect the transmission of diseases (McMichael, this volume; Confalonieri *et al*., this volume). One consequence of changing land use, for example, is the change in the prevalence of infectious disease. Vector-borne diseases are particularly sensitive to changes in land cover and land use, because their spatial distribution is restricted by the geographical range of the vector and by its habitat preferences. Research is needed to better understand correlations between land use, specific vector species and disease patterns. Land use change and vector ecology control the interactions between hosts and vectors, given the use of different land parcels by people, the breeding habitats of specific vectors and their dispersal through the landscape (influenced by landscape pattern and heterogeneity). The impact of land use change on vector-borne disease risks can modify the course of land transformation via a feedback mechanism affecting human decisions on land use. By predicting the effects that changes in land use would have on vector-borne disease, public health funds could be allocated more effectively in the prevention and treatment of these diseases. Vector-borne disease is often most effectively combated through vector control measures but, to be effective, the behavior and ecology of the target species must be fully understood.

Whatever biophysically driven impacts of global change on human health can be postulated, it is clear that the differing vulnerabilities of countries or sectors of society will be often be the decisive factor in determining whether a serious infectious disease pandemic, for example, breaks out or

not. Therefore, indirect changes in the human component of the Earth System may have the most critical effects on health. Deterioration of public health systems due to an increased need of society to cope with the more direct impacts of global change could ironically leave populations more vulnerable to disease. Populations in the developing world weakened by poor nutrition or poor water quality will be also be more vulnerable to health problems as global change accelerates (Shah *et al.*, this volume). Changes in the biophysical component of the Earth System could promote the re-emergence of old diseases or the development of new diseases. As in many aspects of global change, we should expect surprises as the Earth System is subjected to a suite of increasing human pressures. Human health is arguably the most complex of the major types of global change impacts on our societies; understanding how to prepare for the impacts on and improve the resilience of our health systems is surely one of the grand challenges of Earth System science.

REFERENCES

Achard F., Eva H.D., Stibig H.J., Mayaux P., Gallego J., *et al*. (2002), Determination of deforestation rates of the world's humid tropical forests, *Science*, 297:999-1002.

ACIA (2004), *Impacts of a Warming Arctic: Arctic Climate Impacts Assessment*, Cambridge University Press, Cambridge, UK, 139 pp.

Alley R.B., Anandakrishnan S., Jung P. (2001), Stochastic resonance in the North Atlantic, *Paleoceanography*, 16:190-198.

Alverson K., Bradley R., Pedersen T. (eds.) (2003), *Paleoclimate, global change and the future*, The IGBP Global Change series, Springer-Verlag, Berlin Heidelburg New York, 220 pp.

Andreae, M.O., Climatic effects of changing atmospheric aerosol levels. In: *World Survey of Climatology. Vol. 16: Future Climates of the World*, A. Henderson-Sellers (ed.), pp. 341-392, Elsevier, Amsterdam, 1995.

Andreae, M.O., Talaue-McManus, L., and Matson, P.A., Anthropogenic modification of land, coastal and atmospheric systems as threats to the functioning of the Earth System. In: *Earth System Analysis for Sustainability*, H.-J. Schellnhuber, P.J. Crutzen, W.C. Clark, M. Claussen, & H. Held (eds.), pp. 245-264, The MIT Press, Cambridge, Mass., 2004.

Andreae, M.O., Jones, C.D., and Cox, P M., Strong present-day aerosol cooling implies a hot future, *Nature*, 435, 1187-1190, 2005.

Anisimov O. (2004), Global change and the cryosphere: impacts on snow, ice and permafrost. In: Steffen W., Sanderson A., Tyson P., Jäger J., Matson P., Moore III B., Oldfield F., Richardson K., Schellnhuber H.-J., Turner II B.L., Wasson R., *Global Change and the Earth System: A Planet Under Pressure*, IGBP Global Change Series, Springer-Verlag, Berlin Heidelburg New York, pp. 150-151.

Ball J.B. (2001), Global forest resources: history and dynamics. In: *The Forests Handbook*, Vol. 1, J. Evans (ed.), pp. 3-22, Oxford: Blackwell Sci., 418 pp.

Blunier T., Chappellaz J., Schwander J., Barnola J.-M., Desperts T., Stauffer B., Raynaud D. (1993), Atmospheric methane record from a Greenland ice core over the last 1000 years, *Journal of Geophysical Research*, 20:2219-2222.

Bond G., Showers W., Elliot M., Evans M., Lotti R., Hajdas I., Bonani G., Johnsen S.J. (1999), The North Atlantic's 1-2 kyr climate rhythm: Relation to Heinrich events, Dansgaard/Oeschger cycles and the Little Ice Age. In: Clark P.U., Webb R.S., Keigwin L.D. (eds.), *Mechanisms of global climate change at millennial time scales*, American Geophysical Union, Washington, USA, pp. 35-58.

Brasseur G.P., Artaxo P., Barrie L.A., Delmas R.J., Galbally I., Hao W.M., Harriss R.C., Isaksen I.S.A., Jacob D.J., Kolb C.E., Prather M., Rodhe H., Schwela D., Steffen W., Wuebbles D.J. (2003), An integrated view of the causes and impacts of Atmospheric Changes. In: Brasseur G.P., Prinn R.G., Pszenny A.A.P. (eds.), *The Changing Atmosphere: An integration and synthesis of a decade of tropospheric Chemistry research*, The International Global Atmospheric Chemistry Project (IGAC), The IGBP Global Change series, Springer-Verlag, Berlin Heidelburg New York.

Burke L., Bryant D., McManus J.W., Spalding M. (1998), *Reefs at risk: A map based indicator of threats to the world's coral reefs*, World Resources Institute, Washington DC.

Canning D. (2001), World Bank: *A database of world infrastructure stocks, 1950-95*, World Bank, Washington DC, USA.

Charney J., Stone P.H. (1975), Drought in the Sahara: a biogeophysical feedback mechanism, *Science*, 187:434-435.

Crutzen P. (2004), The ozone hole. In: Steffen W., Sanderson A., Tyson P., Jäger J., Matson P., Moore III B., Oldfield F., Richardson K., Schellnhuber H.-J., Turner II B.L., Wasson R., *Global Change and the Earth System: A Planet Under Pressure*, IGBP Global Change Series, Springer-Verlag, Berlin Heidelburg New York, pp. 236-237.

Crutzen, P.J., Overview of tropospheric chemistry: Developments during the past quarter century and a look ahead, *Faraday Discuss.*, 100, 1-21, 1995.

Crutzen P.J., Stoermer E.F. (2001), The 'Anthropocene', *International Geosphere-Biosphere Programme (IGBP) Newsletter*, 41:12-13.

DeFries R., Houghton R.A., Hansen M., Field C., Skole D.L., Townshend J. (2002), Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 90s, *Proceedings of the National Academy of Sciences*, 99(22):14256-14261.

Dicken P. (1992), *Global shift: The internationalization of economic activity*, 2nd ed., Guilford Press, New York.

Döös B.R. (2002), Population growth and loss of arable land, *Glob. Environ. Change: Hum. Policy Dimens.*, 12:303-311.

Dynesius M., Nilsson C. (1994), Fragmentation and flow regulation of river systems in the Northern third of the world, *Science*, 266:753-762.

Eltahir E.A.B., Bras R.L. (1996), Precipitation recycling, *Rev. Geophys.*, 34:367-378.

EPICA community members (2004), Eight glacial cycles from an Antarctic ice core, *Nature*, 429:623-628.

Etheridge D.M., Steele L.P., Langenfelds R.L., Francey R.J., Barnola J.-M., Morgan V.I. (1996), Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn, *Journal of Geophysical Research*, 101:4115-4128.

Falkowski P., Scholes R.J., Boyle E., Canadell J., Canfield D., Elser J., Gruber N., Hibbard K., Högberg P., Linder S., Mackenzie F.T., Moore III B., Pedersen T., Rosenthal Y., Seitzinger S., Smetacek V., Steffen W. (2000), The global carbon cycle: A test of our knowledge of Earth as a system, *Science*, 290:291-296.

FAO (2000), *The state of world fisheries and aquaculture*, Food and Agricultural Organization of the United Nations, Rome. Available at: http://www.fao.org/sof/sofia/index_en.htm [30 April 2002].

FAO (2001a), *FAO (Food and Agriculture Organization of the United Nations) Statistical Databases*. http://apps.fao.org

FAO (2001b), Global forest resources assessment 2000 (FRA 2000): main report, *FAO For. Pap. 140*, FAO, Rome.

FAOSTAT (2002), *Statistical Databases, Food and Agriculture Organization of the United Nations*, Rome. Available at: http://www.apps.fao.org [12 August 2002].

Fu C., Harasawa H., Kasyanov V., Kim J.-W., Ojima D., Wan Z., Zhao S. (2002), Regional-global interactions in East Asia. In: Tyson P., Fuchs R., Fu C., Lebel L., Mitra A.P., Odada E., Perry J., Steffen W., Virji H. (eds.), *Global-Regional Linkages in the Earth System*, IGBP Global Change Series, Springer-Verlag, Berlin Heidelberg New York, pp. 109-149.

Geist H.J., Lambin E.F. (2002), Proximate causes and underlying driving forces of tropical deforestation, *BioScience*, 52:143-150.

Goldewijk K.K. (2001), Estimating global land use change over the past 300 years: the HYDE database, *Glob. Biogeochem. Cycles*, 15:417-434.

Goldewijk K.K., Ramankutty N. (2003), Land cover change over the last three centuries due to human activities: assessing the differences between two new global data sets, *GeoJournal*, in press.

Grübler A. (1994), Technology. In: *Changes in Land Use and Land Cover: A global perspective*, W.B. Meyer, B.L. Turner (eds.), Cambridge, UK: Cambridge Univ. Press, pp. 287-328.

Grübler A. (1998), *Technology and global change*, Cambridge University Press, Cambridge.

Hare B. (2005), Relationship between increases in global mean temperature and impacts on ecosystems, food production, water and socio-economic systems. In: *Avoiding Dangerous Climate Change, a scientific symposium on stabilisation of greenhouse gases*, UK Met Office, Exeter, UK, 1-3 Feb 2005.

Heintzenberg J., Raes F., Schwartz S.E., Ackermann I., Artaxo P., Bates T.S., Benkovitz C., Bigg K., Bond T., Brenguier J.-L., Eisele F.L., Feichter J., Flossman A.I., Fuzzi S., Graf H.-F., Hales J.M., Herrmann H., Hoffman T., Huebert B., Husar R.B., Jaenicke R., Kärcher B., Kaufman Y., Kent G.S., Kulmala M., Leck C., Liousse C., Lohmann U., Marticorena B., McMurry P., Noone K., O'Dowd C., Penner J.E., Pszenny A., Putaud J.-P., Quinn P.K., Schurath U., Seinfeld J.H., Sievering H., Snider J., Sokolik I., Stratmann F., van Dingenen R., Westphal D., Wexler A.S., Wiedensohler A., Winker D.M., Wilson J. (2003), Tropospheric aerosols. In: Brasseur G.P., Prinn R.G., Pszenny A.A.P. (eds.), *The changing atmosphere: An integration and synthesis of a decade of tropospheric chemistry research*, The International Global Atmospheric Chemistry Project (IGAC), IGBP Global Change Series, Springer-Verlag, Berlin Heidelburg New York.

Hong S., Candelone J.P., Patterson C.C., Boutron C.F. (1994), Greenland ice evidence of hemispheric lead pollination two millennia ago by Greek and Roman civilization, *Science*, 265:1841-1843.

Houghton R.A., Boone R.D., Melillo J.M., Palm C.A., Myers N., *et al.* (1985), Net flux of carbon dioxide from tropical forest in 1980, *Nature*, 316:617-620.

Howarth R.W., Billen G., Swaney D., Townsend A., Jaworski N., Lajtha K., Downing J.A., Elmgren R., Caraco N., Jordan T., Berendse F., Freney J., Kudeyarov V., Murdoch P., Zhao-Liang Z. (1996), Regional nitrogen budgets and riverine N and P fluxes for the drainages to the North Atlantic Ocean: Natural and human influences. In: Howarth R.W. (ed.), *Nitrogen cycling in the North Atlantic Ocean and its watersheds*, Kluwer, Dordrecht, 304 pp. (Reprinted from Biogeochemistry, 35(1), 1996).

International Fertilizer Industry Association (2002), *Fertilizer Indicators*. Available at: http://www.fertilizer.org/ifa/statistics/indicators/ind_cn_world.asp [25 Oct 2002].

IPCC (2001), *Climate Change 2001, The Scientific Basis, Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, J.T. Houghton *et al.* (eds.), Cambridge University Press, Cambridge, UK and New York, N.Y., USA, 881 pp.

Jaworski N.A., Howarth R.W., Hetling L.J. (1997), Atmospheric deposition of nitrogen oxides onto the landscape contributes to coastal eutrophication in the northeast United States, *Environmental Science and Technology*, 31:1995-2004.

Kanakidou, M., Seinfeld, J.H., *et al.*, Organic aerosol and global climate modelling: a review, *Atmos. Chem. Phys.*, 5, 1053-1123, 2005.

Kasperson J.X., Kasperson R.E., Turner B.L. (eds.) (1995), *Regions at Risk: Comparisons of Threatened Environments*, Tokyo: UN Univ., Press. 588 pp.

Keeling C.D., Whorf T.P. (2000), Atmospheric CO_2 records from sites in the SIO air sampling network. In: *Trends: A compendium of data on global change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, TN, USA.

Kelleher G., Bleakley C., Wells S. (1995), *A global representative system of marine protected areas*, Volume 1, World Bank, Washington, DC.

Klein Goldewijk K., Battjes J.J. (1997), *One hundred year database for integrated environmental assessments*, National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands.

Kleypas J.A., Buddemeier R.W., Archer D., Gattuso J.-P., Langdon C., Opdyke B.N. (1999), Geochemical consequences of increased atmospheric carbon dioxide on coral reefs, *Science*, 284:118-120.

Kremer H., Crossland C. (2002), Coastal change and the 'Anthropocene': Past and future perspectives of the IGBP-LOICZ project. In: *Deutsches Nationalkomitee fuer das Internationale Hydrologische Programm (IHP) der UNESCO und das Operationelle Hydrologische Programm (OHP) der WMO*, Herausgeber, Koblenz, Germany, IHP/OHP Berichte, Sonderheft 13, ISSN 0177-9915:3-19.

Lambin E.F., Geist H., Lepers E. (2003), Dynamics of land use and cover change in tropical regions, *Annual Review of Environment and Resources*, 28:205-241.

Lambin E.F., Turner B.L., Geist H., Agbola S., Angelsen A., *et al*. (2001), The causes of land-use and land-cover change: moving beyond the myths, *Glob. Environ. Change*, 11:261-269.

Lee J.A., Tallis J.H. (1973), Regional and historical aspects of lead pollution in Britain, *Nature*, 245:216-218.

Legrand M., Hammer C., Deangelis M., Savarino J., Delmas R., Clausen H., Johnsen S.J. (1997), Sulfur-containing species (methanesulfonate and SO4) over the last climatic cycle in the Greenland Ice Core Project (central Greenland) ice core, *Journal of Geophysical Research*, 102(C12):26,663-26,679.

Lepers E., Lambin E., DeFries R., Janetos A., *et al*. (2003), Areas of rapid land-cover change of the world, MEA Rep., *Millenium Ecosystem Assessment*, Penang, Malaysia.

Livett E.A., Lee J.A., Tallis J.H. (1979), Lead, zinc and copper analyses of British blanket peats, *Journal of Ecology*, 67:865-891.

Machida T., Nakazawa T., Fujii Y., Aoki S., Watanabe O. (1995), Increase in the atmospheric nitrous oxide concentration during the last 250 years, *Geophysical Research Letters*, 22:2921-2924.

Mackenzie F.T., Ver L.M., Lerman A. (2002), Century-scale nitrogen and phosphorus controls of the carbon cycle, *Chemical Geology*, 190:13-32.

Mann M.E., Jones P.D. (2003), Global surface temperatures over the past two millennia, *Geophys. Res. Lett.*, 30:1820, doi: 10.1029/2003GL 017814.

Martínez-Cortizas A., Pontevedra-Pombal X., García-Rodeja E., Nóvoa-Muñoz J.D., Shotyk W. (1999), Mercury in a Spanish peat bog: Archive of climate change and atmospheric metal deposition, *Science*, 284:939-942.

Mayewski P.A., Lyons W.B., Spencer M.J., Twickler M.S., Buck C.F., Whitlow S.I. (1990), An ice core record of atmospheric response to anthropogenic sulphate and nitrate, *Nature*, 346:554-556.

McDonalds (2002), *Homepage*. Available at: http://www.mcdonalds.com [28 Oct. 2002].

Meybeck M., Vörösmarty C. (1999), Global transfert of carbon by rivers, *Global Change Newsletter No. 37*, International Geosphere Biosphere Programme, Stockholm, Sweden pp. 12-14.

Milly P.C.D., Wetherald R.T., Dunne K.A., Delworth T.L. (2002), Increasing risk of great floods in a changing climate, *Nature*, 415:514-517.

Mittermeier R., Mittermeier C.G., Gil P.R., Pilgrim J., Fonseca G., *et al.* (2003), *Wilderness: Earth's Last Wild Places*, Chicago: Univ. Chicago Press, 576 pp.

Naylor R.L., Goldburg R.J., Mooney H., Beveridge M., Clay J., Folke N., Kautsky N., Lubcheno J., Primavera J., Williams M. (2000), Nature's subsidies to shrimp and salmon farming, *Science*, 282:883-884.

Nordhaus W. (1997), Do real wage and output series capture reality? The history of lighting suggests not. In: Bresnahan T., Gordon R. (eds.), *The economics of new goods*, University of Chicago Press, Chicago.

Oldfield F., Steffen W. (2004), The Earth System. In: Steffen W., Sanderson A., Tyson P., Jäger J., Matson P., Moore III B., Oldfield F., Richardson K., Schellnhuber H.-J., Turner II B.L., Wasson R., *Global Change and the Earth System: A Planet Under Pressure*, IGBP Global Change Series, Springer-Verlag, Berlin Heidelburg New York, p. 7.

Orr, J.C., Fabry, V.J., *et al.*, Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, *Nature*, 437, 681-686, 2005.

Otterman J. (1974), Baring high-albedo soils by overgrazing: a hypothesised desertification mechanism, *Science*, 86:531-533.

Parmesan C., Yohe G. (2003), A globally coherent fingerprint of climate change impacts across natural systems, *Nature*, 421:37-42.

Pauly D., Christensen V. (1995), Primary production required to sustain global fisheries, *Nature*, 374:255-257.

Petit J.R., Jouzel J., Raynaud D., Barkov N.I., Barnola J.-M., Basile I., Bender M., Chappellaz J., Davis M., Delaygue G., Delmotte M., Kotlyakov V.M., Legrand M., Lipenkov V.Y., Lorius C., Pépin L., Ritz C., Saltzman E., Stievenard M. (1999), Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, *Nature*, 399:429-436.

Pulp and Paper International (1993), PPI's international fact and price book. In: *FAO forest product yearbook 1960-1991*, Food and Agriculture Organization of the United Nations, Rome.

Rahmstorf S., Stocker T.F. (2004), Thermohaline circulation: Past changes and future surprises. In: Steffen W., Sanderson A., Tyson P., Jäger J., Matson P., Moore III B., Oldfield F., Richardson K., Schellnhuber H.-J., Turner II B.L., Wasson R., *Global Change and the Earth System: A Planet Under Pressure*, IGBP Global Change Series, Springer-Verlag, Berlin Heidelburg New York, pp. 240-241.

Ramanathan, V., Chung, C., Kim, D., Bettge, T., Buja, L., Kiehl, J.T., Washington, W.M., Fu, Q., Sikka, D.R., and Wild, M., Atmospheric brown clouds: Impacts on South Asian climate and hydrological cycle, *Proc. Natl. Acad. Sci. USA*, 102, 5326-5333, 2005.

Ramanathan, V., Crutzen, P.J., Kiehl, J.T., and Rosenfeld, D., Atmosphere – Aerosols, climate, and the hydrological cycle, *Science*, 294, 2119-2124, 2001.

Ramankutty N., Foley J.A. (1999), Estimating historical changes in global land cover: croplands from 1700 to 1992, *Glob. Biogeochem. Cycles*, 13:997-1027.

Ramankutty N., Foley J.A., Olejniczak N.J. (2002), People on the land: changes in global population and croplands during the 20th century, *Ambio*, 31:251-257.

Reid W. & M.A. Core Writing Team (2005), *Millennium Ecosystem Assessment Synthesis Report*, pre-publication final draft, March 2005.

Renberg I., Wik-Persson M., Emteryd O. (1994), Pre-industrial atmospheric lead contamination detected in Swedish lake sediments, *Nature*, 368:323-326.

Rial J.A., Pielke R.A. Sr, Beniston M., Claussen M., Canadell J., Cox P., Held H., de Noblet-Ducoudre N., Prinn R., Reynolds J., Salas J.D. (2004), Nonlinearities, feedbacks and critical thresholds within the Earth's climate system, *Climatic Change*, 65:11-38.

Richards J. (1990), Land transformation. In: Turner II B.L., Clark W.C., Kates R.W., Richards J.F., Mathews J.T., Meyer W.B. (eds.), *The Earth as transformed by human action: Global and regional changes in the biosphere over the past 300 years*, Cambridge University Press, Cambridge pp. 163-201.

Roderick M.L., Farquhar G.D. (2004), The pan evaporation paradox. In: Steffen W., Sanderson A., Tyson P., Jäger J., Matson P., Moore III B., Oldfield F., Richardson K., Schellnhuber H.-J., Turner II B.L., Wasson

R., *Global Change and the Earth System: A Planet Under Pressure*, IGBP Global Change Series, Springer-Verlag, Berlin Heidelburg New York, p. 167.

Rühlemann C., Mulitza S., Müller P.J., Wefer G., Zahn R. (1999), Warming of the tropical Atlantic Ocean and slow down of thermocline circulation during the last deglaciation, *Nature*, 402:511-514.

Sagan C., Toon O.B., Pollack J.B. (1979), Anthropogenic albedo changes and the earth's climate, Science, 206:1363-1368.

Sala O.E., Chapin F.S., Armesto J.J., Berlow E., Bloomfield J., *et al.* (2000), Biodiversity – global biodiversity scenarios for the year 2100, *Science*, 287:1770-1774.

Schellnhuber H.J. (2002), Coping with Earth System complexity and irregularity. In: Steffen W., Jäger J., Carson D., Bradshaw C. (eds.), *Challenges of a changing earth: Proceedings of the Global Change Open Science Conference*, Amsterdam, The Netherlands, 10-13 July 2001, Springer-Verlag, Berlin Heidelberg New York, pp. 151-156.

Scholes M.C., Matrai P.A., Andreae M.O., Smith K.A., Manning M.R. (2003), Biosphere-atmosphere interactions. In: Brasseur G.P., Prinn R.G., Pszenny A.A.P. (eds.), *The changing atmosphere: An integration and synthesis of a decade of tropospheric chemistry research*, The International Global Atmospheric Chemistry Project (IGAC), IGBP Global Change series, Springer-Verlag, Berlin Heidelburg New York.

Shiklomanov I.A. (1990), Global water resources, *Nature and Resources*, 26(3).

Shotyk W., Cheburkin A.K., Appleby P.G., Fankhauser A., Kramers J.D. (1996), Two thousand years of atmospheric arsenic, antimony, and lead deposition recorded in an ombrotrophic peat bog profile, Jura Mountains, Switzerland, *Earth and Planetary Science Letters*, 145(1-4):E1-E7.

Shotyk W., Weiss D., Appleby P.G., Cheburkin A.K., Frei R., Gloor M., Kramers J.D., Reese J., Van Der Knaap W.O. (1998), History of atmospheric lead deposition since 12,370 14C yr BP from a Peat Bog, Jura Mountains, Switzerland, *Science*, 271:1635-1640.

Steffen W., Andreae M.O., Bolin B., Crutzen P.J., Cox P., Cubasch U., Held H., Nakicenovic N., Scholes R., Talaue-McManus L., Turner II B.L. (2004a), Abrupt changes: the Achilles heels of the Earth System, *Environment*, 46(3):9-20.

Steffen W., Sanderson A., Tyson P., Jäger J., Matson P., Moore III B., Oldfield F., Richardson K., Schellnhuber H.-J., Turner II B.L., Wasson

R. (2004b), *Global Change and the Earth System: A Planet Under Pressure*, IGBP Global Change Series, Springer-Verlag, Berlin Heidelburg New York, 336 pp.

Steffen W., Love G., Whetton P. (2005), Approaches to defining dangerous climate change: a southern hemisphere perspective. In: *Avoiding Dangerous Climate Change, a scientific symposium on stabilisation of greenhouse gases*, UK Met Office, Exeter, UK, 1-3 Feb 2005.

Tilman D. (1999), Global environmental impacts of agricultural expansion: the need for sustainable and efficient practices, *Proc. Natl. Acad. Sci. USA*, 96:5995-6000.

Tolba M.K., El-Kholy O.A. (eds.) (1992), *The world environment 1972-1992: Two decades of challenge*, United Nations, Chapman & Hall, London.

Trimble S.W., Crosson P. (2000), Land use – US soil erosion rates: myth and reality, *Science*, 289:248-250.

Turner II B.L., Clark W.C., Kates R.W., Richards J.F., Mathews J.T., Meyer W.B. (1990), *The Earth as transformed by human action: Global and regional changes in the biosphere over the past 300 years*, Cambridge University Press, Cambridge.

UN Center for Human Settlements (2001), *The state of the world's cities*, 2001, United Nations. Available at: http://www.unchs.org [4 Oct 2002].

UN Population Division (2000), *World population prospects: The 2000 revision*, United Nations, Department of Economic and social affairs.

UN Population Division, Dep. Econ. Soc. Aff., UN Secr. (2002), *World Urbanization Prospects: The 2001 Revision* (ESA/P/WP.173), New York: UN Publ., 328 pp. http://www.un.org/esa/population/publications/wup 2001/wup2001dh.pdf

UNEP (2000), *Global environmental outlook 2000*, Clarke R. (ed.), United Nations Environment Programme.

UNEP/WMO (2002), *Scientific assessment of ozone depletion*, prepared by the Scientific assessment panel of the Montreal Protocol on Substances that deplete the ozone hole, United Nations Environment Programme/World Meteorological Organization.

US Bureau of the Census (2000), *International Database*. Available at: http://www.census.gov/ipc/www/worldpop.htm. Data updated 10 May 2000.

USEPA (1989) *Report to Congress: Water quality of the nation's lakes*, EPA 440/5-89-003, US Environmental Protection Agency, Washington, USA.

Vitousek P.M., Mooney H.A., Lubchenco J., Melillo J.M. (1997), Human domination of earth's ecosystems, *Science*, 277:494-499.

Vörösmarty C.J., Meybeck M. (2003), Responses of continental aquatic systems at the global scale: New paradigms, new methods. In: Kabat P., Claussen M., Dirmeyer P.A., Gash J.H.C., de Guenni L.B., Meybeck M., Pielke Sr. R.A., Vörösmarty C., Hutjes R.W.A., Luetkemeier S. (eds.), *Vegetation, water, humans and the climate: A new perspective on an interactive system*, IGBP Global Change Series, Springer-Verlag, Berlin Heidelburg New York, in press.

Vörösmarty C.J., Meybeck M., Fekete B., Sharma K., Green P., Syvitksi J. (2003), Anthropogenic sediment retention: Major global-scale impact from the population of registered impoundments, *Global and Planetary Change*, in press.

Vörösmarty C.J., Sharma K., Fekete B., Copeland A.H., Holden J., Marble J., Lough J.A. (1997), The storage and aging of continental runoff in large reservoir systems of the world, *Ambio*, 26:210-219.

Wilson E.O. (1992), *The diversity of life*, Allen Lane, the Penguin Press.

WMO (1999), Scientific assessment of ozone depletion: 1998, *Global Ozone Research and Monitoring Project – Report No. 44*, World Meteorological Organization, Geneva, Switzerland, 732 pp.

Woodwell G.M., Hobbie J.E., Houghton R.A., Melillo J.M., Moore B., *et al*. (1983), Global deforestation: contribution to atmospheric carbon dioxide, *Science*, 222:1081-1086.

World Bank (2002), *Data and statistics*. Available at: http://www1.worldbank.org/economicpolicy/globalization/data.html [4 Oct 2002].

World Commission on Dams (2000), Dams and development: A new framework for decision-making, *The Report of the World Commission on Dams*, Earthscan Publications Ltd, London and Sterling, VA, USA.

World Tourism Organization (2001), *Tourism industry trends, Industry Science Resources*. Available at: http://www.world-tourism.org [22 Oct 2002].

WRI (1990) Forest and rangelands. In: *A Guide to the Global Environment*, World Resources Institute, Washington DC, pp. 101-120.

WRI (1996), *World resources: A guide to the global environment, The Urban Environment*, World Resources Institute, Oxford University Press, Oxford and New York.

WRI (2000), *World resources: People and ecosystems: The fraying web of life*, World Resources Institute, Oxford University Press, Oxford and New York.

WRI (2001), *Earth trends 2001*, World Resources Institute, Washington DC.
WRI (2003), *A guide to world resources 2002-2004: Decisions for the Earth*, A joint publication with UN Development Program, UN Environmental Program, World Bank and World Resources Institute, Washington DC, USA.
Young A. (1999), Is there really spare land? A critique of estimates of available cultivable land in developing countries, *Environ. Dev. Sustain.*, 1:3-18.

BIODIVERSITY AND INFECTIOUS DISEASE: WHY WE NEED NATURE

ANDREW P. DOBSON

Preface

The large leather bible hit the congressional committee room table with a crash that commanded the attention of congressmen, their aides, and the press assembled for the Endangered Species Act hearings. 'Extinction is a sin – We are destroying God's creation', announced the President of Christians for the Environment. This produced a nervous shuffling in seats. Congressmen unmoved by economic or scientific arguments were plainly disconcerted by this sharp reminder that our moral responsibility to other species puts parts of their constituency in blunt conflict with each other. The rest of us breathed a little more easily. The current US Congress, curiously immune to economic and scientific arguments, has a visceral response to the Old Testament. Here I will expand on this testimony and outline the broader economic, scientific and ethical arguments for 'Why we need nature'.

Biodiversity as Food-Webs

One of the first things we learn about in school nature classes is food chains. My five-year-old son can happily arrange krill, fish, big fish, and sharks into a logical hierarchy of producers and consumers. Ecologists and economists have long been fascinated with the mathematical properties of simplified food webs (Dunne, 2005). The Italian mathematician, Vito Volterra, initially examined the properties of the simplest one predator and one prey food web in the 1920s (Volterra, 1926). His stimulus was the worry that his daughter would marry a fisherman. When he created a mathematical model of the interactions between fisherman and fish he realized that

that even this simple model had a pathological tendency to undergo sustained cycles of abundance of predators (fisherman) and prey (fish). While this was an unsuccessful deterrent to his daughter's marriage, it catalyzed a massive expansion in mathematical interest in the properties of ecological systems. This early work of Volterra and Lotka (Lotka, 1923) advanced our understanding of consumer-resource relationships considerably beyond Malthus, whose worries about the geometric and arithmetic rates of increase of consumer and resource are but a sub-set of this broad and complex set of problems. The mathematical study of predator-prey systems is at the heart of many natural processes; lions chasing zebra, insects eating plants and each-other, right through to how the immune system interacts with viral and other pathogens that invade our bodies (May, 2001; Nowak and May, 1991). These problems rapidly become more complex as we increase the numbers of species that interact as hierarchical networks of consumer and resource species within a food web (Cohen, 1989; Cohen *et al.*, 1990; Cohen *et al.*, 2003; Dunne *et al.*, 2002).

'There Ain't Have Been Some Clever Geezers…'

To understand the scientific complexity of the problem we have to compare ecology and the other life sciences to physics and chemistry. One way to do this is to reduce these sciences to the level of the numbers of different particles they consider, the spatial and temporal scales at which those particles interact, and the complexity, or non-linearity, of these interactions. In their purest form physics and chemistry consider limited numbers of particles interacting at either extremely small or rapid scales, or at huge scales over rapid timescales. The sheer magnitude of these scales creates a mystical awe. Ironically this masks their lack of utility to any aspect of human health or well being! In contrast, biologists consider interactions between large numbers of particles of different types (molecules, cells, species) embedded in a complex hierarchy of webs, where processes occur at timescales from seconds and minutes to millennia. Because many of the species involved are cute, charismatic, or pests, and studied at scales closer to fish-tanks and fields, there is less mystical awe in the study of their interactions.

When we convert forest or savannas to agricultural land we simplify the local food webs and focus on cultivating a handful of domesticated species as resources. This simplification reduces our potential to benefit from the services supplied by the natural web. How we value nature is

intimately linked with understanding the trade-offs between agriculture, trade and nature (Arrow *et al.*, 1995; Balmford *et al.*, 2002; Costanza, 1991). Inherently asking the question 'Do we need Nature?' assumes we can realistically compare the long-term value of natural habitats, whose benefits are diffusely appreciated by a variety of people, with the instantaneous and focused annual yields of agriculture, timber products, or golf course revenues. Some of the problems associated with this calculation boil down to simply not knowing if the land is worth more as agricultural land or natural habitat (Balmford *et al.*, 2002). If we focus on a single service or product we will always underestimate its natural value. If we try and completely enumerate its value, we run the risk of exaggeration. Even if we could undertake these calculations, we'd then need to quantify the dependence of the modified habitat on services produced by the natural habitat (Dobson, 2005b).

In the absence of a comprehensive pricing of nature's services, then the natural habitat will appear to be worth less than when converted to agricultural land. This creates further pressure to convert all of the land to agriculture. As an example, consider an area of land containing a patch of woodland or prairie – this could range in size from a small farm, to many hundreds of square miles, the Amazon basin, or the Serengeti. If we convert the forest or savanna to agricultural land (or a golf course), its economic value will be determined by the quality and quantity of products and services produced, minus the cost of converting the land. In contrast, if we leave the land in an undisturbed state it will have an economic value determined by a complex calculation that sums the resources people remove from it for food or fiber, plus the added value we obtain as it cleanses the surrounding air and water. We may also have to add its value as a sink for greenhouse gases. Finally we must include a more complex value determined by the pleasure people receive from spending time there, or even simply knowing it is there. Alternatively, we could convert a proportion of the land to agriculture and leave the rest as forest. This may be particularly sensible if the agricultural land is dependent upon the forest for services such as a constant water supply, or a source of pollinators that ensure crop fertility (Figure 1, see page 407).

Determining the proportion of land to convert and the proportion to leave available for other species is the central problem of natural resource management. Unfortunately the less we value Nature then the smaller the proportion of land that will be kept set aside for other species. Many international conservation treaties recommend leaving ten percent of the

land as set aside for Nature. Their logic stems from the well-known 'species-area' relationship discovered by Robert MacArthur and E.O. Wilson in the 1960s (MacArthur and Wilson, 1963; MacArthur and Wilson, 1967); this relationship suggests that each successive 90% loss in habitat dooms 50% of remaining species to local extinction. From a purely utilitarian perspective, the amount of biodiversity we save is one answer to the question 'Will this level of biodiversity be sufficient to supply all the necessary services we are used to receiving from Nature?'

Unfortunately, the benefits of maintaining biodiversity are most sharply appreciated in their absence! This usually occurs when we notice a decline in air or water quality, reductions in crop pollination rates, or increased pest and disease outbreaks. We are then stuck with the problem of either replacing 'Nature's services' with a technological fix, or of trying to restore the natural habitat. Both are likely to be expensive; in the case of restoration the significant current costs may not be met by future discounted potential benefits (Dobson, 2005b; Dobson *et al.*, 1997; Simberloff *et al.*, 1998). Restoration also assumes we know enough about ecosystem functioning to put food-webs back together (Bradshaw, 1983; Bradshaw, 1984). This illustrates one of the deepest scientific ironies of the 21st Century: while we know much about the structure of atoms and of the universe, we have only the most rudimentary understanding of the structure of salt-marshes, forests, or even the soil beneath our crops.

Reasons to be Careful... Part 3

When land is converted from natural habitat, different species will go extinct at different rates (Seabloom *et al.*, 2002; Tilman *et al.*, 1994). Species are lost as their habitats are converted to agricultural land, as we exploit them for food, or as they are out-competed for space and resources by a suite of invasive plants and animals that are the camp followers of human expansion (Crawley, 1986; Soulé, 1986). We are thus seeing a transformation from a world of tropical forests, tigers, orchids and pandas, to one of weeds, mosquitoes, goats, and sparrows. This sequential loss of biodiversity means that the economic services will be lost at different rates. Services that are predominantly supplied by species with large area requirements, rare species, or species with specialist habitat requirements, will be lost more rapidly than those mediated by species that can persist in a handful of soil, or those able to rapidly adapt to the suburban environment (Dobson *et al.*, 2006; Jenkins, 2003; Kremen,

2005). In some cases agricultural or weedy species may replace some of the ecosystem services performed by the original native species. In general, the services provided by species at the top of the food chain will be lost more rapidly than those provided by those at the base of the web (Dobson *et al.*, 2006; Loreau *et al.*, 2001). This creates a hierarchical loss of natural services, initially the aesthetic benefits will decline, orchids and pandas will become rare, and tigers and other predators will disappear locally (Figure 2, see page 408). Except in areas that are strongly dependent upon ecotourism, this will have only a limited impact on the local economy. It will create problems, the loss of predators will cause the species they prey upon to increase in abundance and become pests (Duffy, 2003). In turn this will lead to an increase in thorny, inedible vegetation that encourages the now abundant deer, rabbits and insects to feed upon crops. The increased contact of these species with humans and domestic livestock increases the potential for disease transmission. Laws such as the Endangered Species Act in the United States can prevent this cascade of events. While focusing attention on individual species, the law seeks to protect entire functioning food webs (Carroll *et al.*, 1996; Eisner *et al.*, 1995; Mann and Plummer, 1995).

Biodiversity and Infectious Diseases

A central question in the field of biodiversity and conservation biology is whether increased diversity of host species tends to either buffer or amplify disease outbreaks (Dobson, 2004; Dobson, 2005a). In the simplest case the potential for disease outbreaks will be determined by the magnitude of R_0 – this number is formerly defined as the number of secondary infections produced by the initial infectious individual introduced into a population of susceptible hosts (Anderson, 1982; Anderson and May, 1986; Dietz, 1993).

We can think of R_0 as a heuristic mathematical device that allows us to examine the biological conditions that lead to an outbreak; quantifying the relative contribution of the ecological factors that determine the magnitude of R_0 considerably focuses our understanding of how different pathogens might be controlled. In cases where increases in species diversity lead to increases in the number of contacts between infected individuals and potentially susceptible hosts, increased host diversity will always lead to increased values of R_0 and a greater potential for disease outbreaks (Figure 3, see page 409). In contrast, where increases in inter-spe-

cific transmission lead to reductions in within-species transmission, then it is possible for increased host species diversity to lead to reductions in R_0. This will be the case for vector-transmitted pathogens such as malaria, yellow fever, dengue fever and Lyme disease (Dobson, 2004). On one hand, increased host diversity will lead to increases in the resources available to the vector population (e.g., more blood meals). However, as most vectors take a finite number of blood meals per lifetime, this will lead to an increased proportion of bites wasted on hosts that may be less viable resources for the pathogen (Figure 3, page 409). Whether or not the pathogens are buffered or amplified by the increased host diversity will depend on whether increases in the size of the vector population are sufficient to compensate for the 'wasted bites' on less viable hosts. Ecologists have called this reduction in disease risk as host species diversity increases 'the dilution effect' (LoGuidice *et al.*, 2003; Schmidt and Ostfeld, 2001). In a world where global climate change may lead to range expansion of vector transmitted pathogens from the tropics into the sub-tropics and temperate zones, the dilution effect creates an important utilitarian argument for conserving biological diversity (particularly vertebrates) – as long as these species are present in abundance, the biting rate of mosquitoes on humans should be reduced.

Rinderpest in the Serengeti

As a final example of the role that infectious diseases can play in modifying the structure of complex ecosystems, let us consider the impact that a single introduced pathogen has had on species that live in the African savanna; an ecosystem where a large proportion of the human population are highly dependent upon a range of ecosystem services to supply most of their food and economic well-being (Homewood and Rodgers, 1991). The savannas of East Africa support a large pastoralist population, their herds often share grazing habitat with wild antelope species, and the millions of ecotourists that visit the region each year provide the major input into the local economy. Yet the region also provides perhaps the best example of a pathogen completely modifying the structure of a food web: the introduction of the rinderpest virus into sub-Saharan Africa in the 1890s (Branagan and Hammond, 1965; Plowright, 1982). Rinderpest is a morbillivirus that infects hoofed animals: cattle, wild buffalo, wildebeest, giraffe, and other large antelope. It is closely related to both canine distemper (CDV) and measles, two of the com-

monest diseases of humans and their domestic dogs; the recent evolution of these three pathogens is intimately entwined with domestication of dogs and cattle, this created the opportunities for the pathogen to establish in new host species, where a few mutations allowed it to differentiate itself from rinderpest, which is the ancestral main trunk of the morbillivirus tree (Barrett, 1987). The split between the three pathogens is so recent (<5000 years) that there is still strong cross immunity between them; inoculation of dogs with rinderpest vaccine will protect them against distemper. This again raises interesting questions about how we classify pathogens in food webs where they may fail to establish a dependence upon a host, but stimulate an immunological response that allows the host to protect itself against invasion by a potentially lethal natural enemy.

Rinderpest caused one of the largest pandemics in recorded history, it took 10 years to spread from the Horn of Africa to the Cape of Good Hope, during this time it reduced the abundance of many ungulate / artiodactyl species by as much as 80% (Plowright, 1982). This in turn produced a transient glut of food for decomposers and scavengers, such as vultures and jackals, however, this quickly lead to a massive reduction in food supply for the predators that relied on wildebeest and other game for food. The removal of the ungulates changed the grazing intensity on both shrubs and grasses. This seems to have allowed some tree species to undergo a pulse of recruitment, thus many of the fever trees that create woodlands in damper areas of the savanna seem consist mainly of individual trees that are now just over a hundred years old (Dobson, 1995; Dobson and Crawley, 1994; Prins and Weyerhaeuser, 1987). In contrast, reduced levels of grass grazing led to an increased fire frequency, which prevented the establishment of miombo bushland that had previously covered the savanna. This in turn modified the habitat for many of the predators that require thicker bush coverage to successfully attack their prey.

The development of a vaccine for rinderpest in the 1950s allowed these processes to be reversed (Dobson, 1995; Plowright and Taylor, 1967). There is an instructive irony here: the presence of rinderpest in wildlife was blamed as the major reason why it had proved almost impossible to establish large-scale cattle ranches in East Africa. The rinderpest vaccine was largely developed to help the cattle industry, it was only ever applied to cattle, but this in turn led to its disappearance from wildlife. Thus cattle had been the reservoir and the repeated epidemics observed in wildlife were in response to constant spillovers from cattle. Rinderpest

vaccination has successfully eradicated the disease from most parts of Africa, except in times of civil unrest, when declines in vaccination coverage allow it to resurge. The impact on wildlife has been spectacular: in the Serengeti, wildebeest numbers have grown from around 250,000 to over 1.5 million; buffalo have appeared in areas where they previously unrecorded, and lion and hyena numbers have increased dramatically in response to the enhanced food supply (Sinclair, 1979). This observation strengthens our contention that predators are less effective than pathogens in regulating host abundance. The numbers of some species have declined, for example there are fewer Thompson's gazelles perhaps because of more competition for grassland forage, but more likely because of increased predation pressure from the more numerous hyenas. African wild dogs have also declined since widescale rinderpest vaccination allowed their prey to increase in abundance; this may be primarily due to competition with hyenas, though there may also be increased risk of infectious disease, particularly distemper (Creel and Creel, 1996; Dobson and Hudson, 1986). In the absence of rinderpest, it may be that wild dogs (and other carnivores such as lions), no longer acquire cross-immunity to distemper. The increased wildebeest abundance has reduced the excess of dried grass during the dry season; this has in turn led to a reduction in fire frequency, which has allowed the miombo bushland to return to some areas of the Serengeti (Sinclair and Arcese, 1995; Sinclair and Norton-Griffiths, 1979).

The key points to take from this example is that while the biomass of the rinderpest virus in the Serengeti was always less than 10 kg, the virus had an impact on the abundance and dynamics of nearly all of the dominant plant and animal species in the system – even those which it did not infect. It strengthens the notion that indirect effects may be as important in shaping species abundance and web linkages as are direct interactions. It also reinforces the earlier comments that pathogens may be as powerful as predators in regulating abundance and distribution of free-living species in natural ecosystems. Finally the rinderpest epidemic illustrates the importance of considering not just infectious diseases of humans when we examine the epidemiology of future human well-being. The pastoralist population of East Africa was heavily impacted by the great rinderpest epidemic, as their herds died, the pastoralists either starved, or changed lifestyles and changed from pastoralists to farmers who settled and permanently farmed the same area for crops. In the absence of wildlife, tsetse flies switched to feeding on humans and created a major

epidemic of sleeping sickness (Rogers and Randolph 1988); a vivid illustration of the dilution effect described above. The recovery of wildlife, itself a direct consequence of the control of rinderpest in livestock, led to the huge expansion of tourism which provides most of the financial input into the local economy.

Outro

My friend's Bible again crashes on the congressional committee room table, 'Extinction is a sin!'. We can now add that the changes in land use that create endangered species also create bell-ringers for the decline of a variety of economic goods and services that will become increasingly limiting. This creates a final irony. The most vociferous critics of environmental protection implicitly assume humans and their domestic livestock will be amongst the last species to go extinct. This opinion is inherent in the belief that human existence is independent of the welfare of other species. It is a dangerous form of naivety. Our dependence upon other species makes it highly unlikely that humans will be the last species left alive on the planet. The simplest bet is that we will go extinct about halfway down the list. With a bit of luck, and a deeper understanding of food webs, our technological skills may delay this inevitable demise. This is less naïve, but still assumes we can develop replacements for the services provided by biodiversity. The alternative is to conserve a significant amount of biodiversity and explicitly acknowledge that Nature has a value beyond our current ethical, economic and scientific understanding. Ultimately, our greatest need from Nature may be the challenge it still presents to human creativity. This is both beyond value and essential to our long-term health and economic and spiritual welfare.

REFERENCES

Anderson, R.M. (1982), Transmission dynamics and control of infectious disease agents, pp. 149-176 in R.M. Anderson and R.M. May (eds.), *Population biology of infectious diseases*, Berlin, Springer-Verlag.

Anderson, R.M. and R.M. May (1986), The invasion, persistence and spread of infectious diseases within animal and plant communities, *Philosophical Transactions of the Royal Society of London, Series B-Biological Sciences*, 314:533-570.

Arrow, K., B. Bolin, R. Costanza, P. Dasgupta, C. Folke, C.S. Holling, B.O. Jansson *et al*. (1995), Economic growth, carrying capacity, and the environment, *Science*, 268:520-521.

Balmford, A., A. Bruner, P. Cooper, R. Costanza, S. Farber, R.E. Green, M. Jenkins *et al*. (2002), Economic Reasons for Conserving Wild Nature, *Science*, 297:950-953.

Barrett, T. (1987), The molecular biology of the morbillivirus (measles) group, *Biochem. Soc. Symp.*, 53:25-37.

Bradshaw, A.D. (1983), The reconstruction of ecosystems, *Journal of Applied Ecology*, 20:1-17.

– (1984), Land restoration: now and in the future, *Proc. R. Soc. Lond. B.*, 223:1-23.

Branagan, D., and J.A. Hammond (1965), Rinderpest in Tanganyika: A review, *Bull. Epiz. Dis. Afr.*, 13:225-246.

Carroll, R., C. Augspurger, A. Dobson, J. Franklin, G. Orians, W. Reid, R. Tracy *et al*. (1996), Strengthening the use of science in achieving the goals of the Endangered Species Act: an assessment by the Ecological Society of America, *Ecological Applications*, 6:1-11.

Cohen, J.E. (1989), Food webs and community structure, in Roughgarden *et al*., *Perspectives in Ecological Theory*, 181-203.

Cohen, J.E., F. Briand, and C.M. Newman (1990), *Community Food Webs: Data and Theory: Biomathematics*, v. 20, Berlin, Springer.

Cohen, J.E., T. Jonsson, and S.R. Carpenter (2003), Ecological community description using the food web, species abundance, and body size, *PNAS*, 100:1781-1786.

Costanza, R. (1991), *Ecological Economics. The Science and management of Sustainability*, New York, Columbia University Press.

Costanza, R., R. d'Arge, R. de Groot, S. Farber, M. Grasso, B. Hannon, K. Limburg *et al*. (1997), The value of the world's ecosystem services and natural capital, *Nature*, 387:253-260.

Crawley, M.J. (1986), The population biology of invaders. Philosophical *Transactions of the Royal Society of London B*, 314:711-731.

Creel, S., and N.M. Creel (1996), Limitation of African wild dogs by competition with larger carnivores, *Conservation Biology*, 10:526-538.

Daily, G.C. (1997), *Nature's Services. Societal Dependence on Natural Ecosystems*, Washington, Island Press.

Daily, G.C., S. Alexander, P.R. Ehrlich, L. Goulder, J. Lubchenco, P.A. Matson, H.A. Mooney *et al*. (1997), Ecosystem Services: Benefits Supplied to Human Societies by Natural Ecosystems, *Issues in Ecology*, 2:1-16.

Dietz, K. (1993), The estimation of the basic reproductive number for infectious diseases, *Statistical Methods in Medical Research*, 2:23-41.

Dobson, A.P. (1995), The ecology and epidemiology of rinderpest virus in Serengeti and Ngorongoro crater conservation area, pp. 485-505 in A.R.E. Sinclair and P. Arcese (eds.), *Serengeti II: Research, Management and Conservation of an Ecosystem*, Chicago, University of Chicago Press.

– (2004), Population dynamics of pathogens with multiple hosts, *The American Naturalist*, 164:S64-S78.

– (2005a), Infectious Disease and Conservation Biology, pp. 67-68 in M.J. Groom, G.K. Meffe and C.R. Carroll (eds.), *Principles of Conservation Biology*, 3rd Edition, Sunderland, Mass, Sinauer Associates.

– (2005b) Monitoring global rates of biodiversity change: Challenges that arise in meeting the 2010 goals, *Phil. Trans. R. Soc. Lond. B*, 360:229-244.

Dobson, A.P., A.D. Bradshaw and A.J.M. Baker (1997), Hopes for the future: Restoration ecology and conservation biology, *Science*, 277:515-521.

Dobson, A.P. and M.J. Crawley (1994), Pathogens and the structure of plant communities, *Trends in Ecology and Evolution*, 9:393-398.

Dobson, A.P. and P.J. Hudson (1986), Parasites, disease and the structure of ecological communities, *Trends in Ecology and Evolution*, 1:11-15.

Dobson, A.P., D.M. Lodge, J. Alder, G. Cumming, J.E. Keymer, H.A. Mooney, J.A. Rusak *et al*. (2006), Habitat loss, trophic collapse and the decline of ecosystem services, *Ecology in Review*, In Press.

Duffy, J.E. (2003), Biodiversity loss, trophic skew and ecosystem functioning, *Ecology Letters*, 6:680-687.

Dunne, J.A. (2005), The Network Structure of Food Webs, pp. (In Press) in M.M. Pascual and J.A. Dunne (eds.), *Ecological Networks: Linking Structure to Dynamics*, Oxford, Oxford University Press.

Dunne, J.A., R.J. Williams and N.D. Martinez (2002), Network structure and biodiversity loss in food webs: robustness increases with connectance, *Ecology Letters*, 5:558-567.

Eisner, T., J. Lubchenco, E.O. Wilson, D.S. Wilcove and M.J. Bean (1995), Building a scientifically sound policy for protecting endangered species, *Science*, 268:1231-1232.

Homewood, K.M. and W.A. Rodgers (1991), *Maasailand Ecology*, Cambridge, Cambridge University Press.

Jenkins, M. (2003), Prospects for Biodiversity, *Science*, 302:1175-1177.

Kremen, C. (2005), Managing ecosystem services: what do we need to know about their ecology?, *Ecology Letters*, 8:468-479.

LoGuidice, K., R.S. Ostfeld, K.A. Schmidt and F. Keesing (2003), The ecology of infectious disease: Effects of host diversity and community composition on Lyme disease risk, *PNAS*, 100:567-571.

Loreau, M., S. Naeem, P. Inchausti, J. Bengtsson, J. P. Grime, A. Hector, D.U. Hooper *et al*. (2001), Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges, *Science*, 294:804-808.

Lotka, A.J. (1923), *Elements of physical biology*, Baltimore, Williams & Wilkins.

MacArthur, R.H. and E.O. Wilson (1963), An equilibrium theory of insular zoogeography, *Evolution*, 17:373-387.

– (1967), *The Theory of Island Biogeography*, v. Princeton University Press, Princeton, New Jersey.

Mann, C.C. and M.L. Plummer (1995), *Noah's Choice: The Future of Endangered Species*, New York, Alfred A. Knopf.

May, R.M. (2001), *Stability and complexity in model ecosystems*, Princeton University Press, 265 p.

Nowak, M.A. and R.M. May (1991), Mathematical biology of HIV infections: Antigenic variation and diversity threshold, *Mathematical Biosciences*, 106:1-21.

Palmer, M., E. Bernhardt, E. Chornesky, S. Collins, A. Dobson, C. Duke, B. Gold *et al*. (2004), ECOLOGY: Ecology for a Crowded Planet, *Science*, 304:1251-1252.

Palmer, M.A., E.S. Bernhardt, E. A. Chornesky, S.L. Collins, A.P. Dobson, C.S. Duke, B.D. Gold *et al*. (2005), Ecology for a Crowded Planet: a Science Action Plan for the 21st Century, *Frontiers in Ecology*, 3:4-12.

Plowright, W. (1982), The effects of rinderpest and rinderpest control on wildlife in Africa, *Symposia of the Zoological Society of London*, 50:1-28.

Plowright, W. and W.P. Taylor (1967), Long-term studies of immunity in East African cattle following inoculation with rinderpest culture vaccine, *Researches in Veterinary Science*, 8:118-128.

Prins, H.H.T. and F.J. Weyerhaeuser (1987), Epidemics in populations of wild ruminants: anthrax and impala, rinderpest and buffalo in Lake Manyara National Park, Tanzania, *Oikos*, 49:28-38.

Rogers, D.J. and S.E. Randolph (1988), Tsetse flies in Africa: Bane or boon?, *Con. Biol.*, 2:57-65.

Schmidt, K.A. and R.S. Ostfeld (2001), Biodiversity and the dilution effect in disease ecology, *Ecology*, 82:609-619.

Seabloom, E.W., A.P. Dobson, and D.M. Stoms (2002), Extinction rates under nonrandom patterns of habitat loss, *PNAS*, 99:11229-11234.

Simberloff, D., D. Doak, M. Groom, S. Trombulak, A.P. Dobson, S. Gatewood, M. Soule *et al*. (1998), Regional and Continental Restoration in M. Soule, and J. Terborgh (eds.), *Continental Conservation: Scientific Foundations of Regional Reserve Networks*, Washington, DC, Island Press.

Sinclair, A.R.E. (1979), The eruption of the ruminants, pp. 82-103 in A.R.E. Sinclair and M. Norton-Griffiths (eds.), *Serengeti: Dynamics of an Ecosystem*, Chicago, University of Chicago Press.

Sinclair, A.R.E. and P. Arcese (1995), *Serengeti II. Dynamics, Management, and Conservation of an Ecosystem*, Chicago, Chicago University Press.

Sinclair, A.R.E. and M. Norton-Griffiths (1979), *Serengeti: Dynamics of an Ecosystem*, Chicago, University of Chicago Press.

Soulé, M.E. (1986), *Conservation Biology. The Science of Scarcity and Diversity*, pp. 584, Sunderland, Massachusetts, Sinauer Associates, Inc.

Tilman, D., R.M. May, C.L. Lehman, and M.A. Nowak (1994), Habitat destruction and the extinction debt, *Nature*, 371:65-66.

Volterra, V. (1926), Fluctuations in the abundance of species, considered mathematically, *Nature*, 118:558-560.

PRE-INDUSTRIAL DEPOPULATION, ATMOSPHERIC CARBON DIOXIDE, AND GLOBAL CLIMATE

WILLIAM F. RUDDIMAN[1] & ANN G. CARMICHAEL[2]

The topic addressed by this volume – interactions among human health, global change, and socio-economic factors – is immensely broad and complex. In this paper, we focus on one of many related issues – the link between population size (one of many possible indices of human well-being), atmospheric carbon dioxide (CO_2), and climate during the pre-industrial portion of the historical interval. Specifically, we test a new hypothesis that intervals of significant human depopulation (at the scale of tens of millions of deaths) caused reforestation of abandoned farmland, and thereby reduced atmospheric CO_2 concentrations and cooled global climate.

In the first section of this paper, we focus on the major multi-regional depopulation intervals identified in historical records. We summarize the likely impacts of famine, war, and disease on depopulation and conclude that disease is the largest factor in most major depopulation intervals. We conclude that the correlation between major pandemics and intervals of decreased CO_2 supports a causal link between mass mortality and carbon levels in the atmosphere. In the second section, we outline the methods used to analyze and quantify possible pandemic-climate links. In the third section, we model the response of atmospheric CO_2 concentrations to carbon input and removal caused by reforestation, by decreases in rates of deforestation, and by decreases in early coal use. We find that reforestation was probably the major contributor to CO_2 decreases during depopulation intervals, while reductions in rates of deforestation and coal use were likely secondary factors.

[1] Department of Environmental Sciences, University of Virginia, Charlottesville, VA.
[2] Departments of History and HPSc, Indiana University, Bloomington, IN.

1. Links Between Depopulation, Disease, and Climate

All dates cited in this paper are in years CE (current era), unless otherwise specified.

1.1. *Pandemics and Depopulation*

Historical changes in human populations for Europe, China, and the Americas, as well as all other regions combined, are shown in Figure 1. Three areas – Europe, China, and the Americas – were plotted individually because they show major depopulation during several pre-industrial intervals. For most of Eurasia, these population data are from McEvedy and Jones (1978), who noted that their estimates varied by ±20% relative to other compilations based on similar sources. More recent estimates of Chinese population from Lee and Feng (1999) were used, even though in most cases continental scale estimates differ little from those in McEvedy and Jones for the period before the late 1600s. Early population estimates for several countries that are lumped together as 'other' in Figure 1 are poorly constrained: India prior to the Mughal era (1500s), and Indo-China, sub-Saharan Africa and Oceania even in later centuries.

Population estimates for the Americas are taken from Denevan (1992), who proposed an indigenous population of 55 million on the eve of European contact, followed by a precipitous drop over the next two centuries to 5 or 6 million. Formerly, such calculations depended on backward projections from Spanish and Portuguese censuses of native populations and on mortality rates extrapolated from local examples. More recently, new forms of archeological study such as air-photo and remote-sensing surveys of ancient structures built for agriculture and transportation have provided additional support for higher population estimates. The backward projection of the American population shown in Figure 1 prior to European contact is based on a 1500-year doubling time for population growth. We make no attempt to address regional fluctuations in population in the Americas before contact with Europeans, such as the collapse of Mayan civilization in the 800s and 900s or the smaller-scale contraction and displacement of the Anasazi in the 1300s.

The generally exponential rise of populations through historical time reflects a host of factors that affected people and their food supplies. Innovations in agricultural methods and technology, introduction of non-indigenous foods (plant and animal), and social policies all played a role.

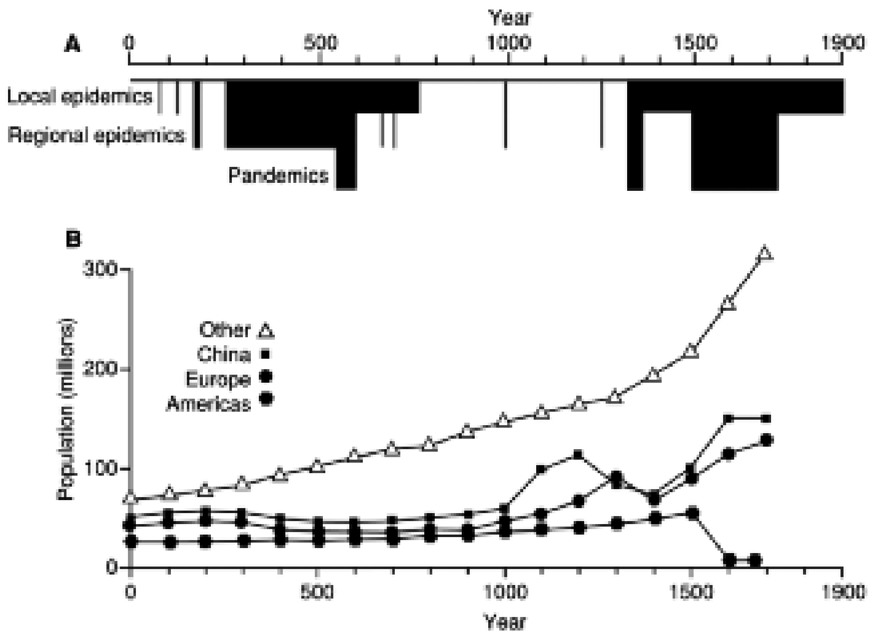


Figure 1. Link between pandemics (A) and major depopulation episodes (B) during the last two millennia. Major population drops in Europe and China from McEvedy and Jones (1978) and in the Americas from Denevan (1992).

Despite general population growth, every pre-industrial civilization also experienced intervals of stagnation or depopulation (Landers, 2004). The classical Malthusian checks of famine and disease, enhanced by warfare, are frequently assumed to have driven periods of stagnation and decline. Chinese populations uniquely limited fertility through infanticide (Lee and Feng, 1999; Lavely and Wong, 1998).

Among the intervals of decline lasting a century or more, we can distinguish in the sketchy historical records before 1800 a few dramatic episodes of population loss that were concentrated within a decade. Here we first explore the possible causes of these major depopulation episodes – famine, warfare, or acute exogenous epidemic disease.

The hypothesis that famine accounts for intervals of large-scale depopulation has largely been rejected. Even the worst famine of the last millennium in Western Europe, from 1315 to 1322, claimed no more than 10% of the population in northwestern Europe, where its effects were

strongest (Jordan, 1996; Watkins and Menken, 1985, 1988). Where not complicated by warfare or the loss of livestock to pestilence, population numbers rebounded before the catastrophic Black Death a few decades later. Population recovery occurred rapidly because famine and the additional diseases it causes primarily take the lives of the very young and the 'elderly' (at that time, those over 45 years of age). Large famines did not recur in pre-industrial European populations; only the difficult years from 1815-1818 compare in geographical extent with the 1315-1322 period. Although significant short-term famine mortality also occurred elsewhere, including large areas of China during the early 1580s and 1640s (Dunstan, 1975), the Chinese imperial state early initiated interregional transfers of food to avert famine (Marks, 1998; Shiue, 2005). By the late eighteenth century, famine crises in Europe, North Africa, and Asia had also been largely solved by regional transportation of grain and by credit financing that gave the affected populations and governments purchasing power in larger markets (Ó Gráda, 2005).

War is also not a first-order factor in most of the intervals of major depopulation, although two periods of intensive warfare merit comment. During the Thirty Years' War in Europe (1618-1648), mortality was abnormally high because civilians were frequently targets of military actions. An estimated 8 million people died in German regions and in Belgium (Gutmann, 1980), but even that number did not reach the levels during the major depopulation intervals plotted in Figure 1b. War and political upheaval, however, did precipitate major depopulation during at least one interval in Chinese history. From 1278 to 1369, the occupying Mongol rulers of northern China (the Yuan dynasty) destroyed much of the ecological and economic infrastructure, leading to catastrophic population losses in the north, as well as localized epidemics and famines. In this case, 25 to 50 million people are thought to have died. Millions likely fled south, to the southern Song dynastic region, with some estimates of regional depopulation in the once prosperous north as high as 86% (Deng, 2003).

With the exception of China, disease is the primary explanation of the great depopulation episodes of human history. Infectious and non-infectious diseases regularly pruned pre-industrial populations before the sanitary revolution of the late nineteenth century, but they did not for the most part result in major fluctuations over decades or centuries (Bourdelais, 2003; Livi-Bacci, 2001). This pruning can be regarded as a 'steady-state' phenomenon: always in action, always culling a share, but not causing discrete population drops.

In contrast, a few acute infectious diseases – principally plague in Europe and various viral rashes in the Americas – were 'pandemic' agents that led to a few great depopulation episodes. Demographically, bubonic plague affects at-risk populations indiscriminately with respect to age and gender. Plagues killed people during their reproductive years and were accompanied or followed by non-plague epidemics; the combination had a lasting negative effect on population recovery (Hatcher, 1977; Paine, 2000). Smallpox and measles introduced in the Americas had similarly extensive, multi-regional effects. Temporally these epidemics lasted a century or more, and the losses were much higher than those caused by other diseases. The impacts of three major pandemics are obvious in the population trends in Figure 1.

The first major interval of large-scale depopulation in Mediterranean and Western Europe began with the so-called Antonine plague of 165-180. Although the principal microbial cause of this epidemic is debated, population loss was geographically and demographically extensive within the Roman Empire, including populous North Africa (Duncan-Jones, 1996; Scheidel, 2001). During the centuries after 200, McEvedy and Jones (1978) estimate a cumulative population loss of 40% (~10 million people) in southern and western Europe, but lesser mortality in the more sparsely populated north and east. The interval between 200 and the early 500s was one of general depopulation, punctuated by brief periods of partial recovery. Other diseases may have been involved in these ongoing losses, as well as socio-economic factors. Sallares (2002) suggests that malaria had a substantial demographic impact on the later Roman Empire.

The 'Plague of Justinian' in 540-542 was far more severe and spread across the entire continent of Europe (Stathakopoulos, 2004). Narrative accounts of Justinian's plague suggest that bubonic plague, caused by *Yersinia pestis*, was the principal agent of recurrent epidemic crises (Sarris, 2002). Plague recurred repeatedly from the sixth to the eight centuries from the Mediterranean west to Ireland and east to the lands of Islam, and probably spread farther east along the Silk Road (Twitchett, 1979).

The 'Black Death' pandemic of 1347-1353 in Europe, the Middle East, and North Africa offers better evidence of abrupt and protracted population decline. In regions where depopulation ratios can be calculated, Black Death losses reached the 30 to 45% level (a total of 25 to 33 million people). Plague, most likely caused by *Yersinia pestis*, and other epidemic infectious diseases recurred frequently in Europe during the century following this catastrophe, and population stagnation persisted well into the

1400s (Hatcher, 1994). The region now comprising Poland, Lithuania, Latvia, and northern Russia initially remained relatively unscathed, but populations in these regions collapsed half a century later.

Regional as well as dramatic localized plagues persisted in Europe until the mid-seventeenth century, with a final epidemic in Marseilles and Languedoc in 1720-22. Specific urban plagues during the period 1550 to 1720 caused mortality rates as high as 25 to 40% in parts of southern and central Europe, the only region where plagues seemed to have caused overall declines in population after 1600. Parish registers in German-speaking regions show that smaller settlements suffered even greater depopulation than urban centers (Eckert, 1996).

Even the extraordinarily high losses from recurrent plague and pestilence in Europe and the lands of Islam were still less devastating than the effects of common 'old world' infections on the native populations of the Americas. In addition to the diseases to which European, Asian and African populations had partial immunity, Native Americans were subjected to a host of diseases introduced by European livestock. The result was unprecedented mortality – depopulation without even partial recovery. An estimated 80 to 90% of the pre-Columbian population (50 to 60 million people) died between 1500 and 1800, with the highest losses in heavily populated regions probably occurring in the 1500s and 1600s (Denevan, 1992; Richards, 2003).

The role of disease in depopulating China was less predominant than in the west (Elvin, 1993). Of the two periods of great epidemics mentioned in Chinese historical sources, that of 636-655 seems to correspond with the spread of plague eastward along the Silk Road (Twitchett, 1979). A second wave of epidemics spread through more than half of North China from 698 to 713, reaching Korea and possibly Japan. Outbreaks in the 730s are more securely identified as smallpox. In areas where the best records survive, Japan's registered population in 726 ranged from 23 to 90% below that of 609. Over the next millennium, until the 1600s, the worst losses occurred in areas of longstanding warfare. Dunstan (1975) and Chang (2002) show that the Manchu conquest of the late 1630s, displacing the Ming dynasty by 1644, was particularly troubled by smallpox epidemics. The extensive and widespread mortality among those being incorporated into this new regime prompted the first extensive treatise on large-scale epidemic diseases within traditional Chinese medicine.

The overall doubling of human global population from 400-500 million in 1500 to 850-950 million in 1800 owed much to improvements in

both regional and global transportation of food and to the availability of credit (Richards, 1997). Most of the population rise during the modern period in India and Japan (late sixteenth-seventeenth centuries) and in China and Europe (eighteenth century) occurred through consolidation and stability of political power. The Americas did not regain pre-Columbian population levels until the nineteenth century, and this only by virtue of European immigration and expansion.

1.2. *Linking Pre-Industrial Depopulation Intervals, CO_2, and Climate Change*

Recent advances in climate science have provided valuable new information on climate change during the historical era. The first proxy-based reconstruction of temperature spanning the last millennium at a hemispheric scale was that of Mann *et al*. (1999). This and subsequent efforts have been based on records in climatic archives such as ice cores, tree rings, and corals. Dozens of records have been analyzed to extract a common temperature signal weighted so as to avoid bias from the uneven spatial distribution of the records.

Several other reconstructions of northern hemisphere temperature changes during the last millennium show similar patterns but varying amplitudes (Crowley and Lowery, 2000; Jones *et al*., 2001; Jones and Mann, 2004). In higher (Arctic) latitudes, reconstructed temperature changes over decadal intervals are generally larger by a factor of 2 or 3 than hemispheric averages (Overpeck *et al*., 1997; Esper *et al*., 2002). This poleward amplification is consistent with the larger temperature variability observed today during cold seasons near sensitive snow and sea-ice boundaries. Moberg *et al*. (2005) recently reconstructed changes in temperature much larger than those in Mann *et al*. (1998) and Crowley and Lowery (2000). But the disproportionate number of proxies from polar latitudes used in the Moberg reconstruction may invalidate it as a representative hemispheric-mean signal.

The first attempt to extend the northern-hemisphere temperature reconstruction towards the start of the historical era in Europe is shown in Figure 2a (Mann and Jones, 2003). The shaded region indicates the large uncertainties resulting from the sparse coverage of sites and the imprecise links of the proxy indicators to temperature. A comparison of the trends in Figures 1b and 2a hints strongly at a correlation between cooler climates and intervals of major regional depopulation. Cooler temperatures prevailed during the slow European population decrease between 200 and 800,

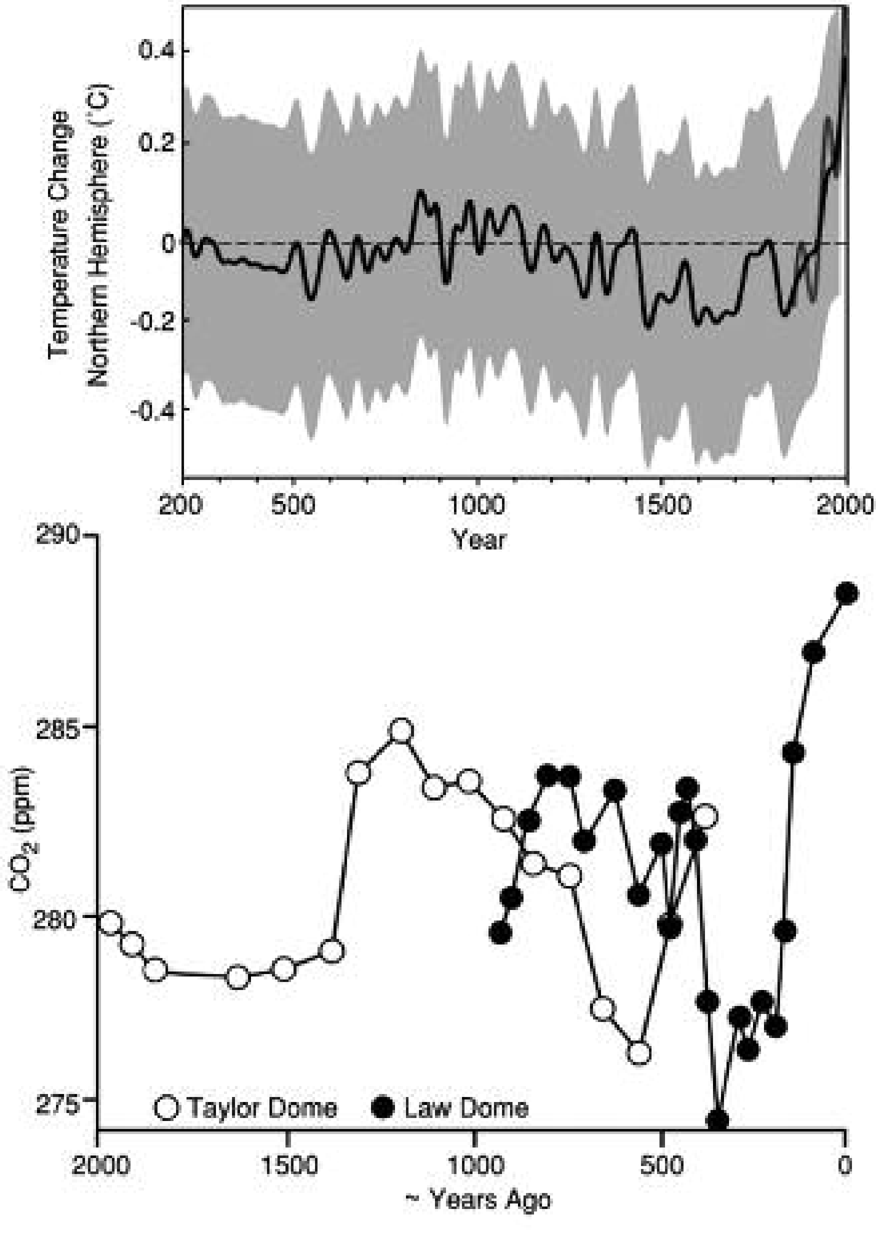


Figure 2. Reconstructed trends in northern hemisphere temperature (top) from Mann and Jones (2003) and atmospheric CO_2 concentrations (bottom) from Etheridge *et al.* (1996) and Indermuhle *et al.* (1999).

and again from 1500 to 1750, the time of the American pandemic. Warmer conditions prevailed during the intervening pandemic-free era.

Given this apparent link between intervals of major depopulation and large-scale climate change, an obvious question is what kind of mechanism might account for it. The greenhouse gas, carbon dioxide (CO_2), plays an important role in climate change, and its past concentrations in the atmosphere can be measured in air bubbles trapped in ice drilled in Antarctica. Plotted in Figure 2b are records of CO_2 concentrations from two sites. The record from Law Dome (Etheridge *et al.*, 1996) is the better dated because it contains layers of volcanic ash particles from explosions of known age (Palmer *et al.* 2001). Dating of the record from Taylor Dome (Indermuhle *et al.*, 1999) is far less secure: it is based on correlating the methane signal at Taylor Dome to that of an annually layered (and thus well dated) ice core from Greenland. Uncertainties for the Taylor Dome record derived in this way are estimated at ~500 years prior to 1000, and ~100 years during the most recent millennium.

The conventional interpretation of these historical CO_2 trends is that they are driven by natural variations in the frequency of volcanic explosions and by small changes in the radiative output of the Sun over decadal and century time scales (Crowley, 2000; Gerber *et al.*, 2003). In this view, both the CO_2 oscillations and the temperature changes shown in Figure 2 are natural responses of the climate system to the same solar-volcanic forcing.

This explanation can be evaluated with carbon cycle models that quantify both the size of the CO_2 oscillations and the amount of global (or hemispheric) temperature change that would result from solar-volcanic changes (Gerber *et al.*, 2003). The models indicate that solar-volcanic forcing sufficient to cause a CO_2 oscillation of 1 ppm (part per million) should produce a temperature response of 0.08°C. Gerber *et al.* (2003) compiled a composite CO_2 curve by stacking and averaging individual CO_2 records from four ice cores. Because several of the largest CO_2 oscillations are offset among individual cores (as in Figure 2b), this procedure yielded a smoothed record with maximum CO_2 variations of ~4 ppm. This value seemed roughly consistent with the CO_2 change that the model predicted should accompany the temperature variations in Figure 2.

Ruddiman (2003) viewed these CO_2 records in a different way. He noted that both ice-core records in Figure 2b have CO_2 oscillations as large as 8 ppm, as did an early record from Adelie Dome (Barnola *et al.*, 1995). He thus proposed that the large CO_2 changes are real but that inadequate dating control caused these oscillations to be misaligned (offset in age) from record to record.

If the CO_2 drops really were as high as 8 ppm, the solar-volcanic explanation cannot account for them without violating the constraint imposed by the small changes in reconstructed temperature. For example, the average cooling between the intervals 900-1100 (the 'Medieval warm interval') and 1500-1800 (the 'Little Ice Age') is no more than 0.3ºC (Figure 2a), consistent with a CO_2 drop of ~3.5 ppm. Yet the CO_2 drop between those two intervals in the well-dated Law Dome ice core is ~8 ppm (Figure 2b), leaving 4 to 5 ppm of the observed change to be explained by factors other than natural forcing.

Subsequent work has further highlighted the shortcomings of natural factors as an explanation of these CO_2 decreases. CO_2 changes in well-dated ice from Dronning Maud Land and South Pole confirm a CO_2 drop of at least 8 ppm from 1100 to 1700 (Siegenthaler *et al*., 2005). Those authors also concluded that the Little Ice Age cooling, which is small in amplitude and centered in the North Atlantic region, is unlikely to explain so large a CO_2 drop. This apparent failure of the solar-volcanic explanation to account for the large observed changes in CO_2 invites a different explanation.

1.3. *The Pandemic-CO_2-Climate Hypothesis*

An alternative explanation for the large CO_2 oscillations-human activities arose as a corollary to a recent hypothesis. Ruddiman (2003) pointed out that the gradual rise in CO_2 and methane values during the last several millennia is anomalous compared to trends during similar intervals in the early portions of previous interglaciations (Figure 3, see page 410). Previously, CO_2 values rose to a strong peak just before the start of each interglaciation and then declined steadily through the next 10,000 to 15,000 years of each early-interglacial interval. Early in the present interglaciation, CO_2 reached a similar peak 10,500 years ago and began to decline as it had before. Beginning around 8000 years ago, however, a large CO_2 increase occurred that had no precedent in previous interglaciations.

Ruddiman (2003) proposed that humans were responsible for this anomalous CO_2 increase, primarily by clearing and burning forests in southern Eurasia to open land for farming. Carcallet *et al*. (2002) had independently inferred a possible human role in the CO_2 increase of the last 8000 years, based on increased deposition rates of charcoal on several continents. Ruddiman (2003) also suggested that major increases in rates of human mortality might be the explanation of the century-scale CO_2 oscillations near the end of the Holocene CO_2 rise. He proposed that pandemics –

outbreaks of bubonic plague and other diseases on a multi-continental scale – are the most plausible mechanism for linking changes in human populations, atmospheric CO_2, and global climate. The link proposed was large-scale reforestation of farms abandoned as a result of pandemics, and subsequent sequestration of atmospheric carbon in the growing forests.

The correlations among pandemics, depopulation episodes, decreases in CO_2, and cooler temperatures in Figures 1 and 2 support this explanation, particularly if allowance is made for the uncertainties inherent in historical estimates of human mortality, in dating the CO_2 signals, and in the pattern and amplitude of the temperature reconstruction. In addition, the evidence from China discussed earlier further indicates that at least one depopulation episode (during the Yuan dynasty, 1278-1369) could have resulted from conquest and social policy. The population losses in that episode also had the potential to affect the global carbon budget because prior to population collapse, northern China had turned to coal as a primary fuel source, having exhausted the accessible forest reserves (Hartwell, 1962, 1967).

If depopulation episodes contribute to changes in atmospheric CO_2 concentrations, it makes sense to ask whether or not the observed CO_2 changes could have driven the reconstructed changes in northern hemispheric temperature. Based on a widely used estimate of Earth's temperature sensitivity to CO_2 changes (IPCC, 2001), the CO_2 drop of 8 ppm between the maximum at 1100-1200 and the minimum at 1500-1800 should have caused global and hemispheric temperature to fall by ~0.1°C. Over this same interval, the average temperature decrease (using century-scale averages and ignoring decadal-scale variations) was 0.15-0.2°C (Fig. 2a). As a result, the observed CO_2 decrease appears to explain half or more of the drift toward colder temperatures from the warmer medieval era to the peak expression of the so-called little ice age (1550 to 1800).

In summary, the model simulation of Gerber *et al.* (2003) indicates that natural forcing cannot explain a CO_2 drop of 8 ppm during the last millennium without violating the small temperature cooling reconstructed for the northern hemisphere. A factor new to the 'natural' workings of the climate system – human history – appears to have played a significant role in both the CO_2 and temperature changes of recent millennia. This hypothesis immediately raises the question of whether human populations really were large enough centuries or millennia ago to have had an impact on atmospheric CO_2 concentrations. The purpose of the rest of this paper is to make a quantitative assessment of this idea.

1.4. *CO_2 Target Signal for Testing the Pandemic-CO_2-Climate Hypothesis*

The CO_2 trends at Taylor and Law Domes from Figure 2b are combined into a composite target signal in Figure 4 (see page 410). The well-dated Law Dome record subsequent to 1100 years ago serves as the target signal. Prior to that time, we are forced to rely on the less securely dated record from Taylor Dome. The intervals of large-scale depopulation discussed earlier correspond to three periods of CO_2 decreases in Figure 3: (1) a small but long-lasting drop of ~2 ppm during the epidemics and pandemics of the late Roman Era between 200 and 600 AD; (2) a short ~2 ppm decrease during the Black Death (bubonic plague) pandemic of 1350 and subsequent outbreaks in the next century; and (3) an abrupt drop of ~5 ppm soon after the start of the American pandemic era in the 1500s.

For the analysis that follows, we need to consider how faithfully the ice cores might record very rapid changes in atmospheric CO_2. The initial Black Death pandemic (1347-1353) represents the most extreme case: large changes in population occurred within a few years, and resulting changes in carbon emissions and sequestration within a few decades. In contrast, the large changes between 200 and 600 and between 1500 and 1750 occurred over centuries.

Because of exchanges of air in the uppermost layers of snow, firn, and ice, CO_2 values recorded in ice cores have been smoothed over intervals of decades. Law Dome ice was deposited rapidly, and the estimated smoothing interval is short enough (~20 years, Etheridge *et al.*, 1996) to register even short-term changes in atmospheric CO_2. As a result, the Law Dome CO_2 values in Figure 3 are well suited for analyzing the effects of the Black Death and American pandemics on CO_2 and climate. The smoothing interval for the slowly deposited Taylor Dome ice is estimated at ~140 years (Indermuhle *et al.*, 1999). This resolution should suffice for analyzing the slow-developing and long-lasting depopulation that occurred from 200 to 600.

2. Methods Used to Test the Depopulation-CO_2-Climate Link

This section describes methods and assumptions used to evaluate the hypothesis that changes in carbon emissions caused by depopulation affected CO_2 oscillations during the last two millennia (Ruddiman, 2003). This analysis requires two key links: (1) quantifying the amount of carbon sequestration on abandoned land produced by the major intervals of depopulation; and (2) translating this carbon sequestration into changes in atmospheric CO_2.

2.1. *Deforestation and Reforestation: Human 'Forest Footprints'*

Deforestation by humans occurred in two major phases that had different impacts on the environment. The first phase, here termed agricultural deforestation, was the clearing of old-growth forests to open land for croplands and pastures. This phase accounted for the largest part of forest removal (Williams, 2003). In addition to making new land available to grow food, habitat of dangerous mammalian predators was reduced. Because the negative environmental impacts of this kind of clearance were relatively small and not much noted at the time, this early phase of deforestation received little comment from contemporary observers, particularly because much of it preceded the advent of written records. This phase of deforestation was featured in the hypothesis of Ruddiman (2003) and is the focus of this paper.

The second phase of deforestation, here referred to as resource deforestation, pertains to forested areas that remained on higher steeper terrain when most arable land had been cleared. Wood taken from these regions was used for building homes and ships, for cooking and heating, and for smelting of iron as societies advance in technological sophistication. Each of these uses of wood was roughly an order of magnitude smaller than the earlier losses to agricultural clearance (Williams, 2003). In many cases, the remaining forests became managed woodlots that were cyclically harvested.

This later phase of resource deforestation was widely noticed by contemporary observers, in part because the impending loss of the residual forests seemed threatening, and in part because of highly visible environmental damage (Bechmann, 1990; Elvin, 1993). Without forest cover, rainwater tended to run off steep slopes rather than being absorbed into the soil and water table. Hillsides eroded, rivers carried mud-laden runoff, and natural springs disappeared because of reduced ground-water flow. These two phases of deforestation overlap to some extent. Even in the early phases of agricultural deforestation, some timber was used for home building and some wood was burned for heating and cooking. But clearance of arable land was the predominant form of deforestation.

History provides a vital 'data point' for the analysis that follows. In 1086, William the Conqueror ordered a survey of his newly won land (England) that by chance gives a snapshot of an iron-age country very near the transition between the two phases of deforestation. The survey results, reported in 1089 in the Domesday Book, counted 1.5 million citi-

zens and determined that 90% of the arable land was in pasture or croplands. A meticulous analysis by Rackam (1980) based on wide-ranging historical and ecological methods confirmed both the pattern and extent of deforestation indicated by the Domesday survey. Perhaps as a result of the Domesday survey, William's heirs enacted laws restricting the access of commoners to the remaining forests.

The Domesday Book provides a basis for two kinds of quantitative estimates used in this study. The Domesday census and land-use numbers indicate that a population density of ~11 persons/km² was sufficient to deforest 90% of the arable land in England in the 11th century. With this value in hand, we can use historical population data from McEvedy and Jones (1978) and compilations of the amount of arable land (lying at <1km elevation) to estimate the time at which other regions inhabited by iron-age people reached a population density of 11/km² and thus (arguably) the point of full deforestation of arable land. Such an assumption is of course, crude, but in many regions it can be evaluated against historical and ecological sources.

Table 2 shows the estimated time when heavily populated regions of Europe and Asia would have reached the Domesday population threshold. In Europe, numerous sources suggest that most arable land in the Greco-Roman centers of early Mediterranean civilization was deforested by the start of the historical era (Hughes, 1975; Thirgood, 1981; Simmons, 1996; Roberts, 1998; Williams, 2003), in agreement with the estimates in table 2. Part of the expansion of the early Roman Empire entailed a search for timber to build ships and maintain dominance in the Mediterranean Sea.

Table 2 indicates that much of Western Europe reached the Domesday threshold between 1000 and 1300, an interval of rapid population growth. These estimates are consistent with data on forest losses in France and Germany between 1100 and 1350 (Bechmann, 1990) and with laws enacted to protect forest preserves in those regions during the 1200s and 1300s (Williams, 2003). In contrast, Scandinavia and countries in the easternmost Baltic were never fully deforested during medieval centuries. Exploitation of timber resources in those regions only occurred in the early modern period, when shipbuilding demands in Western Europe expanded.

We assume that the Domesday density threshold is also applicable to northern and central China where 'dry' farming of grain crops (primarily millet) and pasturing of livestock predominated, as in Europe. Chao (1986) reported that only a little more than 1 ha of land per capita was

tilled as of 2000 years ago but his estimates omitted pastures, hay fields, woodlots, dwelling areas, and (in some cases) fields left fallow between intervals of cultivation. In any case, the Yellow and Yangtze River valleys both passed the Domesday density threshold well before 2000 years ago, and historical ecologists have concluded that they were largely deforested at an early date (Hughes, 1975; Grigg, 1994; Simmons, 1996; Williams, 2003; Elvin, 2004). Historical evidence also suggests very early (pre-historical) deforestation of dry seasonal forest in grain-growing areas of the Indus River valley (Fairservis, 1971; Hughes, 1975).

It is unlikely that the Domesday density threshold can be applied to the rice-growing areas of southern China, the Ganges River area of northeastern India, and intervening lowland areas of southeastern Asia. In these regions, irrigation permitted very high food production per unit area, and people probably had a smaller per-capita impact on forested lands. Nevertheless, Chao (1986) reported that much of the deforestation of South China for agriculture was already confined to mountainous areas by the eighth to eleventh centuries, implying that lowland (rice-growing) regions were already fully occupied by rice farmers. The estimated date of 1100 for agricultural deforestation in this area (Table 2) agrees with this analysis (McNeill, 1997).

In any case, rice consumption favored such rapid population growth that population densities in such regions soon grew to several times the Domesday threshold. As a result, very early deforestation of the Ganges River valley of India and Bangladesh (prior to 2000 years ago) has been widely deduced by several studies (Fairservis, 1971; Hughes, 1975; Grigg, 1994; Simmons, 1996; Williams, 2003). Other regions in Southeast Asia (Japan, Indonesia, and Korea) appear to have passed the Domesday population density threshold by ~1000 or soon after (Table 2).

The Domesday density value of 11 people/km^2 is useful in a second way. Inverted, it yields an average per-capita 'forest footprint' for each iron-age human of 9 hectares (9 ha, or 0.09 km^2). Each human clears that much land, primarily for agriculture. This average per-capita value implicitly allows for the small fraction of England's population that was not engaged in farming and also for the fact that some members of farm families did not personally clear any land. We assume that Europeans living in iron-age cultures and in regions where the arable land had not yet been deforested occupied a farmland footprint of 0.09 km^2 prior to the onset of the late-Roman depopulation interval and the Black Death pandemic waves.

In contrast, the depopulation episode caused by the American pandemic between 1500 and 1750 killed people who were living in a late Neolithic ('stone-age') culture. The per-capita forest clearance for these people was lower than that for iron-age cultures.

Gregg (1988) estimated the amount of land needed by a typical Neolithic European village for crops, pastures, livestock, and fuel wood. She arrived at a per-capita estimate of 3 ha (0.03 km^2) of cleared land. Most of this amount represented pastures and hayfields. Woodlot needs were smaller, and croplands smaller still. The three-fold increase between this 3-ha estimate and the 9-ha Domesday value presumably reflects the intervening appearance of iron-age axes and plows, as well as the domestication of draft animals that dislodged stumps and pulled plows.

Estimates of the per-capita 'forest footprint' in the pre-Columbian Americas range between a few tenths of one ha to more than 2 ha (Maxwell, 1910; Heidenrich, 1971; Hurt, 1987; Williams, 2003). Values toward the lower end of this range tend to focus only on areas under 'current' cultivation for major crops and thus omit other factors: fields that lie fallow, frequent movements to new plots, and the use of forest wood for fuel.

Another reason this range is lower than the 3-ha value for Neolithic Europe is that most indigenous Americans kept no livestock and had no need of pastures and hay fields. But, presumably to fill their nutritional needs, they made wide use of fire to open and maintain clearings so that they could attract game and encourage growth of berries, nuts, and other foods. Although the extent of burning is difficult to estimate (Vale, 2002), it may have been one of the primary alterations of the landscape in pre-Columbian North America.

Plausible allowance for areas that were burned repeatedly (and thus kept 'deforested') would presumably bring the forest-footprint estimates for the Americas more in line with that from late Neolithic Europe. We assume here that every indigenous American occupied a forest footprint of 2 ha (0.02 km^2), half or less in croplands and woodlots, and the balance in areas burned repeatedly (and thus kept open). These forest-footprint estimates (9 ha for iron-age Europe and 2 ha for late Neolithic America) bracket estimates of 5-8 ha for primitive modern cultures living in conditions intermediate between a stone-age and iron age existence (Lawrence *et al*, 1998; personal communication, 2004).

To convert these 'forest footprint' estimates of early farmers to specific amounts (tons) of carbon, we need to know the average carbon density of specific types of forest in various regions. Houghton (1999) estimat-

ed the amount of carbon released by converting specific types of forest to pasture or cropland. The values ranged from less than 1000 tons per km^2 for dry forest to almost 3000 tons of carbon per km^2 for rain forest. We use Houghton's estimates to calculate the amount of carbon sequestered per unit area when cropland or pasture reverted to the type of forest natural to each area.

To estimate the total amount of carbon sequestered on abandoned farms, we calculate the product of the number of people killed during a pandemic interval, the average 'footprint' of farmland occupied per person, and the carbon content per unit area of land that reverts to forest:

Total carbon sequestered = (# people) x (km^2/person) x (tons carbon/km^2).

The very large amounts of carbon involved in these calculations are referred to as GtC, or billions of tons of carbon (1 Gigaton=1 Petagram=10^{15} g).

2.2. *Converting Changes in Atmospheric Carbon to Atmospheric CO_2 Anomalies*

The second major step in assessing the effects of major depopulation episodes is to transform estimates of the amount of carbon sequestered in reforestation into net changes in atmospheric CO_2 concentration. If all of the carbon sequestered by the vegetation were only exchanged with the atmosphere, the exchanges would occur in the ratio 2.13 GtC/1ppm CO_2, based on the weight of CO_2 in the atmosphere. For example, the pre-industrial CO_2 concentration of 282 ppm is equivalent to 600 GtC (282 ppm x 2.13 GtC/ppm = 600 GtC).

In nature, however, exchanges of terrestrial carbon are not limited to the atmospheric reservoir. Rapid exchanges occur between the atmosphere, the surface ocean, and the vegetation and soil reservoirs over years to decades to centuries, while exchanges between these surface reservoirs and the deep ocean occur over centuries to millennia.

If a sudden decrease in terrestrial carbon emissions occurs, it will produce an abrupt drop in atmospheric CO_2, but within 50 years the signal in the atmosphere will have relaxed nearly half way back toward the pre-input concentrations (Fig. 5; based on Fig. 1 of Joos *et al.*, 2004). The later phase of the relaxation is slower much because of the slow carbon exchanges with the deep ocean and the even slower exchanges between deep water and sea-floor $CaCO_3$. Even after several millennia, an estimated 10% of the initial carbon perturbation still remains in the atmosphere. In summary, the CO_2 concentration of the atmosphere at any 'current'

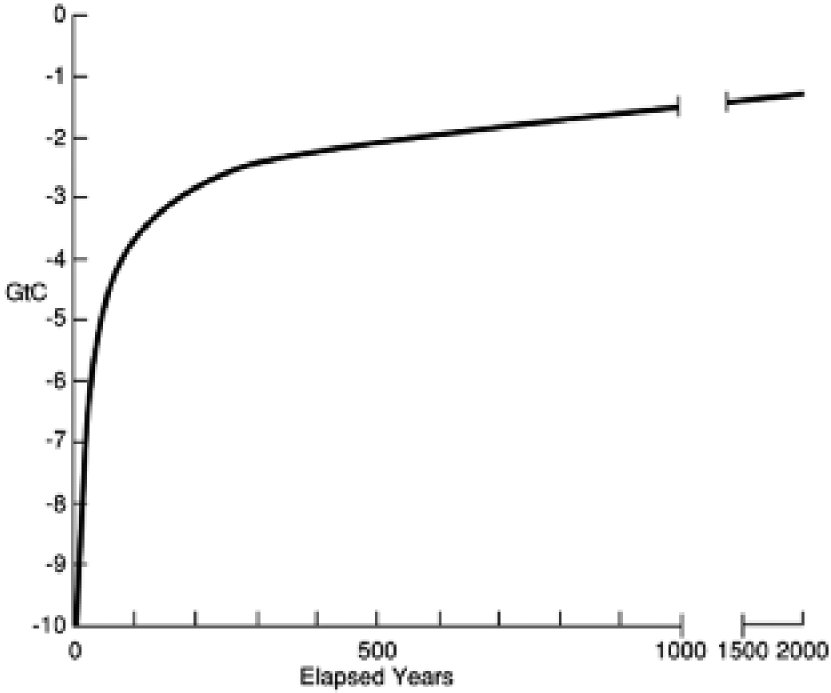


Figure 5. Impulse response based on Joos *et al*. (2004) showing the effect of carbon exchanges on the carbon content of the atmosphere. After sudden removal of 10 GtC, the atmospheric perturbation decreases rapidly for decades, then much more slowly for centuries and millennia.

time is significantly effected by carbon emissions in preceding decades, but it also retains a significant 'memory' of emissions from centuries ago.

To model the changes in atmospheric carbon and CO_2 concentrations that would occur in response to carbon sequestration during historical intervals of major depopulation, we use historical population data, human 'forest footprint' estimates, and carbon-density information to establish the time-varying pattern of carbon sequestration. Because most historical population data are available at no better than multi-decadal resolution, and because ice cores smooth changes in atmospheric CO_2 over similar intervals, we estimate changes in 50-year time steps.

We then move the impulse response function shown in Figure 5 across these time-varying estimates of carbon sequestration. The pertur-

bations caused by carbon sequestration after pandemics and other depopulation episodes die off in the same manner as the function in Figure 5: quickly at first and then much more slowly for centuries to millennia. This pattern simulates the operation of complex carbon-climate models by distributing carbon quickly among fast-reacting surface reservoirs and then more slowly with the deep-ocean and ocean-sediment reservoirs. The filter moves forward in time steps of 50 years, equivalent to the resolution of the historical depopulation estimates.

3. Modeling the Effect of Depopulation on Carbon Budgets

Three possible links exist between intervals of large-scale depopulation and decreases in atmospheric CO_2. As noted in sections 1 and 2, pandemics may have caused reforestation of abandoned farms, thus sequestering large amounts of carbon, and driving atmospheric CO_2 to lower values (Ruddiman, 2003). Here we use the approach outlined in the previous section to evaluate the effects of reforestation.

In addition, we evaluate the potential impact of two additional factors on atmospheric CO_2 concentrations – reduced rates of global-mean deforestation during and after the three major pandemics, and reduced rates of pre-industrial coal burning during the major depopulation of China in the 1200s and 1300s. Given the uncertainties involved in all of these analyses, the results presented are best regarded as a first-order test of the plausibility of these links, rather than as a definitive quantification of their size.

3.1. *Reforestation*

Ruddiman (2003) cited evidence from historians and historical ecologists for widespread abandonment of farms and farm villages during the two plague pandemics in Europe (McNeill, 1976; Rackam, 1980; Thirgood, 1981; Taylor, 1983). Reforestation is a foregone conclusion for the American pandemic, given the mortality rates of 85-90% (Denevan, 1992; Williams, 2003).

Evidence from land-use modeling (Houghton, 1999) indicates that forests will reclaim abandoned farmland in 50 to 100 years as dense stands of carbon-rich brush and saplings. Well within 100 years, the vegetation regains the equivalent carbon density of a mature forest. If successive waves of pandemic disease recur over a half-century or more,

large amounts of carbon will continue to be sequestered on land that had previously been devoted to agriculture.

Reforestation occurs only if no major population ‘excess’ exists to reoccupy the abandoned farms. If an excess does exist, farms are unlikely to revert to nature because other people soon resettle them. Based on the estimates in Table 1, we eliminate north-central China from this reforestation analysis, despite the large population drops shown in Figure 1a. The high population densities in northern and central China since before 2000 years ago should have provided a large excess of people to occupy abandoned land. Population densities in southeastern China were also high enough by 1000 years ago to leave it out of consideration.

For Europe during the late Roman Era and Black Death pandemics, we use estimates of pandemic mortality from McEvedy and Jones (1978), who subdivided population estimates for Europe along the boundaries of modern countries. We use the footprint of 9 ha (0.09 km^2) for European iron-age cultures. For the interval of depopulation spanning the late Roman era and subsequent centuries, we assume that each death led to per-capita farm abandonment in all of Europe except Greece and Italy, the only two regions that had already surpassed the Domesday population threshold (Table 2). For these regions, we assume no farm abandonment.

By the time of the Black Death, southwestern Europe had reached the Domesday threshold (Table 2). We assume that all farms in this region were re-occupied because population densities exceeded the threshold, but that none were re-occupied in northeastern Europe, Scandinavia and European Russia, where densities were low. Williams (2003) showed substantial reforestation of the northern plains of Germany after the Black Death.

For the American pandemic, Denevan (1992) subdivided pre-Columbian American populations at continental to subcontinental scales. We use the 2 ha (0.02 km^2) forest footprint for the American Neolithic cultures. In view of the 85-90% mortality rate for the American pandemic, we assume no reoccupation of abandoned regions, except for Aztec Mexico. In that region, the Spanish moved in quickly with their livestock, thus preventing major reforestation in lower-lying areas, but not in more heavily forested high terrain. We assume a reforestation rate of 75% for Mexico.

The product of the total mortality, the human ‘footprint’, and the forest-specific carbon sequestration yields estimates ranging from ~8 to ~14 Gt of carbon sequestration for the three pandemics (Table 3). These totals need to be allocated as time-varying amounts of sequestration across the duration of each pandemic.

McEvedy and Jones (1978) estimated that the population of Europe decreased nearly linearly between the geographically extensive Antonine 'plague' in 165-180 and the severe plague of Justinian in 540-542. Afterward, populations remained low for several centuries as lesser plague outbreaks continued. To approximate this continuous loss of population (and reforestation of abandoned land), we apportion the estimated 7.7 Gt of sequestered carbon from reforestation (Table 3) evenly across the interval 200 to 600 at a rate of 0.96 GtC per half century. We add a lag of 50 years to allow for the time required for abandoned farms to be covered by young forests.

The model simulation indicates that the atmosphere would have registered a maximum decrease of 2.2 GtC in the atmosphere by the interval 600-650 (Fig. 6a). Because the worst phase of this depopulation episode ended by 600, the major phase of reforestation should have come to an end by that time as well, even though plague outbreaks continued afterward. In the following centuries, carbon levels in the atmosphere slowly rose toward their prior concentration.

The severe Black Death outbreak from 1347 to 1353 was followed by smaller recurrences of plague over the next century. Of the total of 8.2 GtC estimated to have been sequestered during this pandemic (Table 3), we allocate 6 GtC to the interval 1350-1400 and the other 2.2 GtC to the interval 1400-1450. The model simulation (Fig. 6a) indicates that the atmosphere would have registered a maximum decrease of ~3.2 GtC between 1400 and 1450. For a brief interval between 1450 and 1500, following the Black Death but before the start of the American pandemic, the model simulates a brief rebound in atmospheric carbon (Fig. 6a).

Historical evidence suggests that the effects of European diseases were most devastating during the early phase of the American pandemic. Very high mortality is recorded by the 1520s and 1530s among the Aztecs, Incas, and populations in river valleys of southern North America (Denevan, 1992; Newson, 2001). We allocate the estimated 13.8 Gt of total carbon sequestration during the American pandemic as follows: 5 GtC in the half-centuries from 1550 to 1600 and from 1600 to 1650, and 1.9 GtC in the intervals from 1650 to 1700 and from 1700 to 1750. The model simulates a maximum decrease in atmospheric carbon of ~4.3 Gt from 1600 to 1650 (Fig. 6a). Much of this anomaly lingers into the 1700s and 1800s, by which time Europeans had begun to settle in the Americas in large numbers and deforest (or 're-deforest') many regions that had already been deforested prior to the pandemic.

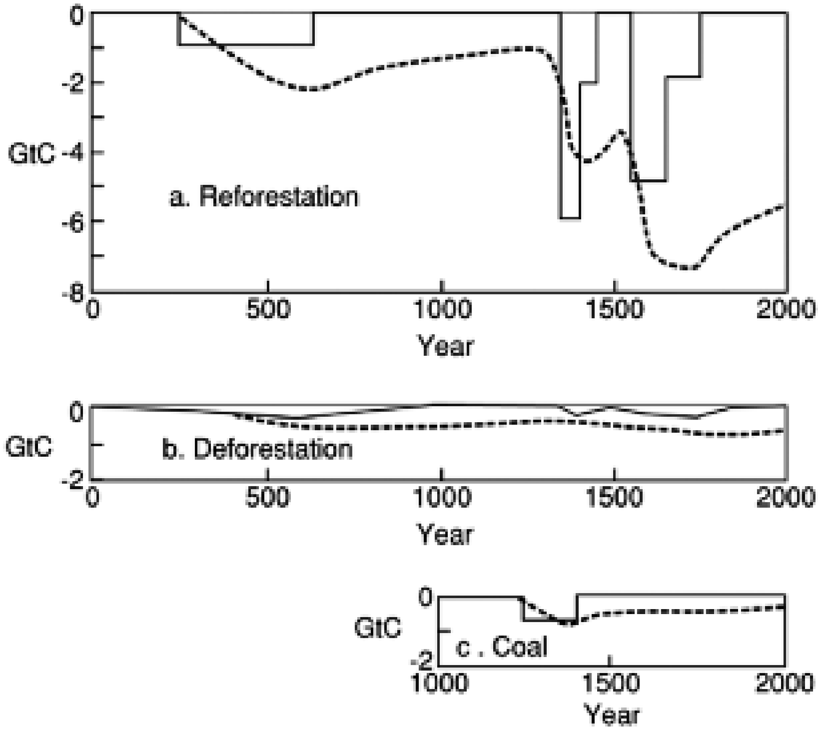


Figure 6. Model simulation of effects of major depopulation intervals on atmospheric carbon. Solid lines are estimated decreases in atmospheric carbon based on historical and ecological data. Dashed lines are model simulation of effects on the atmosphere after exchanges with other carbon reservoirs. Reforestation (A) has the largest effect on carbon reductions, with smaller effects from reduced deforestation (B) and coal burning (C).

3.2. *Reduced Deforestation*

In addition to causing reforestation, pandemics can also affect carbon budgets by reducing the rate of global deforestation. By killing a sizeable fraction of the humans that had previously been cutting forests, pandemics reduce the amount of carbon emitted to the atmosphere.

It is not possible to make region-by-region estimates of rates of deforestation during the historical era prior to industrial times. Such estimates would require detailed knowledge of growth rates of each regional population through time, yet population estimates for these times are inexact.

Calculating first derivatives (rates of change) of these already-uncertain numbers would greatly amplify the errors.

We can, however, make a very crude estimate of the reductions in deforestation rates by taking an indirect approach. First, we estimate the mean global rate of carbon emissions from deforestation across the last two millennia. For this, we use the estimate from Ruddiman (2005) that anthropogenic emissions prior to the industrial era amounted to a total of ~200 GtC. The rising CO_2 concentration in the atmosphere shown in Figure 3 can be used as a guide to partition this 200 GtC total across the last 8000 years. We assume that emissions began at 0 GtC/century 8000 years ago and rose to ~4 GtC/century by 2000 years ago. It then fluctuated around that value until the start of the industrial era. These estimates yield total emissions of just under 200 GtC for the pre-industrial era. The value of 4 GtC/century fits within estimates from Joos *et al*. (1999) of terrestrial carbon emissions (natural plus anthropogenic) during the last millennium.

Next we estimate the fractional reduction in global-mean carbon emissions caused by each pandemic. For simplicity, we assume that deforestation ceased entirely in pandemic-stricken regions because of pervasive mortality and related social disruption. We use Table 2 as a guide to eliminate regions that had previously reached the end of the agricultural phase of deforestation. In Table 4, we list the populations of all naturally forested regions that were still being actively cleared for agriculture at the time each pandemic struck. We calculate the size of the pandemic-stricken population as a percentage of this 'global' estimate. Finally, we allow for the fact that the first two (European) pandemics struck iron-age cultures, while the last (American) one ravaged a stone-age (Neolithic) culture. We assume that the rate of forest clearance in stone-age cultures was 22% of that in iron-age cultures, in proportion to the ratio of the iron-age and stone-age per-capita footprints (9 ha and 2 ha) and we adjust the calculations to incorporate this factor.

Using this method, we estimate fractional reductions in global carbon emissions at the peak of each pandemic: 13% for the late Roman era depopulation, 5% for the 14th century (Black Death) pandemic, and 14% for the American pandemic (Table 4). We then multiply the global-mean rate of carbon emission from deforestation (4 GtC/century) by the maximum fractional reduction in each pandemic to estimate the (maximum) net reduction in carbon emissions from deforestation. The values calculated are: ~0.5 GtC/century for the pandemic that began in the Roman Era, 0.2 GtC/century for the Black Death pandemic, and 0.6 GtC/century for the American pandemic.

We distribute these carbon-emissions reductions across the span of each pandemic interval in Figure 5b based on the mortality rates estimated by McEvedy and Jones (1978) and Denevan (1992). The effect of deforestation on carbon budgets differs from that of reforestation in two respects. First, the reductions in carbon emissions to the atmosphere are immediate, unlike the multi-decadal delay of carbon sequestration due to slow reforestation and carbon uptake after the pandemics. Second, the reduced rate of carbon emissions continues until the stricken population recovers. After the late Roman Era pandemic, European populations did not fully recover until 1000 (McEvedy and Jones, 1978). In contrast, recovery was well underway by 1500, only 150 years after the Black Death (Landers, 2004). In the case of the American pandemic, native populations had not even begun to recover by 1800 when Europeans arrived in large enough numbers.

In general, the simulated impact of reduced deforestation on the carbon budget of the atmosphere is considerably smaller than that from reforestation (Fig. 6b). Carbon reductions for all intervals in the last 2000 years are tenths of a GtC, compared to estimated values of several GtC for reforestation. Deforestation contributes a small amount to carbon reductions in the atmosphere during and after pandemics.

3.3. *Reduced Burning of Coal*

A third factor that could have affected pre-industrial atmospheric CO_2 concentrations is burning of coal in China. Because of early deforestation in northern and central China, coal was used as a source of fuel for heating and cooking long before the pre-industrial era. People extracted coal from surface outcrops and shallow pit mines using iron tools and then burned it in small braziers (hibachi-like stoves) in their homes. By the middle of the Song Dynasty (1040-1200), burning of coal was common in the north, while firewood was rationed and charcoal making forbidden. Hartwell (1962, 1967) estimated that all domestic fuel by the twelfth century was coal (apparently referring to northern China). Most of the documented coal use was domestic, although the Chinese may have begun using coal to produce iron.

China experienced substantial depopulation in the early centuries of the modern era, but the magnitude of these decreases is difficult to estimate because fluctuations in registered populations could simply reflect collapse of the census system. We restrict the focus of this section to the Mongol era from 1279 to 1368 (Figure 1b). Estimated population losses

are very high because the Mongols destroyed much of the economic infrastructure and also implemented genocidal policies (Deng, 2003). Recovery began with the Ming dynasty after 1400.

For this analysis, we assume a net mortality of 40 million people in China between 1200 and 1400 (Table 1). A simple but crude calculation can be made of the maximum possible impact of this level of mortality in reducing carbon emissions from coal burning. We assume that each 5-person family that used coal burned a total of 10 kg per day, of which 50% by weight was carbon. Each family would have emitted ~5 kg of carbon/day, or 1.8 tons/year. If the 40 million deaths occurred in ~8 million families that burned coal, the total reduction in carbon emissions would have been ~14.4 million tons/year. Over the course of the 150 years between 1250 and 1400, the total reduction in emissions would have been ~2.2 GtC. We allocate this 2 GtC as reductions of 0.73 GtC for each 50-year interval between 1250 and 1400.

We use the filter in Figure 5 to estimate the effect of this reduction in carbon emissions on the amount of carbon in the atmosphere. The maximum simulated lowering from reduced burning of coal is 0.83 GtC from 1350-1400 (Fig. 6c).

4. The Effects of Depopulation on Atmospheric CO_2

The simulated reductions in the carbon content of the atmosphere (Fig. 6) can be directly converted into estimates of the reductions in atmospheric CO_2 (Fig. 7, see page 411). As noted in section 2, this conversion follows the pre-industrial atmospheric ratio of 2.13 (the change in GtC per 1 ppm change in CO_2). In effect, we simply scale the GtC changes in Figure 6 to this ratio in order to plot Figure 7. In this section, we evaluate the separate (and then combined) effects of reforestation, reduced deforestation, and reduced coal burning on atmospheric CO_2. Estimated changes in atmospheric CO_2 are plotted at the same scale as the CO_2 'target signal' (Fig. 7a).

Reforestation. The CO_2 anomaly that developed during the depopulation interval spanning the late Roman era reached a maximum simulated value of ~1 ppm by 600-650 (Fig. 7b). The gradual development of this anomaly resembles the (poorly-dated) CO_2 anomaly measured at Taylor Dome (Figs. 2b, 7a), and the simulated anomaly accounts for about half of the amplitude of the measured CO_2 anomaly. For the Black Death, the model simulates a short-lived CO_2 drop of ~1.5 ppm between 1400 and 1450. This value represents more than half of the (noisy) CO_2 decrease of ~2 ppm recorded in the

well-dated Law Dome record. For the American pandemic era, the model simulates a CO_2 decrease of just over 2 ppm between 1650 and 1700 (Fig. 7b). This CO_2 drop represents about 40% of the ~5 ppm CO_2 decrease observed between 1500 and 1750 at Law Dome. In all three cases, carbon sequestration caused by reforestation provides a plausible explanation for roughly half (or more) of the CO_2 decreases measured in the ice cores.

Deforestation. Deforestation has only a second-order impact on atmospheric CO_2 values during pandemics (Fig. 7c), although the relative impact of the long-lasting depopulation episodes that followed the late Roman era and American pandemics are not negligible. Together, reforestation and reduced deforestation appear to account for half of the observed CO_2 drops during the three pandemic episodes.

Coal Burning. The reduction in burning of coal from shallow pit mines caused by ~40 million deaths in China between 1250 and 1400 causes an estimated CO_2 drop of ~0.4 ppm between 1350-1400 (Fig. 7d). Like reduced deforestation, reduced coal burning is a second-order factor in past changes in atmospheric CO_2.

4.1. *Combined Effects of Depopulation Episodes on Atmospheric CO_2*

Atmospheric CO_2 concentrations rose at a relatively steady rate from 8000 to 2000 years ago, but then leveled off and began to oscillate erratically (Fig. 3). Ruddiman (2003) proposed that the reason for both the change in trend near 2000 years ago and the subsequent oscillations might be linked to the effects of major pandemics in sequestering carbon and reducing the amount of carbon in the atmosphere.

The model-simulated effects of reforestation, reduced deforestation and reduced coal burning are combined in Figure 8b (see page 412). The short-term effects of the major depopulation episodes – the CO_2 decreases at and after 200, 1350, and 1500 – are the most obvious feature of this trend. But the atmosphere has a long 'memory' for carbon emissions in prior centuries and even prior millennia, and substantial effects of these episodes linger for centuries and millennia afterward.

This trend can be compared against the CO_2 'target signal' shown in Fig. 8a. The plot in Figure 8b does not attempt to simulate the (unknown) changes in agricultural deforestation that were underway in many regions of the world through the last two millennia. Increases in deforestation would have driven atmospheric CO_2 concentrations to higher values. Instead, this plot summarizes only the extent to which depopulation

episodes could have slowed the long-term increase in atmospheric CO_2 that had been underway for millennia (Fig. 3)

The ~1 ppm CO_2 reduction that occurred during the late Roman era had decreased to about 0.7 ppm just before the effects of reduced coal burning in China and the Black Death pandemic began to be felt. An abrupt increase in Eurasian populations between 1000 and 1300 (McEvedy and Jones, 1978) may have caused increased deforestation and driven CO_2 values sharply higher at that time (Fig. 8b).

Subsequently, three depopulation episodes occurred within a relatively short time. The combined effects of reduced coal burning in China, the Black Death pandemic in Eurasia, and the American pandemic drove CO_2 concentrations ~4 ppm lower than the values reached in 1100-1200. This drop represents just over half of the 7-8 ppm decrease in ice-core CO_2 measured between 1100 and 1700 (Fig. 8a). It appears that depopulation (mainly through pandemics) imposed a net CO_2 decrease of ~4 ppm on the rising trend that had been underway for many millennia (Fig. 3). This analysis does not take into account the possible role of climate-system feedbacks in amplifying CO_2 changes imposed by depopulation episodes.

The net simulated decrease of ~4 ppm that might be explained by pandemics and other causes of depopulation between 1000-1200 and 1500-1750 would result in a net global cooling of about 0.05°C. This estimate represents about one third of the century-scale cooling between 1000-1200 and 1500-1750 estimated by Mann *et al.*, (1998) and Crowley and Lowery (2000). A vigorous debate is underway as to whether these temperature estimates for the last millennium are valid or whether the actual cooling was larger. Within this ongoing uncertainty, pandemic-driven effects on reforestation and carbon sequestration could range from a significant factor in a small cooling to a secondary factor in a larger cooling.

An interesting sideline of this analysis is the fact that the simulated (and lingering) effects of these historical pandemic intervals make the CO_2 concentration in the modern atmosphere ~3 ppm lower than it would be in their absence (Fig. 8b).

5. Discussion and Conclusions

Variations in population (Fig. 1) and climate (Fig. 2) appear to have been coupled during the historical era. Invoking a form of environmental determinism, Lamb (1977) proposed a causal chain in which cooler cli-

mates reduced food production, causing famine and death as direct results, and also produced malnutrition and disease that led indirectly to additional mortality. Historical demographers have criticized this hypothesis on the basis that mortality from even the largest pre-industrial famines was neither large enough nor persistent enough to have a first-order effect on populations (Watkins and Menken, 1985, 1988). Climate scientists have voiced doubts that the changes in temperature proposed by Lamb would have had significant effects on food production, as they would have had impacts only in colder regions at high latitudes and altitudes that are at best marginally suitable for agriculture, and thus not heavily occupied by humans (Le Roy Ladurie, 1971; Pfister *et al.*, 1999). The small size of the temperature changes reconstructed for the northern hemisphere (Figure 2a) adds weight to this criticism.

Nevertheless, large-scale depopulation episodes during the historical era do appear to correlate with drops in atmospheric CO_2 levels and decreases in northern hemisphere temperature (Figs. 1, 2). Given the inability of volcanic eruptions and changes in solar irradiance to account for CO_2 drops as large as 8 ppm without violating hemispheric and global temperature constraints, pandemics and depopulation episodes are a potential casual link.

Three plausible mechanisms could link major historical depopulation episodes with CO_2 decreases and climate: reforestation of abandoned lands, reduced rates of deforestation, and reduced burning of coal. Our analysis suggests that reforestation was the primary mechanism that lowered CO_2 concentrations, that reduced rates of deforestation played a lesser role, and that reduced coal burning was also relatively unimportant. Given the large number of assumptions required for these analyses, the results summarized here should be viewed as a demonstration of the feasibility (plausibility) of these links, but not as a demonstration of proof.

If our basic conclusions are correct, disease and climate are linked historically, but in an indirect way. Rather than changes in climate causing changes in population (as in Lamb, 1977), both climate and population respond to an independent driver – the massive level of mortality during pandemics and other major depopulation episodes.

Appendix

Table 1. Population estimates for China (in millions).

Age	McEvedy and Jones (1978)	Biraben (1979) and Livi-Bacci (2001)
200 BCE	40	40
1 CE	50	70
200	60	60
400	50	25
600	45	49
800	50	56
1000	60	56
1100	100	83
1200	115	124
1300	85	83
1400	75	70
1500	100	84
1600	150	110
1650	130	100-150
1700	150	190-225
1750	215	220
1800	320	320-330

TABLE 2. Estimated dates when populations for various regions* would have reached the Domesday density threshold of 11 persons/km^2.

Continent/ Country/ Region	Domesday Population Threshold (10^6 People)	Date Threshold Reached
Asia		
N. China, Yellow River	9	before 0
C. China, Yangtze River	9	before 0
NE India, Ganges River	12	before 0
NW India, Indus River basin	8	before 0
Japan	0.7	300
South-Central India	1.7	700
Indonesia	3.3	800
Southern China	4	1100
Korea	2.2	1100
Europe		
Greece	0.6	before 0
Italy	1.7	before 0
Czechoslovakia	1.1	800
France	6.1	1000
Belgium/Netherlands	0.7	1000
Austria	0.6	1000
British Isles	2.5	1100
Spain/Portugal	5.3	1100 (200)
Germany	5.2	1200
Hungary	1.0	1300
Poland	3.4	1300

* McEvedy and Jones (1978), Denevan (1992).

TABLE 3. Estimated sequestration of carbon (in Gt = 10^9 tons) by reforestation of farmland (crops and pastures) abandoned because of high pandemic mortality.

Pandemic (region)	Forest type	Mortality** (millions)	Per-capita Footprint (km^2)	C Sequestered (Tons/km^2)	GtC Sequestered
Late Roman Era (200-800)					
Scandinavia	Conifer	-0.1	0.09	1350	-0.1
N-C Europe	Deciduous	3.05	0.09	1630	4.5
S Europe	Mediterranean	2.05	0.09	1820	3.3
				Total	~7.7
Black Death (1350-1450)					
Scandinavia	Conifer	0.5	0.09	1350	0.6
N-C Europe	Deciduous	2.6	0.09	1630	7.6
				Total	~8.2
American (1500-1750)					
N. America	Deciduous	3.5	0.02	1630	1.1
Mexico	Seasonal/dry	15	0.02	800	2.4
Amazon	Tropical Wet	15	0.02	2190	6.6
Andes	Montaine	14	0.02	1300	3.7
				Total	~13.8

** Mortality shown only for regions still in process of being deforested (Table 1).

TABLE 4. Rates of deforestation in pandemic-stricken areas as a fraction of all regions still undergoing deforestation on a global basis.

200-800 Epidemics and Pandemic			
Region	Population Deforesting (m)	Correction for stone tools	Adjusted Rate of Deforestation
Pandemic: Europe	*5.2*	-	*5*
Other:			
Asia	25	-	25
Africa (Sub-Sahara)	5	x 0.22	1
Americas	30	x 0.22	7
Global Total			38
Pandemic deforestation as a fraction of the global total: 5/38 = ~13%			

1350-1450 Pandemic			
Region	Population Deforesting (m)	Correction for stone tools	Adjusted Rate of Deforestation
Pandemic: Europe	*3.1*	-	*3*
Other:			
Asia	36	-	36
Africa (Sub-Sahara)	22	x 0.22	5
Americas	49	x 0.22	11
Global Total			55
Pandemic deforestation as a fraction of the global total: 3/55 = ~5%			

1500-1750 Pandemic			
Region	Population Deforesting (m)	Correction for stone tools	Adjusted Rate of Deforestation
Pandemic: Americas	*48*	*x 0.22*	*11*
Other:			
N. Europe	15	-	15
Asia	46	-	46
Afica (Sub-Sahara)	31	x 0.22	7
Total			79
Pandemic deforestation as a fraction of the global total: 11/79 = ~14%			

REFERENCES

Barnola, J.M., Anklin, M., Procheron, J., Raynaud, D., Schwander, J., and Stauffer, B. (1995), CO_2 evolution during the last millennium as recorded by Antarctic and Greenland ice, *Tellus B* 47, 264-272.

Bechmann, R. (1990), *Trees and Man: The Forest in the Middle Ages* [1984], translated from the French by Katharyn Dunham, New York: Paragon House.

Biraben, J.N. (1979), Essai sur l'évolution du nombre des hommes, *Population* 34: 13-24.

Bourdelais, P. (2003), *Les Épidémies terrassées: une histoire de pays riches*, Dijon Quetigny: Éditions de La Martinière.

Carcallet, C., *et al*. (2002), Holocene biomass burning and global dynamics of the carbon cycle, *Chemosphere* 49, 845-863.

Chao, K. (1986), *Man and Land in Chinese History*, Stanford: Stanford University Press.

Chang (2002), Disease and its impact on politics, diplomacy, and the military: the case of smallpox and the Manchus (1613-1795), *Journal of the History of Medicine and Allied Sciences* 57, 177-197.

Crowley, T.J. (2000), Causes of climate change over the past 1000 years, *Science* 289, 270-277.

Crowley, T.J. and Lowery, T.S. (2000), How warm was the Medieval Warm Period? A comment on man-made versus natural climate change, *Ambio* 39, 51-54.

Denevan, W.M. (ed.) (1992), *The Native Population of the Americas in 1492*, 2nd ed., Madison: University of Wisconsin Press.

Deng, K.G. (2003), China: Tang, Song and Yuan Dynasties, *The Oxford Encyclopedia of Economic History*, J. Mokyr (ed.), 5 volumes (New York and Oxford: Oxford University Press) 2, 423-28.

Duncan-Jones, R. (1996), The Impact of the Antonine Plague, *Journal of Roman Archaeology* 9, 108-136.

Dunstan, H. (1975), The late Ming epidemics: a preliminary survey, *Ch'ing-Shih wen-t'i* 3, 1-59.

Eckert, E. (1996), *The Structure of Plagues and Pestilences in Early Modern Europe: Central Europe, 1560-1640*, New York and Basel: Karger.

Elvin, M. (1993), Three thousand years of unsustainable growth: China's environment from archaic times to the present, *East Asian History* 6, 7-46.

Elvin, M. (2004), *The Retreat of the Elephants: an Environmental History of China*, New Haven: Yale University Press.

Esper, J., Cook, E.R., and Schweingruber, F.H. (2002), Low-frequency signals in long tree-line chronologies for reconstructing past temperature variability, *Science* 295, 2250-2253.

Etheridge, D.M., Steele, L.P., Langenfields, R.L., Francey, R.J., Barnola, J.-M., and Morgan, V.I. (1996), Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn, *Journal of Geophysical Research* 101, 4115-4128.

Fairservis, W.A., Jr. (1971), *The Roots of Ancient India*, New York: Macmillan.

Gerber, S., Joos, F., Brugger, P., Stocker, T.F., Mann, M.E., Sitch, S., and Scholze, M. (2003), Constraining temperature variations over the last millennium by comparing simulated and observed atmospheric CO_2, *Clim. Dyn.* 20, 281.

Gregg, S.A. (1988), *Foragers and Farmers: Population Interaction and Agricultural Expansion in Prehistoric Europe*, Chicago: University of Chicago Press.

Grigg, D.B. (1994), *The Agricultural Systems of the World*, Cambridge: Cambridge University Press.

Gutmann, M. (1980), *War and Rural Life in the Early Modern Low Countries*, Princeton: Princeton University Press.

Hartwell, R. (1962), A revolution in the Chinese Iron and Coal Industries during the Northern Sung, 960-1126 AD, *Journal of Asian Studies* 21, 153-162.

Hartwell, R. (1967), A cycle of economic change in imperial China: coal and iron in northeast China, 750-1350, *Journal of the Economic and Social History of the Orient*, 10, 102-159.

Hatcher, J. (1977), *Plague, Population and the English Economy, 1348-1530*, London: Macmillan.

Hatcher, J. (1994), England in the aftermath of the Black Death, *Past and Present* 144, 3-35.

Heidenrich, C. (1971), *Huronia: A history and geography of the Huron, Indians, 1600-1650*, Ontario: McClelland and Stewart.

Houghton, R.A. (1999), The annual net flux of carbon to the atmosphere from changes in land use 1850-1990, *Tellus B* 51, 298-313.

Hughes, J.D. (1975), *Ecology in Ancient Civilizations*, Albuquerque: Univ. New Mexico Press.

Hurt, R.D. (1987), *Indian Agriculture in America*, Kansas: Lawrence University Press.

Indermuhle, A., Stocker, T.F., Joos, F., Fischer, H., Smith, H. J., Wahlen, M., Deck. B., Mastroianni, D., Blunier, T., Meyer, R., Stauffer, B. (1999), Holocene carbon-cycle dynamics based on CO_2 trapped in ice at Taylor Dome, Antarctica, *Nature* 398, 121-126.

IPCC (2001), *Third Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge: Cambridge Univ. Press.

Jones, P.D., and Mann, M.E. (2004), Climate over Past Millennia, *Reviews of Geophysics* 42, 2003RG000143.

Jones, P.D., Osbourn, T.J., and Briffa, K.R. (2001), The evolution of climate over the last millennium, *Science* 292, 662-667.

Joos, F., Meyer, R., Bruno, M., Leuenberger, M. (1999), The variability in the carbon sinks as reconstructed for the last 1000 years, *Geophysical Research Letters* 26, 1437-1440.

Joos, F., Gerber, S., Prentice, I.C., Otto-Bleisner, B.L., Valdes, P. (2004), Transient simulations of Holocene atmospheric carbon dioxide and terrestrial carbon since the last glacial maximum, *Global Biogeochemical Cycles*, 18, GB2002 10.1029/2003GB002156.

Jordan, W.C. (1996), *The Great Famine: Northern Europe in the Early Fourteenth Century*, Princeton: Princeton University Press.

Lamb, H.H. (1977), *Climate: Past Present and Future*, vol. 2, London: Methuen.

Landers, J. (2004), *The Field and the Forge: population, production and power in the pre-industrial West*, New York and Oxford: Oxford University Press.

Lavely, W. and Wong, R.B. (1998), Revising the Malthusian narrative: the comparative study of population dynamics in late imperial China, *Journal of Asian Studies* 57, 714-748.

Lawrence, D., Peart, D.R., and Leighton, M. (1998), The impact of shifting cultivation on a rainforest landscape in West Kalimantan: spatial and temporal dynamics, *Landscape Ecology* 13, 135-148.

Lee, J.Z, and Feng, W. (1999), *One Quarter of Humanity: Malthusian Mythology and Chinese Realities, 1700-2000*, Cambridge, MA: Harvard University Press.

Le Roy Ladurie, E. (1971), *Times of Feast, Times of Famine: A History of Climate Since the Year 1000*, translated by Barbara Bray, Garden City, N.Y.: Doubleday.

Livi-Bacci, M. (2001), *A Concise History of World Population*, 3rd edition, Oxford: Blackwell.

Mann, M.E. and Jones, P.D. (2003), Global surface temperatures over the past two millennia, *Geophysical Research Letters* 30, doi: 10.10129/2003GL017814.

Mann, M.E., Bradley, R.S., and Hughes, M.K. (1999), Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophysical Research Letters* 26, 759-762.

Marks, R.B. (1998), It never used to snow: climatic variability and harvest yields in late-imperial south China, 1650-1850, pp. 411-446. In: *Sediments of Time: Environment and Society in Chinese History*, Mark Elvin and Liu Ts'ui-jung (eds.), Cambridge and New York: Cambridge University Press.

Maxwell, H. (1910), The use and abuse of the forests by the Virginia Indians, *William and Mary College Quarterly* 19, 73-104.

McEvedy, C., Jones, R. (1978), *Atlas of World Population History*, New York: Penguin.

McNeill, J.R. (1997), China's Environmental History in World Perspective, pp. 31-49. In: Mark Elvin and Liu Ts'ui-Jung (eds.), *Sediments of Time: Environment and Society in Chinese History*, New York: Cambridge University Press.

McNeill, W.H. (1976), *Plagues and Peoples*, New York: Doubleday.

Moberg, A., Sonechkin, D.M., Holmgren, K., Datsenke, N.M., Karlen, W. (2005), Highly variable Northern hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature* 455, 613-627.

Newson, L. (2001), Pathogens, places and peoples: geographical variations in the impact of disease in early Spanish America and the Philippines, pp. 167-210. In: Raudzens, G. (ed.), *Technology, Disease and Colonial Conquests, Sixteenth to Eighteenth Centuries*, v. 2 of the Loyola College History of Warfare, Leiden: Brill.

Ó Gráda, C. (2005), Markets and famines in pre-industrial Europe, *Journal of Interdisciplinary History* 36, 143-166.

Overpeck, J.S., *et al*. (1997), Arctic environmental change of the last four centuries, *Science* 278, 1251-1256.

Paine, R.R. (2000), If a population crashes in prehistory and there is no paleodemographer there to hear it, does it make a sound?, *American Journal of Physical Anthropology* 112, 181-190.

Palmer, A.S., van Ommen, T.D., Curran, M.A.J., Morgan, V., Souney, J.M., and Mayewski, P.A. (2001), High-precision dating of volcanic events (A.D. 1301-1995) using ice cores from Antarctica, *Jour. Geophys. Res.* 106, 28, 089-28, 095.

Pfister, C., Brázdil, R., and Glaser, R. (eds.) (1999), Climatic Variability in Sixteenth-Century Europe and its Social Dimension, *Special issue of Climatic Change* 43, (1).

Rackam, O. (1980), *Ancient Woodland: Its History, Vegetation, and Uses in England*, London: Edward Arnold.

Richards, J.F. (1997), Early Modern India and World History, *Journal of World History* 8, 197-209.

Richards, J.F. (2003), *The Unending Frontier: An Environmental History of the Early Modern World*, Berkeley and Los Angeles: University of California Press.

Roberts, N. (1998), *The Holocene*, Oxford: Blackwell.

Ruddiman, W.F. (2003), The anthropogenic greenhouse era began thousands of years ago, *Climatic Change* 61, 261-293.

Ruddiman, W.F. (2005), Cold climate during the closest analog to modern interglaciation, *Quaternary Science Reviews* 24, 1121-1132.

Sallares, R. (2002), *Malaria and Rome: A history of malaria in Ancient Italy*, Oxford: Oxford University Press.

Sarris, P. (2002), The Justinianic plague: origins and effects, *Continuity and Change 17*, 169-182.

Scheidel, W. (2001), *Death on the Nile: Disease and demography of Ancient Egypt*, Leiden: Brill.

Shuie, C.H. (2005), The political economy of famine relief in China, 1740-1820, *Journal of Interdisciplinary History* 36, 33-56.

Siegenthaler, U., Monnin, E., Kawamura, K., Spahni, R., Schwander, J., Stauffer, B., Stocker, T.F., Barnola, J.-M., and Fischer, H. (2005), Supporting evidence from the EPICA Dronning Maud Land ice core for atmospheric CO_2 changes during the past millennium, *Tellus 57B*, 51-57

Simmons, I.G. (1996), *Changing the Face of the Earth*, Oxford: Blackwell.

Stathakopoulos, D.C. (2004), *Famine and Pestilence in the late Roman and Early Byzantine Empire: a systematic survey of subsistence crises and epidemics*, London: Ashgate.

Taylor, C. (1983), *Village and Farmstead*, London: George Phillip.

Thirgood, J.V. (1981), *Man and the Mediterranean Forest*, London: Academic Press.

Twitchett, D. (1979), Population and pestilence in T'ang China, pp. 35-68. In: Wolfgang Baue (ed.), *Studia Sino-Mongolica: Festschrift für Herbert Franke*, Wiesbaden: Franz Steiner Verlag.

Vale, T.R. (ed.) (2002), *Fire, Native Peoples, and the Natural Landscape*, Island Press: Washington, D.C.

Watkins, S.C., and Menken, J. (1985), Famines in historical perspective, *Population and Development Review* 11, 647-675.

Watkins, S.C., and Menken, J. (1988), On the role of crises in historical perspective, *Population and Development Review* 14, 165-170.

Williams, M. (2003), *Deforesting the Earth*, Chicago: Univ. Chicago Press.

SOCIAL AND ENVIRONMENTAL VULNERABILITY TO EMERGING INFECTIOUS DISEASES

U.E.C. CONFALONIERI[1], M.E. WILSON[2] & A.L. NAJAR[1]

1. Introduction

Emerging Infectious Diseases (EIDs) have been broadly defined as 'diseases that have recently increased in incidence or geographic range, recently moved into new host populations, recently been discovered or are caused by newly-evolved pathogens' (Daszak *et al.*, 2001). In the past three decades, at least 40 EIDs have been identified, not including the regional resurgence of the widespread endemic diseases of the tropics (malaria, dengue fever, leishmaniasis, etc.) and the new antibiotic-resistant strains of bacterial infections (tuberculosis, staphylococcus, etc.) and protozoans (malaria).

There have been several papers and reports dealing with factors involved in the emergence (IOM, 1992; 2003; Molyneux, 2003; Mayer, 2000; Patz *et al.*, 2000; Daszak *et al.*, 2000; Wilson *et al.*,1994; Morse, 1993; 1995; Taylor *et al.*, 2001) as well as discussions on the technical, political and institutional responses to the EIDs (Broome, 1998; Binder *et al.*, 1999; Plotkin & Kimball,1997; Winch, 1998; LeDuc & Tikhomirov, 1994; Heymann & Rodier, 2001).

In this paper we use updated information on diseases that have emerged in the past few decades and discuss aspects of the contemporary social and environmental vulnerability to the emergence of infections. We present a conceptual framework to the study of the EIDs and we apply the concept of vulnerability to the study of the emergence and global dissem-

[1] The Oswaldo Cruz Foundation, Brazil.
[2] Harvard University, USA.

ination of infections. Also, via a case study we discuss some aspects of the global interconnection of social and environmental drivers for the emergence/dispersal of infectious and parasitic diseases. Special emphasis is placed on how mechanisms of emergence are determined by vulnerability factors; on what makes an emerging infection capable of global spread, and the current characteristics of urban structures that make large settlements places for reception and global dispersal of EIDs.

2. Disease Emergence in the Past Three Decades

In regard to the EIDs identified in the last thirty years we can make these general statements:

1. For many of these EIDs the basic drivers and mechanisms of emergence are not known. For several others, solid hypotheses are accepted. Also, the exact place of first emergence and the year that the phenomenon has occurred can rarely be tracked (years in the table refer to the first isolation of the agent).

2. Disease processes have emerged due to several drivers and mechanisms, both in developed as well as in developing countries. In developed countries, among the six major categories of drivers identified by the IOM (1992, 2003), those associated with changes in technology/industry, microbial adaptation and change, and international travel and commerce tend to predominate. On the other hand, those EIDs clearly identified as originating in low income countries have usually been associated with the encroachment of agriculture on natural systems (land use changes); with direct contact with infected animals; and with the breakdown of public health measures and changes in demographics and human behavior.

3. The more immediate drivers and mechanisms can be discerned with some reliability or at least hypothesized. However, the more distant drivers are rarely discussed or even sought. Emergence is associated with disturbances of previously stable microbiological equilibria caused by various social and environmental activities by humans. The primary drivers of change (social, economic, cultural, political and behavioral) have so far received little research attention.

4. There is a variable epidemiological pattern in the EIDs. Some have become widespread pandemics (AIDS, cholera); others are globally distributed but occurring at low levels or as eventual outbreaks (Campylobacter; Rotavirus; E. coli, Legionella, etc.); while a few are restricted to groups of

countries (SARS, West Nile Virus fever, HTLV-1, Monkeypox). A few others have been reported occasionally from restricted localities (Nipah, Hendra, Pfiesteria, Avian Influenza, Guanarito, etc.), though some, such as avian influenza, seem to be developing the capacity to spread widely.

3. Vulnerability to Emerging Infectious Diseases

Vulnerability is a concept derived from geography and social sciences and, as far as human well-being is concerned, it has been developed initially to study the impacts of natural disasters upon human communities. It has been defined as 'the characteristics of a person or group related to the capacity to resist, cope with, and recover from the impacts caused by hazards' (Blaikie *et al.*, 1994). Another definition is 'the differential capacity of groups and individuals to deal with hazards based on their positions within the physical and social worlds' (Dow, 1992).

Vulnerability is broadly defined in terms of exposure, capacity and potentiality (Watts & Bohle, 1993), and is also distinguished as both a biophysical condition (geographic space) and defined by political, social and economic conditions of society (Liverman, 1990). In general it is recognized that the most vulnerable social groups are those that experience the most exposure to a hazard; that are the most sensitive to it (e.g., the most likely to suffer from it), and who have the weakest capacity to respond and ability to recover.

In relation to EIDs the concept of vulnerability can be applied to three different levels: drivers and mechanisms of emergence; dissemination of infections; and the social responses to control them. This expanded framework of vulnerability to biological hazards (EIDs) should include the characteristics of societies that produce the drivers and mechanisms of emergence and also render the receptive populations exposed to and sensitive to the infections. It would also include the conditions that facilitate the amplification of infections as well as those that determine the capacity of the social groups to respond to these hazards in a timely and effective way. By drivers we mean the social, economic and cultural dynamics that direct the human interventions in the physical world. By 'mechanisms' of emergence we consider the ensuing reactions in the environment and in the microbial world.

At the ecological level, mechanisms of emergence always involve the transfer of a pathogen from an 'extra-human' environment (biophysical

environment; domestic or wild animal) to the human hosts ('host transfer' in Figure 1). The early process of emergence of novel human infections can be approached from different perspectives: biological, ecological, geographical, social and behavioral. Vulnerability factors may operate in different moments of emergence: microbial changes, human exposure, human infection and local dissemination of the infection (Figure 2).

At the level of microbial changes human interference can facilitate genetic shifts through the use of drugs for therapy (microbial resistance) or by facilitating the interchange of agents between different hosts, such as the case of avian influenza, or by changing the natural ecosystems, such as the case of deforestation and the outbreaks of Venezuelan equine encephalitis (Brault *et al.*, 2004).

Human contact with potentially novel pathogens (= exposure) may be determined by a host of factors, which range from the place of settlement, feeding habits (e.g., bushmeat), behavior (e.g., HIV) and medical and other technologies (e.g., blood transfusion, etc.).

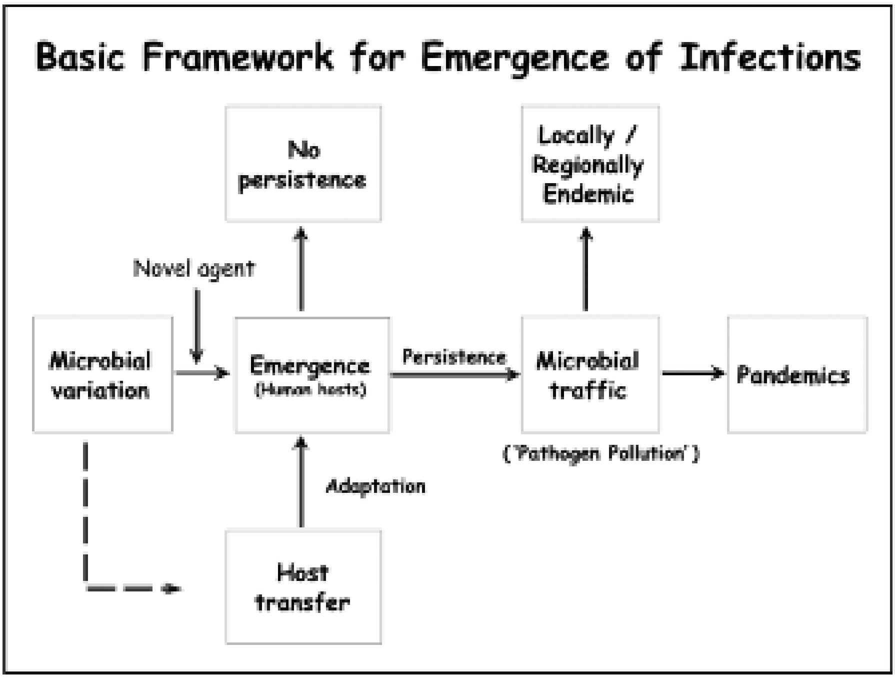


Figure 1. Basic framework for the emergence of infections.

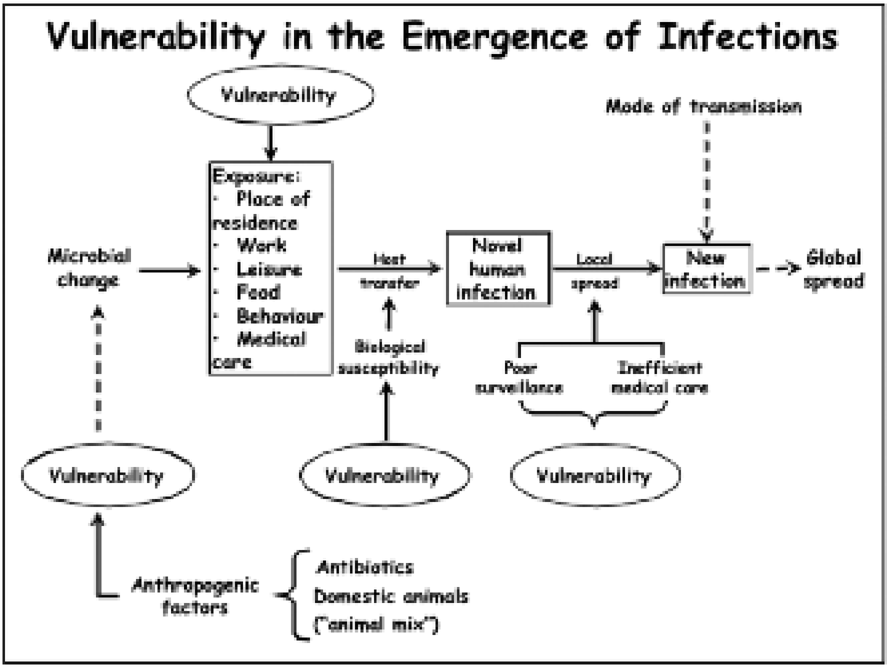


Figure 2. The role of vulnerability in the emergence of infections.

In the following step – host transfer (i.e., the invasion of human hosts) – vulnerability is determined basically by biological susceptibility (immunity). Local amplification of infection is primarily determined by the capacity of the organism to pass from one human host to another (infectivity) – but it is also a result of the capacity of the epidemiological surveillance systems to detect and control the spread of infection.

As for the phase of global (or regional) spread of EIDs, vulnerability is linked to the factors that cause people and goods to move from one geographical location to the other, either for economic reasons (commerce, work) or because of political factors (conflicts) and even tourism, an aspect which will be discussed later. In this regard, large metropolitan areas are especially vulnerable places as they are 'receptive' to the introduction and redistribution of pathogens undergoing global spread, also to be discussed in another section.

The capacity to respond to a disease emergence involves both the early detection of the disease and the implementation of control meas-

ures. This means a combination of clinical attentiveness, good laboratory facilities and comprehensive epidemiological surveillance systems for infections in both humans and animals (Desselberguer, 2000). Factors that may cause greater vulnerability in this area (= poor capacity to react) are political instability, lack of technical infrastructure and lack of professional expertise.

4. The Vulnerability of Developing Countries

Developing countries are especially vulnerable to the emergence and spread of novel human infections for several reasons, which can be summarized in Figure 3.

With respect to the biological processes involved, the probability that microbial change and adaptation will occur is higher in developing countries since they contain most of the existing natural systems in the world

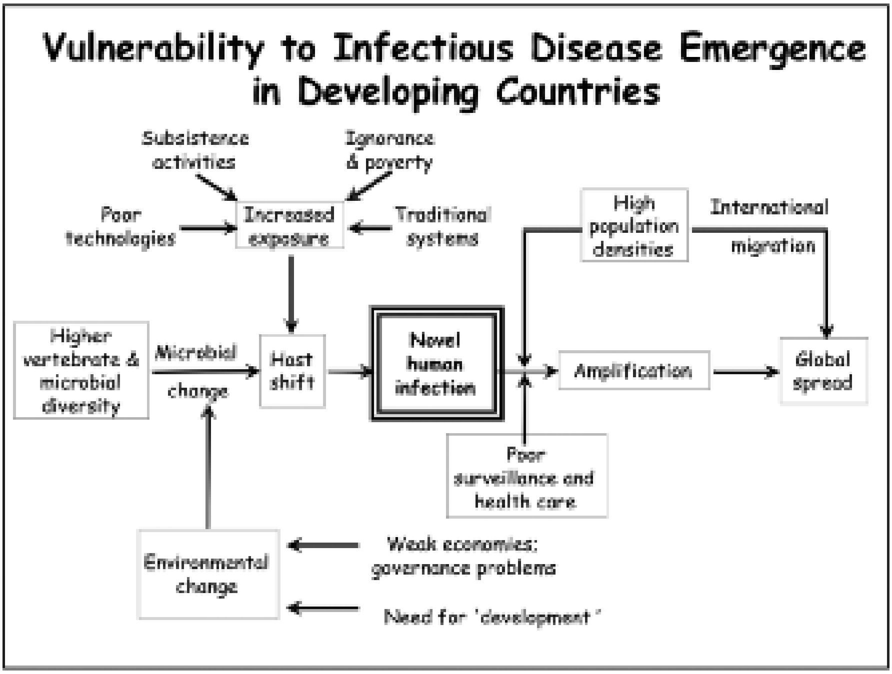


Figure 3. Vulnerability to emergence of infectious disease in developing countries.

that provide a rich environment for emergence, especially in the tropics, due to their high vertebrate and microbial biodiversity (Wolfe *et al.*, 2000). In this context the increased probability of a successful emergence event is determined by the intimacy and frequency of contact of people with the microbe-rich environment due to immediate subsistence needs, through direct contact with infected vertebrate animals, such as the case of bushmeat hunting (De Merode *et al.*, 2004). Human contact with potential new infections in these ecosystems can also be associated to the exposure to vectors during the human invasion of their established ecological niches, as part of their work. It has been demonstrated that environmental changes such as deforestation can result in microbial genetic and population shifts that may favor disease emergence (Borneman & Tripplet, 1997; Brault *et al.*, 2004). This is due to major shifts in the ecological niches of microbial populations and to changes in the selection pressures in the environment.

People are driven to enter these microbe-rich niches by their need for jobs (frequently 'unregulated'). The activity of exploitation of natural systems is often associated to international economic demands, especially for agricultural products and logging. However, traditional systems may also be associated with enhanced human-animal-environment contacts that may facilitate host transfer.

The major factors favoring the amplification of novel infections are demographic and those related to health systems. The new infection may escape early detection due to the chronic deficiencies in health surveillance in developing countries. On the other hand, high population densities (especially in urban areas) and migration (rural to urban; developing to developed countries) contribute to the spread of the pathogen on a wider scale.

The following case of the relationship between bovine spongiform encephalopathy (BSE) in Europe and the subsequent expansion of soybean cultivation in Brazil illustrates the vulnerability of developing countries to infectious disease emergence. It describes the possible linkages between disease-driven changes in livestock production systems in high-income countries and environmental changes favoring the emergence of infections in low-income countries.

Due to the emergence of the BSE in cattle in the UK (1986) and the subsequent discovery of the zoonotic aspect of the causative prion, producing the new variant Creutzfeldt Jakob Disease (Prusiner, 1998), new regulations for feeding cattle were enforced and rendered foods for animals were banned in Europe and elsewhere. This produced in the inter-

national market of agricultural commodities an increased demand for soybean grain, which is a 'clean' plant protein adequate to feed cattle and other domestic animals. Also, soy growing has been progressively transferred from temperate to tropical areas, where land is cheaper (Fearnside, 1999). In Brazil, one of the most important soybean producers in the world, this increased demand is causing the expansion of the soybean growing area in the central part of the country to the north, encroaching the Amazon forest. (Fearnside, *loc. cit.*). The country contains the largest block of original rainforest cover in the world and this poses a new risk of outbreaks of emergent arboviruses since this ecosystem is known to be rich in these microorganisms (Vasconcelos *et al.*, 2001), and environmental changes such as deforestation may place humans in contact with previously inaccessible pathogens (Figure 4).

This example illustrates the linkages between economic globalization and agricultural policies in developing countries and their adverse consequences for the natural systems and the resulting increase in the risk of disease emergence.

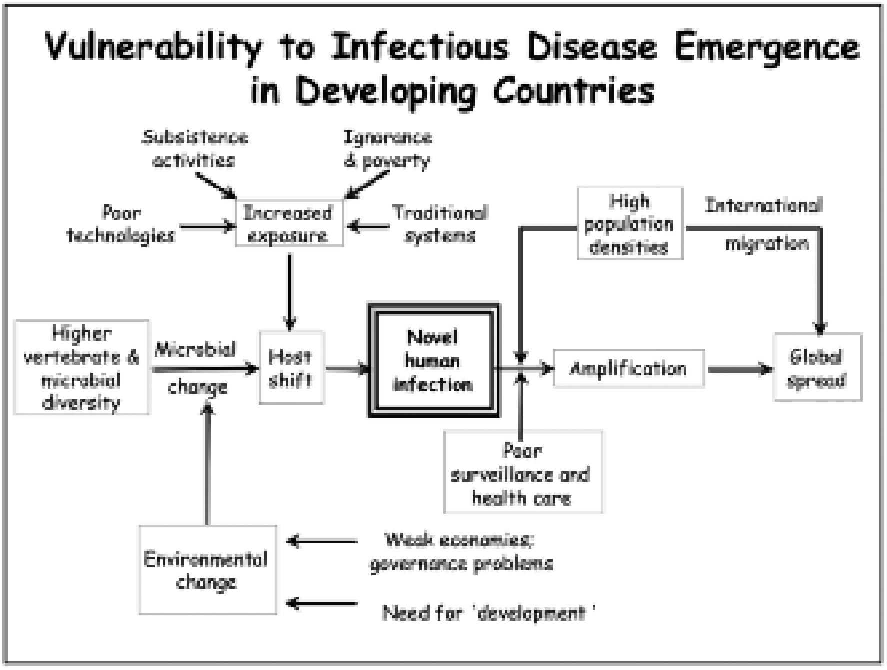


Figure 4. Soybean planting and disease risk in South America.

5. Globalization of Emerging Infections

Processes relevant to the globalization of infectious diseases, which is the last stage in the disease emergence process, will be discussed under these two general headings:

1) movement of humans and associated biological material;
2) trade.

5.1. *Movement of Humans and Associated Biological Material*

International travel and commerce facilitate rapid, massive, and global dispersal of biological material, including microbial genetic material. The scale, speed, and reach of movement of people and goods today is unprecedented and shapes the appearance, spread, and distribution of infectious diseases in humans, plants, and animals (Wilson, 1995a). Travel, trade, and the wide availability of mass transport have vastly expanded the movement of biological material that occurs naturally via migration of animals, flight of birds, movement of species such as insects, plant seeds, and marine life over land, in winds and air currents, and in streams and oceans.

International travel is growing in numbers and speed. Today more than 1.4 million people cross international borders on airplane flights each day. In 1950 the number of international tourist arrivals was 50 million; in 2002 more than 700 million international tourist arrivals were registered (WTO, 2003). Although most travel occurs along well-established corridors, more travelers are reaching remote areas where they may have contact with potentially pathogenic microbes that are not yet well characterized in animals or environmental reservoirs. In 1999, more than 75 million people traveled from industrialized to developing countries. By the 1990s more than 5000 airports had regularly scheduled international flights. This means that dense urban centers throughout the world are connected by a steady flow of humans.

Migrating humans (soldiers, merchants, explorers, refugees, missionaries, others) have long played a role in the introduction and spread of infectious diseases. (Berlinger, 1992; Bruce-Chwatt, 1968; Crosby, 1989). The difference today is the scale, speed, and reach of interconnections. A human can be viewed an interactive biological unit, carrying an assemblage of microbial genetic material and immune responses and immunological memory that reflect past exposures (Wilson, 2003). Travelers often spend time in a sequence of shared environments (e.g., bus, train, terminal, plane,

ship, etc.), which link the traveler to a wide network of contacts. Relevant exposures can occur in transit as well as at the final destination. Travelers can pick up, carry, and transmit microbial genetic material during and after travel – in some instances long after travel. Microbes transmitted from person to person (especially respiratory pathogens and sexually transmitted infections) can be carried to any part of the world by travelers. HIV is a prominent contemporary example of an infection that was carried by humans and transmitted widely, primarily via sexual contact.

Humans who are sick often attract attention (and medical interventions), but humans can carry pathogens and transmit them in the absence of symptoms. Examples include asymptomatic infection with HIV and hepatitis B virus. Pathogens that regularly colonize mucosal surfaces (e.g., *Neisseria meningitidis, Streptococcus pneumoniae, Staphylococcus aureus*) may be carried (often transiently) and transmitted by healthy people who have no awareness that they carry these potential pathogens (Wilder-Smith *et al.*, 2002). Humans can also carry resistance genes, sometimes found in commensal bacteria of the gut or skin, which may be transferred to other microbes (O'Brien, 2000; Okeke, 2001). A carrier who introduces a pathogen, a new clone, or a resistance gene into a community may never experience any illness. Some infections, such as HIV, can remain asymptomatic for prolonged periods, allowing transmission to occur repeatedly, over a long period of time, and in places remote from the site of acquisition.

5.2. *Global Trade and the Spread of Diseases*

Global trade is another mechanism for the worldwide dispersal of pathogens (that can infect humans, plants, and animals), insect vectors (such as mosquitoes), and intermediate hosts. Imported food now comprises a substantial part of the diet for many populations, especially in North America and parts of Europe. For example, the U.S. imports more than 30 billion tons of food per year, including fruit, vegetables, meat, and seafood (Murphy, 1999). The food chain has become very long; fresh produce may be grown and packed thousands of kilometers from where it is consumed. Mass production and mass processing of many foods and wide distribution networks mean that microbial contamination at one point in the production site or during processing may reach thousands or millions of people in multiple countries. Contamination can occur because of human ignorance or error, breakdown in equipment, or it may be introduced intentionally.

Trade in exotic animals is huge, much of it illegal; animals may be sought as pets or for food. In the U.S. the annual trade in illegal plants and animals is estimated at US$3 billion, a significant portion of it from trade in exotic reptiles (for pets). In one port (Miami) in one year (1996) more than 30 million animals were legally imported. Among the arrivals were 28.6 million fish, 1.1 million reptiles, 108,000 amphibians, 70,000 mammals, and 1,400 birds.[1] In 1999, the year that an outbreak of West Nile virus (WNV) was first identified in the U.S., with its epicenter in New York, 2,770 birds legally entered the country through JFK International Airport in New York. In addition, almost 13,000 birds passed through this airport in transit (Rappole, 2000).

Multiple wildlife species, many of them exotic and rare, are often collected from diverse, remote areas and kept in close proximity in zoos or wildlife conservation parks, adjacent to large human populations. These captive wildlife populations undergo inconsistently intense observation and monitoring. They are occasional sources of infection but can also serve usefully as sentinel populations for new events. In the WNV outbreak in New York, testing of stored and newly collected sera suggested that birds newly introduced to the Bronx Zoo were not the source of infection, that WNV infection had not been present before 1999, and that it spread widely in the zoo bird population and to a lesser extent in mammals (Ludwig *et al.*, 2002).

In an example of the triumph of trade pressures over good sense, distributors of exotic animals have exported prairie dogs from the U.S. to multiple other countries even though prairie dogs can be infected with the bacteria that cause tularemia, plague, and other potentially life-threatening infections in humans. An outbreak of tularemia in prairie dogs at a distribution center in Texas, which exported these animals to at least 10 states and seven countries, highlights this risk. Microbes in imported animals can be a source of human disease and may also threaten domestic and local wildlife populations (Daszak *et al.*, 2000). In the spring of 2003, prairie dogs housed with exotic African rodents became infected with monkeypox virus (from Africa), resulting in subsequent transmission to humans (Guarner *et al.*, 2004).

A UK government report estimated that in 2003 more than 11.5 tons of bushmeat (including monkey, rat, bat, gorilla, camel, and elephant) were smuggled into the UK (*The Telegraph*, 5 Sept 2004, reported on ProMED). In Africa, the establishment of networks, improved infrastruc-

[1] Stephanie Ostrowski, Division of quarantine, CDC; verbal communication.

ture, and especially the building of roads for logging and other development has facilitated the trade in bushmeat (Nisbett *et al.*, 2001).

Simian immunodeficiency viruses (SIVs) are widespread in primate species. In one study of blood and tissue samples in Cameroon, from 788 monkeys (bushmeat and pet animals), 16.6% of plasma samples tested strongly positive for HIV antigens (Peeters, 2002). There is good evidence that the AIDS pandemic in humans originated from multiple introductions of related SIV viruses from African apes and monkeys and the subsequent evolution of these viruses (Hahn *et al.*, 2000; Sharp *et al.*, 2001). Contact with primates as pets or with primate tissue through butchering animals could allow transfer of SIVs to humans. A recent survey of inhabitants in rural Cameroon found that more than 60% reported contact with fresh blood or body fluids from nonhuman primates, primarily through hunting and butchering chimpanzees, monkeys, or gorillas (Wolfe *et al.*, 2004). Further testing of humans with contact showed evidence of infection with simian foamy virus, a retrovirus endemic in African primates. Although this virus is not known to cause disease in humans, its presence in humans confirms ongoing transmission of primate viruses to humans.

Arthropods that can serve as vectors of human, animal, and plant diseases are regularly transported around the world by ships, planes or other vehicles. They may become established and spread in the new area (Bram *et al.*, 2002; Lounibos, 2002). Historically, water tanks in boats provided a suitable environment for the survival of mosquitoes traveling across oceans. *Aedes albopictus*, a mosquito competent to transmit dengue fever virus and other viral pathogens, was introduced into North America in the 1980s, probably via used tires imported from Asia. It was first identified in Texas and spread to 678 counties in 25 states within 12 years (Moore *et al.*, 1997). Of note, its dispersal followed interstate highways. It has also been introduced into several countries in Latin America. *Aedes albopictus* is the vector implicated in an outbreak of dengue serotype-1 that started in Hawaii in 2001.

Introduced exotic species threaten local populations and ecosystems. Ships used to transport cargo carry all types of biological life in ballast water and have been the source of introductions of invasive, exotic species into ports around the world (Carlton & Geller, 1993). Infectious diseases also threaten plants and animals and thereby have a major economic impact. Examples include bovine spongiform encephalopathy (BSE) in the UK and subsequently in many other countries, and foot and mouth disease, especially in the UK. Infectious diseases in plants and animals may also

have both direct and indirect consequences to human health. The transnational movement of people and goods facilitates the introduction of plant pathogens and their vectors into new communities. Infectious diseases of plants such as Karnal bunt of wheat, potato late blight, and citrus tristeza affect large, commercially important crops (Bandyopadhyay & Frederiksen, 1999). The destruction by disease of major food crops such as rice, wheat, and potatoes, could affect food security.

6. Global Cities and Disease Emergence

Attributes of the world's populations today that favor the emergence and spread of infectious diseases include size, density, location, mobility, vulnerability, inequities, and demographic shifts (Wilson ME, 1995b). The human population is larger than ever in history. Most of the population growth in the coming decades is projected to be in developing countries. Populations of domestic animals are also larger than ever in the past; densely concentrated populations of animals may grow under conditions that favor spread of infections.

More people live in urban areas than ever in history, and urbanization is expected to increase. The most rapidly growing segment in the world is the urban population in developing countries (UN, 2001). Megacities, defined as urban areas with a population greater than 10 million, have increased from 5 in 1975 to 19 in 2000. Most of the new megacities are located in low latitude areas and in developing countries.

Several research programs have compared megacities in the more developed countries, focusing on the paradigm of the success of the global city (Sassen, 1991). However, comparisons of megacities that are at different development levels may also be informative, either to understand better their specific characteristics as highlighted via the contrast, or to investigate common processes that seem to affect them.

The widespread segregation of rich and poor, within cities, is increasingly recognized as a public policy crisis that assails low-cost housing projects in outlying urban areas. This 'dualization' problem can be viewed as the spatial translation of social exclusion. This isolation of underprivileged segments has been criticized for two reasons: first, due to the negative effects of regrouping poverty-stricken communities – resulting in a cumulation of material difficulties and socialization problems, with the risk that the immigrants will return to communitarianism that then hampers their integration

as citizens; and second, the negative physical characteristics of these neighborhoods – decaying buildings, excessively dense and ‘inhuman’ town-planning, enclaves, and a lack of safety and security. There is increasing awareness that segregation is becoming more marked, with two-tier cities appearing, characterized by a gap between the excluded and the others.

In many cases, the spatial dualization hypothesis applied to megacities is based on specific areas that constitute particularly striking examples: fashionable boroughs and gentrified neighborhoods on the one hand, with decaying downtown areas or huge outlying townships assailed by unemployment and poverty on the other. Although striking, these examples of marked contrasts do not offer an overall view of the development of the urban system, and do not confirm that this situation of opposites is due to the dualization process of the urban system as of a whole.

The relationships between EIDs and these recent societal developments in cities deserve more research attention, especially since significant alterations in the differential use of space by different social groups in cities have been occurring at increasingly rapid rates from the 1970s onwards. Of practical epidemiological importance is the fact that many of the large urban areas in developing countries are surrounded by peri-urban slums characterized by poor infrastructure, absence of clean water, inadequate waste disposal and sanitation, poor housing, and close contact between animal and human populations. Many people dwelling in these peri-urban slums have relatives in rural communities. Frequent contact between residents of peri-urban and rural areas offers opportunities for passage of microbes into and out of the urban areas. Medical care may be limited, allowing infections and other medical problems to go undiagnosed, untreated, and unreported.

7. Conclusions

Infectious disease emergence is a complex phenomenon, involving different types of drivers at different stages in the process of emergence. These range from the biological susceptibility of hosts to the economically-driven environmental changes caused by humans and the efficiency of health-care systems. The combination of these factors that influence disease emergence differs in developing and developed countries. In the former we find more often the combination of conditions most conducive to disease emergence, making those countries more vulnerable.

Although not all emergent diseases are able to 'go global' in today's world, international movements of humans and associated biological material as well as the global trade have been important mechanisms for the dissemination of pathogens and vectors around the world. Urbanization is expected to increase globally, especially in the developing world, and segregation within cities (that is, the isolation of underprivileged social segments) is becoming more marked, both in developing as well as in developed countries. The recent examples of SARS and avian influenza point to the role played by large cities in the amplification of emergent infectious diseases. As central nodes in the international network of travelers and of the exchange of goods, as well as densely populated by humans living under conditions of inequality, the global cities have become a matter of special concern in the epidemiology of disease emergence.

REFERENCES

Bandyopadhyay R., Frederiksen R.A., Contemporary global movement of emerging plant diseases. In: Frazier T.W., Richardson D.C. (eds.), *Food and agricultural security*, New York Academy of Sciences, 1999; 892: 28-36.

Berlinguer G., The interchange of disease and health between the old and new worlds, *American Journal of Public Health*, 1992; 82(10): 1407-13.

Binder S., *et al.*, Emerging infectious diseases: public health issues for the 21st Century, *Science*, 1999; 284: 1311-3.

Blaikie P., *et al.*, *At Risk: natural hazards, People's vulnerability and disasters*, Routledge, London, 1994.

Borneman J., Tripplet E.W., Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation, *Appl. Environm. Microbiol.*, 1997; 63: 2547-53.

Bram R.A., George J.E., Reichard R.E., Tabachnick W.J., Threat of foreign arthropod-borne pathogens to livestock in the United States, *Journal of Medical Entomology*, 2002; 39: 405-16.

Brault A., *et al.*, Venezuelan equine encephalitis emergence: enhanced vector infection from a single amino acid substitution in the envelope glycoprotein, *PNAS*, 2004; 101(31): 11344-9.

Broome C., Effective global response to emerging infectious diseases, *Emerg. Infect. Dis.*, 1998; 4(3).

Bruce-Chwatt L.J., Movements of populations in relation to communicable disease in Africa, *East African Medical Journal*, 1968; 45(5): 266-75.

Carlton J.T., Geller J.B., Ecological roulette: The global transport of non-indigenous marine organisms, *Science*, 1993; 261: 78-82.

Crosby A.W. Jr., *The Columbian Exchange*, Westport (CT): Greenwood Press, 1989.

Daszak P., *et al.*, Emerging infectious diseases of wildlife – threats to biodiversity and human health, *Science*, 2000; 287: 443-9.

Daszak P., Cunningham A.A., Hyatt A.D., Anthropogenic environmental change and the emergence of infectious disease in wildlife, *Acta Tropica*, 2001; 78: 103-16.

De Merode E., *et al.*, The value of bushmeat and other wild foods to rural households living in extreme poverty in Democratic Republic of Congo, *Biol. Conserv.*, 2004; 118(5): 573-81.

Desselberger U., Emerging and re-emerging infectious diseases, *J. Infect.*, 2000; 40: 3-15.

Dow K., Exploring the differences in Our Common Future(s): the meaning of vulnerability to global environmental change, *Geoforum*, 1992; 23: 417-36.

Fearnside P.M., Soybean cultivation as a threat to the environment in Brazil, *Environm. Conserv.*, 1999; 28(1): 23-38.

Guarner J., *et al.*, Monkeypox transmission and pathogenesis in prairie dogs, *Emerg. Infect. Dis.*, 2004; 10: 426-31.

Hahn B.H., Shaw G.M., De Cock K.M., Sharp P.M., AIDS as a zoonosis: scientific and public health implications, *Science*, 2000; 287: 607-14.

Heymann D.L., Rodier F., Hotspots in a wired world: who surveillance of emerging and reemerging infectious diseases, *Lancet Infectious Diseases*, 2001; 1: 345-53.

IOM, *Emerging infections. Microbial threats to health in the United States*, Institute of Medicine, National Academy Press, Washington, DC, 1992.

IOM, *Emerging microbial threats to health in the 21st Century*, National Academy Press, Washington, DC, 2003.

LeDuc J.W., Tikhomirov E., Global Surveillance for recognition and response to emerging diseases, *Ann. N. York. Acad. Sci.*, 1994; 740: 341-345.

Liverman D., Vulnerability to global environmental change, pp. 27-44, In: Kasperson RE, *et al.* (ed.), *Understanding Global environmental Change: the contributions of risk analysis and management*, Clark University, The Earth Transformed program, Worcester, MA, 1990.

Lounibos L.P., Invasions by insect vectors of human diseases, *Annual Review of Entomology 2002*; 47: 233-66.

Ludwig G.V., Calle P.P., Mangiafico J.A., Raphael B.L., Danner D.K., Hile J.A., *et al*., An outbreak of West Nile virus in a New York City captive wildlife population, *American Journal of Tropical Medicine and Hygiene, 2002*; 67: 67-75.

Mayer J.D., Geography, ecology and emerging infectious diseases, *Soc. Sci. & Med.*, 2000; 50: 937-52.

Molineaux D.H., Drivers of emergent parasitic diseases, *J. Parasitol.*, 2003; 89(Suppl.): 53-513.

Moore C.G., Mitchell C.J., *Aedes albopictus* in the United States: ten-year presence and public health implications, *Emerging Infectious Diseases*, 1997; 3: 329-34.

Morse S.S. (Ed.), *Emerging viruses*, Oxford University Press, 1993.

Morse S.S., Factors in the emergence of infectious diseases, *EID*, 1995; 1: 7-15.

Murphy F.A., The threat posed by the global emergence of livestock, food-borne, and zoonotic pathogens. In: Frazier T.W., Richardson D.C., (Eds.), Food and Agricultural Security, New York: *Annals of the New York Academy of Sciences*, 1999; 894: 20-7.

Nisbett R.A., Monath T.P., Viral traffic, transnational companies and logging in Liberia, West Africa, *Global Change & Human Health*, 2001; 2(1): 18-9.

O'Brien T.F., Emergence, spread, and environmental effect of antimicrobial resistance: how use of an antimicrobial anywhere can increase resistance to any antimicrobial anywhere else, *Clinical Infectious Diseases*, 2000; 34(Suppl 3): 78-84.

Okeke I.N., Edelman R., Dissemination of antibiotic-resistant bacteria across geographic borders, *Clinical Infectious Diseases*, 2001; 33: 364-9.

Peeters M., Courgnaud V., Abela B., *et al*., Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat, *Emerging Infectious Diseases*, 2002; 8: 451-7.

Plotkin B.J. & Kimball A.M., Designing and international policy and legal framework for the control of emerging infectious diseases: first step. *EID*, 1997; 3: 1-9.

Patz J.A., *et al*., Effects of environmental change on emerging parasitic diseases, *Intern. J. Parasitol.*, 2000; 30: 1395-1405.

Prusiner S.B., Prions, *Proc. Nat. Acad. Sci.*, USA, 1998; 95: 13363-83.

Rappole J.H., Derrickson S.R., Hubalek Z., Migratory birds and spread of West Nile virus in the Western Hemisphere, *Emerging Infectious Diseases*, 2000; 6(4): 319-28.

Sharp P.M., Bailles E., Chaudhuri R.R., Rodenburg C.M., Santiago M.O., Hahn B.H., The origins of AIDS viruses; where and when?, *Philosophical Transactions of the Royal Society of London Series B*, 2001; 356: 867-76.

Sassen S., *The global city: New York, London, Tokyo*, New Jersey, Princeton University Press, 1991.

Taylor L.H., *et al.*, Risk factors for human disease emergence, *Phil. Trans. Roy. Soc. London, Ser. B*, 2001; 356: 983-9.

United Nations Population Division, *World Population Prospects: The 2000 Revision*, New York: United Nations; 2001, ESA/P/WP, 165, 2001.

Vasconcelos P.F.C., *et al.*, Inadequate management of natural ecosystem in the Brazilian Amazon results in the emergence and resurgence of arboviruses, *Cad. Saúde Pública*, Rio de Janeiro, 2001; 17(supl.): 155-64.

Watts M.J., Bohle J.G., The space of Vulnerability: the causal structure of hunger and famine, *Progr. Human Geogr.*, 1993; 17: 43-67.

Wilder-Smith A., Barkham T.M.S., Earnest A., Paton N.I., Acquisition of W135 meningococcal carriage in Hajj pilgrims and transmission to house contacts: prospective study, *British Medical Journal*, 2002; 325: 365-6.

Wilson M.E., *et al.*, Disease in evolution: global changes and emergence of infectious diseases, *Ann. N. York Acad. Sci.*, 1994; 740: 1-503.

Wilson M.E., Travel and the emergence of infectious diseases, *Emerging Infectious Diseases*, 1995a; 1: 39-46.

Wilson M.E., The traveler and emerging infections: sentinel, courier, transmitter, *Journal of Applied Microbiology*, 2003; 94: 1S-11S.

Wilson M.E., Infectious diseases: an ecological perspective, *British Medical Journal*, 1995b; 311: 1681-4.

Winch P., Social and cultural responses to emerging vector-borne diseases, *J. Vector Ecol.*, 1998; 23(1): 47-53.

Wolfe N.D., *et al.*, Deforestation, hunting and the ecology of microbial emergence, *Global Change & Human Health*, 2000; 1(1): 10-25.

Wolfe NE, *et al.*, Naturally acquired simian retrovirus infections in central African hunters, *Lancet*, 2004; 363: 932-7.

WTO (World Tourism Organization), *World Tourism in 2002*. Available from: URL: www.world-tourism.org, accessed on 2003 Mar. 12.

OUR CHANGING BIOGEOPHYSICAL AND SOCIOECONOMIC ENVIRONMENTS: INFLUENCE ON INFECTIOUS DISEASES

ANTHONY J. McMICHAEL

1. Introduction

The health of human populations – particularly in the medium-to-longer term – depends, fundamentally, on the conditions of the social and natural environments. This statement represents an ecological perspective. It is pitched at the level of whole communities or populations, and thus differs from the view held by most lay individuals and health professionals who typically think of health as a function of personal circumstances and behaviors, genes and access to the health-care system.

During the latter decades of last century there was a growing awareness of environmentally-based risks to health. Today, we must look to even wider environmental horizons. The increasing scale of human impact on the world's environment is now causing unprecedented changes in Earth's ecological and biogeophysical systems, and these necessarily jeopardize the prospects for human health. Of particular concern is the generalized resurgence in infectious diseases. Historically, changes in social and environmental conditions have often fostered the emergence or spread of infections.

Half a century ago, developed countries assumed that infectious disease, the age-old scourge of humankind (the biblical 'Fourth Horseman'), was at last receding. The antibiotic era had begun successfully in the 1940s; vaccine development was accelerating; pesticides were being used to control mosquito populations; and surveillance and control measures (border controls, quarantine, other social controls, public education) were improving and becoming internationally coordinated. Hence, by the early 1970s, various eminent authorities proclaimed the end of the

infectious disease era. In retrospect this was naïve: by late twentieth century a generalized upturn in emerging and resurgent infectious diseases had become evident.

More than 30 new infectious diseases were identified in the final quarter of the 20th century, including HIV/AIDS, Legionnaire's disease, hepatitis C, Nipah virus disease and many viral haemorrhagic fevers [1]. Since 2000, the story has continued worldwide, particularly with the emergence of SARS and the Avian influenza virus in the Asia-Pacific region. Meanwhile, many long-recognized infectious diseases have increased, including tuberculosis, malaria and dengue fever. Cholera, too, continues to enjoy its largest-ever and longest-ever pandemic (see Box 1) – and a new, eighth, cholera pandemic, entailing a new strain of the cholera bacterium, is currently emerging. Meanwhile, mired in poverty and squalor, diarrhoeal disease and acute respiratory infections continue to kill a total of around seven million infants and children every year.

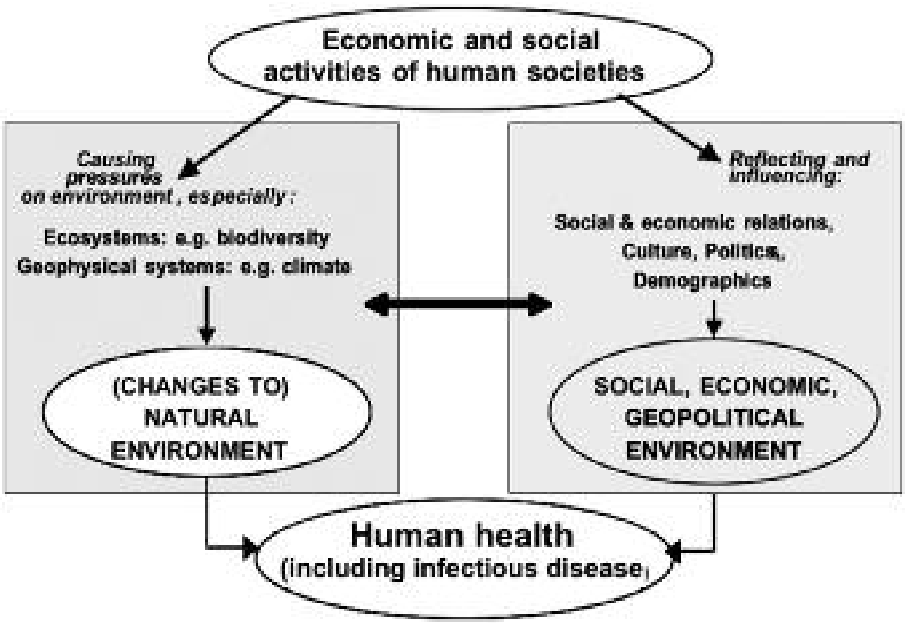


Figure 1. The pathways by which 'social' and 'environmental' changes arise and influence patterns of human health. Note also the interaction between the social and environmental domains.

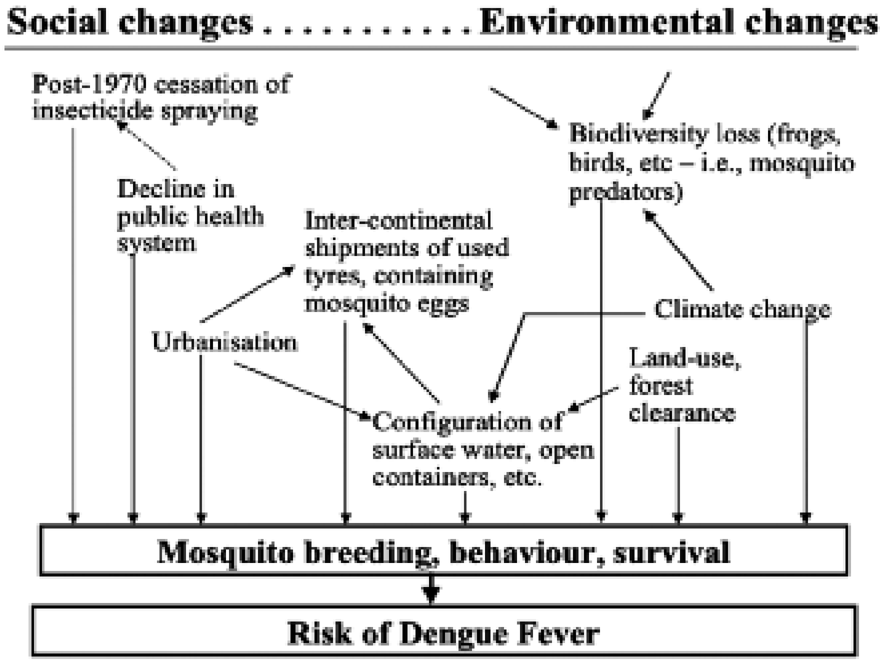


Figure 2. Illustration of the multiple social-cultural changes (e.g., pesticide spraying policy, urbanization) and environmental changes (e.g. land-use, biodiversity loss, climate change) that coexist and interact as influences on the risk of dengue fever occurrence.

Broadly, this upturn reflects a diverse range of distinctive conditions of the modern globalizing world: the growth in population size and density, urbanization, human mobility, long-distance trade, persistent poverty, conflict and warfare, and, increasingly, the advent of human-induced global environmental changes that entail large-scale disturbance and destabilization of ecosystems and biophysical processes (such as the world's climate system) [2].

Figures 1 and 2 show how this mix of (often coexisting) 'social' and 'environmental' factors influences the dynamics of infectious disease emergence, transmission and dissemination.

Figure 1 shows schematically how both 'social' and 'environmental' changes arise as manifestations of human economic and social activity. However, the depiction of the two separate causal pathways is somewhat

misleading, since, as shown with the two-way arrow, there is a variety of interactions between the left and right sides of the schema.

For example, poverty is often associated with over-use of, and damage to, local forests and marginal farmland; and, obversely, local environmental deterioration will often tend to exacerbate poverty and resource deprivation. Further, the levels of material resources, social and human capital, and governmental stability and capacity will influence the extent of vulnerability of local populations to the stressors of environmental change. (In general, poor populations in tropical regions are much more likely to experience an actual increase in vector-borne infectious disease in response to the warming and altered rainfall patterns of climate change than are richer temperate-zone countries with strong public health defenses).

A more specific example of this constellation of large-scale social and environmental influences, as coexistent and interacting determinants of the probability of dengue fever outbreaks, is shown in Figure 2.

The underlying determinants of infectious disease emergence and occurrence can be better understood within both an ecological frame and against an historical background. Since we must continue to live in a microbially-dominated world, we should think in ecological terms, not adversarial militaristic terms, about our relations with bacteria, viruses, protozoa and other tiny parasites. Having learnt that we cannot definitively conquer the disease-inducing microbes, we should now develop a better capacity to anticipate the risks of new or resurgent infectious diseases, while we continue to coexist with an ever-changing microbial world.

In historical terms, there are many recorded episodes and processes in history that can teach us more about the consequences of human actions, of changes in human ecology, for the patterns of occurrence of infectious diseases in human populations.

BOX 1. CHOLERA – THE SEVENTH PANDEMIC

Cholera illustrates how modern social, economic and environmental conditions, having undergone various recent changes at large scale, can affect the pattern of occurrence of infectious diseases.

Cholera originated in the Ganges delta, in India, where epidemics of a cholera-like disease have been described over the past four centuries [3]. Since 1961, a major pandemic of cholera has occurred. This is the seventh pandemic since cholera (apparently reinforced with a newly acquired toxin-producing gene) first extended outside South Asia in 1817. That initial spread followed the Great Kumbh annual religious festival in the Upper Ganges, in which pilgrims from all over India came to bathe in the sacred waters. Their subsequent dispersal, and contacts with British troops mobilizing in the northwest frontier region, caused a cholera epidemic that spread from India to the Arabian Peninsula and along the trade routes to Africa and the Mediterranean coast. In the early 1830s, the faster-traveling steamboats enabled cholera to cross the Atlantic. The disease reached North America in 1832, and spread rapidly around the coastline and inland via major rivers.

This seventh pandemic has reached further than ever before, affecting Asia, Europe, Africa, North America and Latin America. It is by far the longest lasting pandemic to date [4]. This pandemic entails the El Tor strain of the vibrio, which, in mid-twentieth century, replaced the more lethal classical biotype of the nineteenth century pandemics. The scale of this pandemic is likely to reflect the confluence of greatly amplified human mobility between continents, the rapidity and distance of modern shipping-based trade, coincident increases in nutrient enrichment of coastal and estuarine waters by phosphates and nitrates in run-off water (enhancing proliferation of vibrio-harboring phytoplankton and zooplankton), and the growth of urban slums with unsafe drinking water.

2. ECOLOGICAL PERSPECTIVES, WITHIN AN HISTORICAL FRAME

The worldwide dispersal of the human species over the past 50,000-100,000 years, along with cultural evolution and inter-population contacts, has profoundly transformed the relationship between *Homo sapiens* and the microbiological world [5]. During this grand odyssey, there have been several major transitions at key historical junctures, each entailing the emergence of various new or unfamiliar infectious diseases, and occurring at increasingly large scale.

The co-evolution of humans and infectious agents has, of course, a long history. In human prehistory two profound transitions in this relationship occurred: the first when early humans became serious meat-eaters, thereby exposing themselves to various animal parasites, and the second when *Homo sapiens* spread out of Africa into new environments and climates where they encountered unfamiliar microbes. Since the advent of agriculture and animal husbandry, and the dawn of 'history' via written records, three other great transitions have occurred in the human-microbe relationship [2]. First, early agrarian-based settlements enabled enzootic microbes to cross the species barrier and make contact with *Homo sapiens*. Hence, the early city-states of the Middle East, Egypt, South Asia, East Asia and Central and South America each acquired their own distinct repertoire of locally evolving 'crowd' infectious diseases. Second, later, over the course of a thousand years or so the ancient civilizations of greater Eurasia (Egypt, India, Rome, China) made contact, swapped their dominant microbes and painfully equilibrated. (Perhaps a similar process occurred within some other continents, but, if so, the historical record is not available. Eurasia, the largest continent, and with a predominant east-west axis, afforded a particularly auspicious opportunity for inter-civilization contacts.) Third, from the fifteenth century onwards, expansionist sea-faring Europe, technologically pre-eminent in the world, inadvertently exported its lethal, empire-winning, germs to the Americas and later to the south Pacific, Australia and Africa.

Today, we are living through the fourth great transition, which, for the first time, extends globally. The generalized increase in lability in occurrence, spread and biological behavior of infectious diseases largely reflects the combined and increasingly widespread impacts of demographic, commercial, environmental, behavioral, technological and other rapid changes in human ecology. The main characteristics that underlie this contemporary global transition, foreshadowed in the introduction, are summarized in Box 2.

Box 2. The Main Drivers of the Contemporary Global Transition in Human-Microbe Relations

- Increased human mobility: travel, migration (including refugees)
- Long-distance trade
- Ever-larger cities and urban populations:
  - Large, dense populations
  - Freer, non-traditional, sexual contacts
  - IV drug injecting
  - Unsanitary shanty towns and slums
- Intensified food production practices (especially livestock)
- Various modern medical/hospital procedures (injection, transfusion, transplantation)
- Global environmental changes: climate change, land-use change, dam-building and irrigation, ecosystem disruption (including biodiversity loss), etc.

3. Food Sources; Food Production

The spectacular outbreak of 'mad cow disease' (bovine spongiform encephalopathy, BSE) in the UK in the mid-1980s underscored the infectious disease risks inherent in changes in food production methods. Within many cultures, the predilection for eating meat of various exotic species exacerbates the risk of exposure to infectious agents not previously encountered. Indeed, this situation probably triggered the 2003 epidemic of Severe Acute Respiratory Syndrome, SARS, in the Guangzhong region of southern China [6]. Recent studies have implicated bats as the natural reservoir species for the SARS coronavirus. Other wild mammals, including civet cats, have been reported to be infected with SARS-related viruses, and surveys in live markets and restaurants in Guangzhong, southern China, identified various small carnivore species that had been captured in rural China, Laos, Vietnam and Thailand and that were then brought into close proximity with one another [7].

The consumption of bushmeat in Africa poses a serious risk of emerging infectious diseases (as well as a risk for species conservation) [8]. Indeed, it is the sick, infected, animals that are often more easily captured. Molecular genetic studies indicate that the HIV virus has crossed from chimpanzees into humans at least three times [9], and it is plausible that one or more of these cross-over events arose from butchering the animals.

Southern China, with its traditional close-contact animal husbandry, has long been implicated in the origin of epidemic influenza virus strains via the juxtaposition of ducks, chickens, pigs and humans. The generation of novel zoonotic viral strains occurs readily in this environment. Over recent decades, the human population and the animal populations needed to support them have rapidly increased. The numbers of poultry and pigs are now at record levels and, as wealth increases and dietary preferences 'modernize', those numbers are set to double approximately every 10 years [10].

4. Social, Economic and Associated Influences

Each year 17 million people, mostly young children, die from infectious diseases. Worldwide, infectious diseases account for almost one in three of all deaths. However, the discrepancy between rich and very poor countries is huge: infections cause 1-2% of all deaths in the former, but more than 50% in the latter. Diseases such as tuberculosis, leprosy, cholera, typhoid and diphtheria are known to be pre-eminently diseases of poverty. Malaria, tuberculosis and dengue fever have all increased their compass over the past 20 years, particularly within poorer communities or groups, [11], [12]. Similarly, urbanization, long-distance travel and freer sexual relations have all amplified the spread of various such diseases. As happened historically with tuberculosis, HIV infection seems now to be entrenching itself among the world's poor and disempowered, especially in sub-Saharan Africa and South Asia. Much of the spread of HIV has been along international 'fault lines', tracking the inequality and vulnerability that accompany migrant labor, educational deprivation and sexual commerce [12].

The health ramifications of economic disadvantage have been further highlighted by some recent infectious disease outbreaks attributed, in part, to the impacts of free trade agreements. For example, in the 1990s several outbreaks of hepatitis A and cyclosporiasis (a protozoal infection) occurred in the United States from fecally-contaminated strawberries and raspberries imported from Central America. Such contamination is likely to have occurred because of the introduction of the North American Free Trade Agreement. This, like other such agreements, tends to subordinate environmental and labor standards (including worker access to toilet facilities) to the primacy of profitability.

Socially disordered populations living in circumstances of privation, unhygienic conditions and close contact, are susceptible to infectious dis-

ease. History abounds with examples. The severity of the bubonic plague (Black Death) in mid-fourteenth century Europe appears to have reflected, in part, the malnutrition and impoverishment caused by several preceding decades of unusually cold and wet weather with crop failures. This, in conjunction with the incipient destabilization of the hierarchical feudal system, would have heightened the vulnerability of the European populations to epidemic disease.

In modern times, the urban environment and associated ways of life has become an increasingly important influence on infectious disease patterns around the world. Cities are now the dominant human habitat. Urbanism typically leads to a breakdown in traditional family and social structures, and entails greater personal mobility and extended and changeable social networks. These features along with access to modern contraception have facilitated a diversity of sexual contacts and, hence, the spread of STDs. This risk is further amplified by the growth in sex tourism in today's internationally mobile world, which capitalizes on the desperation and ignorance of poverty, combined with exploitative behaviors, in developing countries. More generally, cities act as highways for 'microbial traffic' [13].

Rural-urban migration is fuelled by the primary drive to enter the cash economy, allied with the burgeoning international demand for skilled and unskilled workers in a globalizing marketplace. Rapid, unplanned, urbanization exacerbates old infectious diseases such as childhood pneumonia, diarrhea, tuberculosis and dengue. It also facilitates the spread of various 'emerging' diseases; for example, high-rise housing can create new infectious disease risks, as was recently observed for SARS in Hong Kong. Such housing also increases risks of infection via the consequences of family breakdown and social instability, intravenous drug abuse and sexual transmission of infections [14].

4.1. *Travel and Trade*

Microbes are no respecters of political and administrative borders. The mobility of humans, animals and birds is a constant stimulus to changes in the pattern of infectious disease occurrence. HIV/AIDS has spread quickly around the world in the past twenty years. SARS spread readily from Hong Kong to Vietnam, across to Germany and Toronto [6]. Vector mosquito species can travel with trade and transport – as apparently did the bubonic plague-infected black rat, traveling westwards across the Silk Road towards the Black Sea and Europe in the fourteenth century.

Dengue fever, numerically the most important vector-borne viral disease of humans, illustrates well how patterns of trade, travel and settlement can influence various infectious diseases. Although dengue is primarily a tropical disease, its extension in recent decades into various temperate countries reflects both the introduction of the disease's main mosquito vector species, *Aedes aegypti* (which is behaviorally adaptable to a cold climate), and the increase in imported cases resulting from increased air travel [15]. It also reflects the rapid evolutionary adjustment of this mosquito species to coexistence with urban-dwelling humans, having originated in forest Africa. Indeed, *Aedes aegypti* has followed humankind on its travels and migrations around the world [16]. A major alternate mosquito vector for the dengue virus, *Aedes albopictus* (the 'Asian tiger mosquito'), has been disseminated widely in recent years via the unwitting intercontinental exportation of mosquito eggs in used car tires from Asia into Africa and the Americas [17].

Neisseria meningitidis, a global bacterial pathogen, causes seasonal epidemics of meningitis in the 'meningitis belt' of Sahelian Africa. Molecular marker studies have revealed that, in recent decades, Muslim pilgrims brought an epidemic strain of *N. meningitidis* from southern Asia to Mecca. In Mecca they passed it on to pilgrims from sub-Saharan Africa who, after returning home, initiated strain-specific epidemic outbreaks in several locations [18].

The globalization of the food market has potentiated the movement of pathogens between regions. As examples, first, an outbreak of cholera in Maryland, USA, was traced to imported contaminated frozen coconut milk [19], and, second, alfalfa sprouts grown from contaminated seed sent to a Dutch shipper caused outbreaks of infections with *Salmonella* species in both the USA and Finland [20].

4.2. *Land Use and Environmental Change*

Like other very large mammals, we humans are 'patch disturbers'. This we do via tropical deforestation, irrigation, dam building, urban sprawl, road building, intensified food production systems, and pollution of coastal zones. The increasing scale of this encroachment on the natural environment accelerates the emergence of new infectious diseases. As we spread into the last corners of Earth's tropical forests, new contacts occur between wild fauna and humans (and their livestock), increasing the risk of cross-species infection.

Habitat fragmentation and biodiversity changes alter the risks of infectious disease. This is well illustrated by deforestation that results in habitat fragmentation – and an increase in the 'edge effect' that promotes pathogen-vector-host interaction. This process has contributed, in recent decades, to the emergence of a number of viral haemorrhagic fevers in South America. Various such viral infections have been caused by arenaviruses. Major examples, mostly in rural populations, have been described in Argentina (Junin virus), Bolivia (Machupo virus) and Venezuela (Guanarito virus) [21], [22], [23]. The Machupo virus is an example. Forest clearance in Bolivia in the 1960s, accompanied by blanket spraying of DDT to control mosquitoes, led, respectively, to infestation of cropland by *Calomys* mice and to the poisoning of the rodents' usual predators (village cats). The consequent proliferation of mice and their viruses resulted in the appearance of a new viral fever, the Bolivian (Machupo) haemorrhagic fever, which killed around one seventh of the local population.

More generally, rodents, responding to environmental disturbances, are a prime source of new and re-emerging infections. Consider the emergence of Hanta viruses as a source of human disease in the USA in the 1990s. In mid-1993, an unexpected outbreak of acute, sometimes fatal,

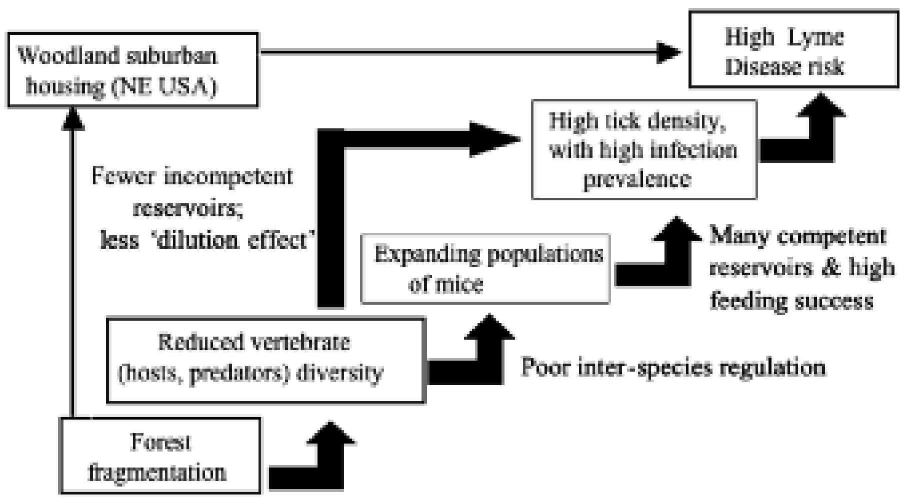


Figure 3. Schematic representation of the sequence of environmental and ecological changes that, along with social-residential changes, increase the probability of Lyme Disease transmission in northeast USA (based on the work of R. Ostfeld).

respiratory disease occurred in southwest USA [24]. This ‘hantavirus pulmonary syndrome’ was caused by a previously unrecognized virus, maintained in the natural environment primarily within native deer-mice and transmitted via rodent excreta. The 1991-92 El Niño event, causing unseasonally heavy summer rains and a proliferation of piñon nuts within southwest USA, hugely amplified local rodent populations, leading to the 1993 outbreak of ‘hantavirus pulmonary syndrome’ [24], [25].

In the USA, nature conservation and increased contact with woodland in the Eastern states has led to the emergence of Lyme disease (borreliosis) – summarized in Figure 3. The ticks that transmit the spirochaete *Borrelia burgdorferi* normally feed on deer and white-footed mice. The latter are the more competent host species. However, forest fragmentation has caused the loss of various predator species – wolves, raptors, and others – with a resultant shift of ticks from the less competent to the more competent host species (i.e., from deer to white-footed mice). These changes, along with suburban sprawl into woodlands, have interacted in the emergence of this disease [26], [27].

5. Global Environmental Changes and Infectious Disease Risk

The large-scale environmental changes that humans are now imposing on the biosphere have great implications for the future pattern of infectious disease. Much research in this emerging topic area has been done specifically in relation to global climate change – although it is now well understood that there will be widespread interactions between the impacts of these environmental changes. A further consideration is that the disease impact of changes in environmental conditions will usually be modulated by the level of susceptibility of the human population – reflecting population density, immune status, nutritional status, extent of mobility, level of social organization/disorganization, flexibility of political systems and of governance, and various other such social-environmental factors.

5.1. *Global Climate Change*

The now-certain prospect of human-induced global climate change raises long-term questions about how infectious diseases will respond over the coming century (and beyond). Many infectious agents, their vector organisms and their reservoir non-human species, are sensitive to cli-

matic conditions and to resultant environmental changes. The example of hantavirus pulmonary syndrome has been discussed above.

Salmonella food-poisoning is known to be very temperature-sensitive, and diarrhoeal diseases peak in summer. In the Asia-Pacific region, El Niño fluctuations appear to affect the occurrence of dengue fever, the world's most prevalent vector-borne viral disease, spread primarily by the *Aedes aegypti* mosquito. Similarly, interannual variations in climatic conditions in Australia, especially those due to the El Niño cycle, influence the pattern of outbreaks of Ross River virus disease [28], [29].

Climate change, via both a shift in background climate conditions and changes in regional climatic variability, will affect the spatial and seasonal patterns of the potential transmission of various vector-borne infectious diseases. These would include: malaria, dengue fever, various types of viral encephalitis, schistosomiasis (spread by water-snails), leishmaniasis (spread by sand-flies in South America and around the Mediterranean coast), onchocerciasis (West African 'river blindness', spread by black flies) and yellow fever (also spread by the *Aedes aegypti* mosquito). The key phrase here is *potential transmission*. It is relevant to estimate how the intrinsic infectious disease transmission properties of the world would alter in response to climate change. Indeed, such research is in the classic tradition of experimental science, which seeks to hold everything else constant while estimating the effect of varying just one key factor. Nevertheless, we know that the *actual* transmission of diseases such as malaria is, and will be, much affected by economic and social conditions and by the robustness of public health defenses. Hence, we also need to develop methods of modeling that can incorporate other reasonably foreseeable contextual changes.

Both statistical and biologically-based ('process-based') models have been used, to assess how shifts in ranges of temperature and patterns of rainfall would affect the transmission potential of various vector-borne diseases [30], [31]. However, this type of scenario-based modeling of future risks has not yet attempted to address all aspects of the 'scenario'. For example, how will domestic and urban water use (particularly relevant to dengue fever occurrence) change in a warmer world with altered patterns of precipitation? How would an increase in the tempo of extreme weather events and natural disasters affect infectious disease occurrence? We still have many things to learn about how the impending shift into unfamiliar climatic conditions will affect the complex processes of infectious disease transmission, especially the vector-borne diseases.

Meanwhile, there is now appearing suggestive evidence that the climate change that has occurred over the past 30 years has influenced cholera outbreaks in Bangladesh [32], the extension of tick-borne encephalitis in Sweden [33] and, more debatably, the range and seasonality of malaria in some parts of eastern Africa [34].

6. The Modern World: Too Hygienic?

There is, finally, one other aspect of modernity that we should note. It, too, reflects a shift in human ecology. The pattern of exposure to infectious agents in childhood can perturb the human immune system in two ways: by influencing its developmental pathway, and by initiating autoimmune disorders. Modern living entails a more hygienic childhood, with the elimination or deferral of childhood infections that have long influenced the development of the young immune system. Most dramatically, this led to the rise of polio in developed countries in the 1950s because children were no longer exposed to the polio virus in early childhood when the infection is normally harmless. More subtly, this increase in hygiene, along with fewer siblings, has reduced the intensity of childhood exposure to a range of infectious agents, including the many commensal bacteria that, historically, colonized the infant gut. This significant shift in human ecology affects the early-life programming of the immune system, inclining it towards a more 'atopic' (allergic) pattern of response. This may help explain the recent widespread rise of childhood asthma and hay fever. Likewise, the progressive elimination of bowel parasites such as round-worms, whip-worms and pin-worms (parasites which coevolved with hominids over several million years) may have contributed to the rise in inflammatory bowel diseases, including Crohn's disease and ulcerative colitis, in Western populations over the past half century.

Infectious agents can also influence the occurrence of autoimmune disorders, in which the immune system erroneously attacks normal body tissues. A basic evolutionary device of parasitic microorganisms is to acquire an outer protein surface that resembles that of the host's tissue. This 'molecular mimicry' provides camouflage against attack by the host immune system, since proteins recognized by the host immune system as 'self' are not normally attacked. Occasionally the immune system gets it wrong, and attacks both the microorganism and the part of 'self' that it resembles. It is likely that insulin-dependent diabetes (which usually

begins in childhood), multiple sclerosis and rheumatoid arthritis all involve this type of viral infection-triggered autoimmune mechanism.

7. Conclusion

As ever, the world is replete with microbes jostling for supplies of nutrients, energy and molecular building blocks. From the microbe's viewpoint, the appropriate microbe (often in mutant form), fortuitously in the right place at the right time, can extend, re-start or even found a dynasty. It has happened many times before and it will continue to do so.

As the scale of human impact on the biosphere escalates, and as the structures and fluidity of human societies change along with the levels of susceptibility of local human populations, so, perennially, these environmental and social-economic changes create opportunities for infectious agents, both new and resurgent. Today, this process is occurring at an accelerated rate, and, increasingly, on a global scale, as we undergo a fourth, and larger-than-ever, transition in the overall relationship between the human species and the microbial world.

If we are to achieve a more enlightened, ecologically attuned, coexistence with the microbial world, and a capacity to anticipate and minimize infectious disease risks, then the main challenges for researchers are to develop a broadly-based inter-disciplinary collaboration and a capacity to deal with complexities, uncertainties and spatial-temporal scales that extend beyond conventional research practice.

REFERENCES

1. Institute of Medicine (IOM), *Microbial Threats to Health*, Washington DC: IOM, 2003.
2. McMichael, A.J., Environmental and social influences on emerging infectious diseases: past, present and future, *Philosophical Transactions Royal Society London B: Biological Sciences* 2004; 359(1447): pp. 1049-58.
3. Speck, R.S., Cholera, in *The Cambridge World History of Human Disease* (ed. Kiple, K.F.). pp. 642-47 (Cambridge University Press, Cambridge, 1993).
4. Lee, K. and R. Dodgson, Globalization and cholera: implications for global governance, *Global Governance 2000*, 2000. 6: pp. 213-36.
5. McNeill, W., *Plagues and Peoples*, 1976, Middlesex: Penguin.

6. Peiris, J.S. and Y. Guan, Confronting SARS: a view from Hong Kong, *Philos. Trans. R. Soc. Lond. B: Biol. Sci.*, 2004. 359(1447): pp. 1075-9.
7. Bell, D., S. Roberton, and P.R. Hunter, Animal origins of SARS coronavirus: possible links with the international trade in small carnivores, *Philos. Trans. R. Soc. Lond. B: Biol. Sci.*, 2004. 359(1447): pp. 1107-14.
8. Weiss, R. and McMichael, A.J., Social and environmental risk factors in the emergence of infectious disease, *Nature Medicine*, 2004; 10: pp. 70-6.
9. Hahn, B.H., *et al.*, AIDS as a zoonosis: scientific and public health implications, *Science*, 2000. 287(5453): pp. 607-14.
10. Sleigh, A., International Teamwork in Response to Emerging Infections in the Asia-Pacific, in *Biosecure 2004 Conference*, 2004, Canberra.
11. Fineberg, H. and M. Wilson, Social vulnerability and death by infection, *N. Engl. J. Med.*, 1996. 334: pp. 859-60.
12. Farmer, P., *Infections and Inequalities. The Modern Plagues*, 1999, Berkeley: University of California Press.
13. Morse, S.S., Global microbial traffic and the interchange of disease, *Am. J. Public Health*, 1992. 82(10): pp. 1326-7.
14. Cohen, J., *How Many People Can the Earth Support?*, 1995, New York: Norton.
15. Kuno, G., Review of the factors modulating dengue transmission, *Epidemiol. Rev.*, 1995. 17: pp. 321-35.
16. Monath, T., Dengue: The risk to developed and developing countries, *Proc. Natl. Acad. Sci.*, USA, 1994. 91: pp. 2395-2400.
17. Reiter, P. and D. Sprenger, The used tire trade: a mechanism for the worldwide dispersal of container-breeding mosquitoes, *Journal of American Mosquito Control Association*, 1987. 3: pp. 494-501.
18. Moore, P., *et al.*, Intercontinental spread of an epidemic group A Neisseria meningitidis strain, *Lancet*, 1989. 2: pp. 260-3.
19. Taylor, J.L., *et al.*, An outbreak of cholera in Maryland associated with imported commercial frozen fresh coconut milk, *J. Infect. Dis.*, 1993. 167(6): pp. 1330-5.
20. Mahon, B.E., *et al.*, An international outbreak of Salmonella infections caused by alfalfa sprouts grown from contaminated seeds, *J. Infect. Dis.*, 1997. 175(4): pp. 876-82.
21. Maiztegui, J.I., Clinical and epidemiological patterns of Argentine haemorrhagic fever, *Bull. World Health Organ*, 1975. 52(4-6): pp. 567-75.
22. Simpson, D., Viral haemorrhagic fevers of man, *Bull. WHO*, 1978. 56(6): pp. 819-32.

23. Salas, R., *et al.*, Venezuelan haemorrhagic fever, *Lancet*, 1991. 338(8774): pp. 1033-6.
24. Parmenter, R., *et al.*, *The hantavirus epidemic in the southwest: rodent population dynamics and the implications for transmission of hantavirus-associated adult respiratory distress syndrome (HARDS) in the four corners region*, Sevilleta LTER Publication, 1993. 41.
25. Engelthaler, D.M., *et al.*, Climatic and environmental patterns associated with hantavirus pulmonary syndrome, Four Corners region, *United States, Emerg. Infect. Dis.*, 1999. 5(1): pp. 87-94.
26. Glass, G., *et al.*, Environmental risk factors for Lyme disease identified with geographical information systems, *Am. J. Public Health*, 1995. 85(7): pp. 944-8.
27. Schmidt, K. and R. Ostfeld, Biodiversity and the dilution effect in disease ecology, *Ecology*, 2001. 82: pp. 609-19.
28. Woodruff, R.E., *et al.*, Predicting Ross River virus epidemics from regional weather data, *Epidemiology*, 2002. 13: pp. 384-93.
29. Tong, S., *et al.*, Climate variability and Ross River virus transmission, *Epidemiology*, 2002. 13(4): p. 2.
30. Martens, P., *et al.*, Climate change and future populations at risk of malaria, *Global Environmental Change-Human and Policy Dimensions*, 1999. 9: pp. 89-107.
31. Hales, S., *et al.*, Potential effect of population and climate changes on global distribution of dengue fever: an empirical model, *Lancet*, 2002. 360: pp. 830-4.
32. Rodo, X., *et al.*, ENSO and cholera: a non-stationary link related to climate change?, *Proc. Nat. Acad. Sci.*, Early Edition, 2002.
33. Lindgren, E. and R. Gustafson, Tick-borne encephalitis in Sweden and climate change, *Lancet*, 2001. 358: pp. 16-8.
34. Patz, J., *et al.*, Regional warming and malaria resurgence, *Nature*, 2002. 420: pp. 627-8.

FOOD, WATER, HEALTH, AND INFECTIOUS DISEASES: FOCUS ON GLOBAL CHANGE

MAHENDRA M. SHAH, GUENTHER FISCHER,
HARRIJ van VELTHUIZEN

1. Introduction

In spite of numerous international commitments made over the last three decades to improve food, water, and health security in the developing world, some 840 million people remain undernourished, over 1.2 billion lack access to safe water, and more than 2.5 billion are at risk of infectious diseases, of which just six account for over 90% of the 13 million deaths a year. The world population is projected to increase by some 50% to reach 9 billion over the next five decades. This growth, combined with economic growth and increasing consumption and pollution, will result in substantial ecological pressures on land, water, the atmosphere, and biological resources. Global environmental change, particularly climate change, will lead to ecological changes that will likely cause higher levels of disease-causing pathogens and parasites.

There is a critical need for integrated climate, ecology, economy, and demographic (CEED) assessments to identify those ecosystems and populations that will be most at risk under conditions of global environmental change. These assessments require spatial analyses by differential vulnerability of ecosystems and populations (Lutz and Shah, 2002; Lutz, 2006) to target policies that may respond to and mitigate the threats of food, water, and health insecurity.

This paper presents an integrated CEED methodology and a policy-modeling framework for the assessment of the evolution of the world food system in the 21st century, under various future scenarios of population growth, economic growth, and environmental change. A global ecological database, comprising spatial data on land, water, forests, population, and

habitation for all countries, is combined with national and regional general equilibrium models that are embedded in the global agricultural economy. This CEED systems analytic approach, particularly the spatial global ecological database and methodology, is also suited to the assessment of the extent, intensity, and location, as well as to the analysis of policy options, for future water insecurity and infectious disease distribution and risks.

2. Food and Hunger

Adequate nourishment is universally recognized as a fundamental human right. At the global level, there is enough food to meet everyone's need, and yet a fifth of the world's population continues to be chronically undernourished. Every year some 15 million people die from hunger alone, and over 200 million suffer health consequences due to nutritional deficiencies including those of proteins, micronutrients, and essential amino acids. Balanced nutrition is the foundation of good health, and healthy people are less susceptible to many infections and diseases.

In 1974, world political leaders set the goal of eradicating hunger within a decade. A quarter of a century later, at the Millennium Summit in 2000 and the World Food Summit in 2002, world leaders endorsed the less ambitious goal of reducing hunger by half by 2015, even though the rate of progress over the previous ten years indicated that it would take more than 60 years to reach this target. Political goals have a role and relevance in this effort, but in the face of continued shortfalls there is a limit to the hope and trust of the hundreds of millions who spend their lifetimes in debilitating hunger.

In addition to the intractable problem of world hunger, there is also an emerging problem of over-consumption, which is resulting in obesity and related health disorders such as diabetes and cardiovascular disease. More than 800 million people worldwide are estimated to be obese. The globalization of the world food system and the increasing influence of large food corporations – processors, distributors, and retailers – are affecting food consumption patterns through the increased marketing of processed foods that contain unhealthy levels of fat, salt, and sugar.

The next world food crisis will no doubt be a human health crisis, with conditions of too little food or too much unhealthy food affecting people differently in all countries, developing and developed. The scientific community, civil society, national governments, and the international development

community bear the fundamental responsibility to achieve nutritionally sound, productive, and sustainable food agriculture. The key challenge will be in linking food sciences, agricultural technology, land, water, biodiversity management, and national and international policies and actions through all stages of the food chain, from production to marketing to consumption.

Agriculture must be given the highest political commitment and attention, because producing food is the predominant use for environmental and natural resources, and it has the greatest impact on the sustainability of ecosystems and their services. The trend in reduced allocation of national development budgets to agriculture – for example, in extension services and training, marketing, and infrastructure – together with declining multilateral lending and bilateral aid for this sector exemplifies the fact that agriculture is not given the attention and commitment it requires. There can be no progress toward reducing hunger and poverty without sustainable resource management and a science-based policy commitment to agriculture (Shah, 2001).

2.1. *The Challenge for Agricultural Science*

The promise of science to improve societies and human well-being cannot be delivered unless science is relevant to real, practical, and people-centered issues. The scientific community and the development policy community at the national and international levels must work expeditiously towards the goal of achieving health-enhancing food systems that are socially, economically, and environmentally viable and sustainable. This will require a systemic combination of the relevant sciences, including biology and biochemistry, agro-ecology and environmental science, social and economic science, as well as the fields of informatics and knowledge communication.

The challenge to the biological sciences is to combine the best of conventional breeding technology with biochemistry and safe, ethical molecular and cellular genetic research to develop nutritionally enhanced and productive germplasm (Shah and Strong, 2000). The specific food crops used by the poor, including coarse grains, roots and tubers, and plantains and bananas, should be given the highest priority. There is considerable scope for the development or improvement of environmentally sound fish farming and intensive livestock production, with due consideration of potential health hazards and animal welfare.

Targeted research, including biotechnology, has the potential to overcome many environmental constraints such as soil toxicity, water limita-

tions, and pests and diseases, as well as to increase the nutritional content of crop plants. Vitamin A and iron fortification of rice can reduce the risk of blindness that affects 8 million people annually, prevent the death of 2 million infants disposed annually to diarrhea and measles, and lower the prevalence of anemia that affects 30% of the world's population.

Agro-ecological sciences, together with recent developments in geographical information systems, including remote sensing and the increasing quality and spatial coverage of sub-national, national, and global resource databases – of soils, water resources, land cover, thermal regimes and rainfall patterns, population distribution, etc. – has enabled productivity assessment tools to identify the potential of and environmental constraints to crop production at regional and national levels. The integration of this information into assessments of the global food economy, together with projections of future climate change and variability, enables evaluation of the possible impacts of climate change on agriculture and provides a basis for prioritizing regional and commodity-specific agricultural research for adaptation to and mitigation of the agricultural effects of climate change.

Climate change and variability will result in irreparable damage to arable land and water resources, with serious consequences for food production. Most of these losses will occur in developing countries where the capacity to cope and adapt is limited. Although the international community has focused on climate change mitigation, adaptation to climate change is an equally pressing issue and must be put on the international agenda. Adaptation is of critical importance to the many developing countries that have contributed little to greenhouse gas emissions thus far, yet will bear the brunt of the negative impacts of climate change and variability.

The challenge in the social sciences and economics is to ensure an environment that enables and empowers farmers and consumers to benefit from advances in science. In developing countries, health services, food aid, education and training, extension services, marketing, and infrastructure development are priority areas. The information and communication revolution has a significant role to play in developing global agricultural systems by combining the best of science with traditional knowledge. However, the fact that less than 1% of the population in Africa and developing Asia have access to the Internet, in comparison to about half of the population in the developed world, will further inhibit progress towards sustainable agriculture. This digital divide must be overcome because otherwise, global disparities will widen further.

2.2. *Integrated CEED Assessment of the World Food System*

Sustainability of land and water resources is significant to many central themes and issues in the study of global environmental change. Alterations in the earth's surface have major implications for the global radiation balance and energy fluxes, contribute to changes in biogeochemical cycles, alter hydrological cycles, and influence ecological balances and complexity. These environmental impacts at local, regional, and global levels, driven by human activity, have the potential to significantly affect food and water security and the sustainability of agro-ecological systems.

Food production systems interact with land and water resources, forest ecosystems, and biodiversity, and are susceptible to the effects of climate change. Ensuring soil fertility, genetic diversity, agricultural water resource management, and adapting to the impacts of climate change is critical to enhancing production.

The Food and Agriculture Organization of the United Nations (FAO) and the International Institute for Applied Systems Analysis (IIASA) have over the last two decades developed integrated CEED analytical tools and global databases. The focus of these efforts has been on multi-disciplinary scientific research that analyzes the current and future availability and use of regional and global land and water resources, in the face of local, national, and super-national demographic and socioeconomic change, international trade and globalization, technological development, and environmental changes including climate change and climate variability.

We assess the sensitivity of agro-ecosystems to climate change, as determined by the FAO/IIASA Agro-ecological Zones (AEZ) model (Figure 1, see page 413) within the socio-economic scenarios defined by the Intergovernmental Panel on Climate Change (IPCC) Special Report on Emissions (SRES). For this purpose, IIASA's global linked model of the world food system is used. This modeling framework, referred to as the Basic Linked System (BLS), comprises a representation of all major economic sectors and views national agricultural systems as embedded in national economies, which in turn interact at the international level. The combination of AEZ and BLS provides an integrated ecological-economic framework for the assessment of the impact of climate change. We consider climate scenarios based on experiments with four General Circulation Models (GCM) and assess the four basic socioeconomic development pathways and emission scenarios formulated by the IPCC in its Third Assessment Report.

The AEZ methodology for land productivity assessments follows an environmental approach; it provides a framework for establishing a spatial inventory and database of land resources and crop production potentials. This land-resources inventory is used to assess, for specified management conditions and levels of inputs, the suitability of crops/Land Utilization Types in relation to both rain-fed and irrigated conditions, and to quantify expected production of cropping activities relevant in the specific agro-ecological context. The characterization of land resources includes components of climate, soils, landform, and present land cover. Crop modeling and environmental matching procedures are used to identify crop-specific environmental limitations, under various levels of inputs and management conditions. The framework contains the following basic elements:

– Geo-referenced climate, soil, and terrain data, which are combined into a land-resources database. The computerized global AEZ database comprises some 2.2 million grid-cells;

– Selected agricultural production systems, with defined input and management relationships, and crop-specific environmental requirements and adaptability characteristics. AEZ distinguishes more than 180 crop, fodder, and pasture land-use types, each at three generically defined levels of inputs and management;

– Accounting of spatial land use and land cover, including forests, protected areas, population distribution and density, and land required for habitation and infrastructure;

– Multi-criteria analysis of agricultural production options.

The BLS (Figure 2) is a global general equilibrium model system for analyzing agricultural policies and food system prospects in an international setting. BLS views national agricultural systems as embedded in national economies, which interact through financial flows and trade at the international level. The BLS consists of 18 single-country national models, two models for regions with close economic co-operation (one for the European Union and one for the countries of Eastern Europe and the former Soviet Union), and a further 14 regional models for the rest of the world's countries. The BLS national and regional models simulate the behavior of producers, consumers, and the government. They distinguish two broad sectors, agriculture and non-agriculture, and the agriculture sector produces nine aggregate commodities.

The individual national and regional models are linked together by means of a world market. The model is formulated as a recursively dynam-

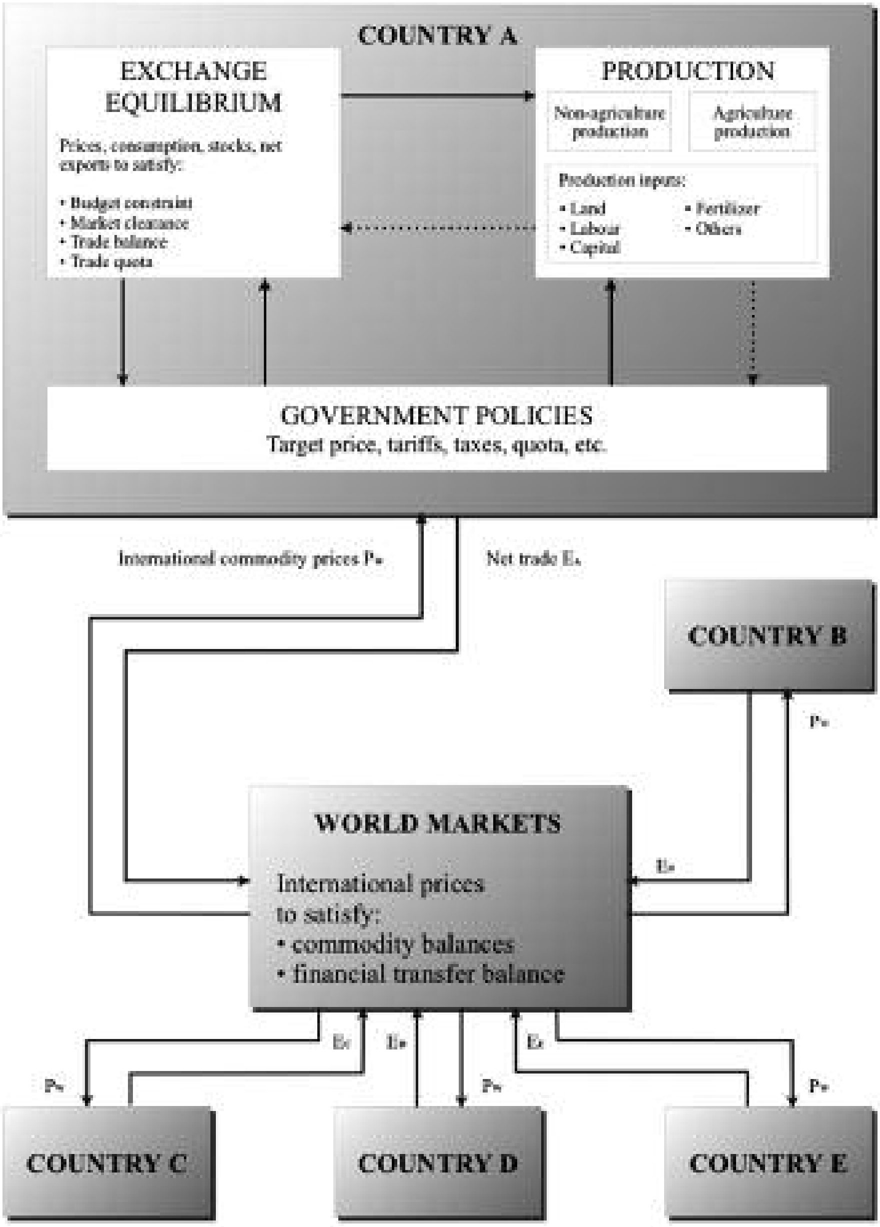


Figure 2. The International Linkage in the BLS (Fischer *et al.*, 2002b).

ic system, working in annual steps, where the outcome of each step is affected by the outcomes of earlier ones. Each individual model covers the whole economy of the respective geographical area. For the purpose of international linkage, production, consumption, and trade are aggregated to nine agricultural sectors and one non-agricultural sector. All physical and financial accounts are balanced and mutually consistent, including the production, consumption, and financial accounts at the national level, and the trade and financial flows at the global level.

The integrated CEED methodology and database (Figure 3, see page 414) provide a foundation for detailed country studies. The main results of the IIASA study include climate-change effects on the prevalence of environmental constraints to crop agriculture; climate variability and the variability of rain-fed cereal production; changes in potential agricultural land; changes in crop production patterns; and the impact of climate change on cereal production potential. Results of the AEZ-BLS integrated ecological-economic analysis of climate change on the world food system includes quantification of the scale and location of hunger, international agricultural trade, prices, production, land use, etc. The analysis assesses trends in food production, trade, and consumption, and the impact on poverty and hunger of alternative development pathways and varying levels of climate change. Some highlights of results are summarized below:

Environmental Constraints to Crop Cultivation: Two-thirds of global land surface – some 8.9 billion hectares – suffers severe constraints for rain-fed crop cultivation: 13.2% is too cold, 26.5% is too dry, 4.6% is too steep, 2.0% is too wet, and 19.8% has poor soils. Climate change will have positive and negative impacts, as some constraints may be alleviated while others may increase. The results for the Hadley HadCM3 climate model and the IPCC A1F1 scenario indicate that the constraints listed above will change respectively to 5.2%, 29.0%, 1.1%, 5.7%, and 24.5%. The agro-ecological changes due to climate change will result in water deficits in some areas and surpluses in others, and will alter the levels of infestation of disease pathogens and parasites.

Land with Cultivation Potential: In Asia and Europe, the rain-fed land currently under cultivation amounts to 90% of the land that is potentially suitable or very suitable for agricultural production. In North America, some 75% of the potentially suitable or very suitable land is currently under cultivation. By contrast, Africa and South and Central America are estimated to have some 1 billion hectares of potentially suitable or very suitable land in excess of the some 350 million hectares of currently cultivated land. However,

most of this additional cultivable land is concentrated in just seven countries – Angola, Congo, Sudan, Argentina, Bolivia, Brazil, and Colombia.

Cultivation Potential in Forest Ecosystems: About a fifth of the world's land surface – some 3 billion hectares – is forest ecosystems. Sixty percent of the world's forestland is located in eight countries – Russia, Brazil, Canada, the United States, China, Australia, Congo, and Indonesia. During the past decade, some 127 million hectares of forests were cleared, while some 36 million hectares were replanted. Africa lost some 53 million hectares of forest during this period, primarily from the expansion of crop cultivation. The study results show that some 470 million hectares of land in forest ecosystems have crop cultivation potential. However, cultivating this land would have serious environmental implications, as forests play a critical role in watershed management and flood control and the protection of biodiversity, and serve as carbon sinks.

Climate Change and Fragile Ecosystems: The world's boreal and arctic ecosystems are likely to decline by 60% due to a northward shift of thermal regimes. The semi-arid and arid land areas in developing countries may increase by about 5 to 8%. Over a fifth of Africa's population, some 180 million people, currently live in these areas, where they derive their livelihoods from agriculture.

Climate Change Impact on Food Production: Although strong gains in crop production potential occur in North America and the Russian Federation, significant losses are projected for Africa, particularly North and Southern Africa. Currently there are some 78 food insecure countries. The current population of these countries is 4.2 billion, of which 18% are undernourished, and their total population in the 2080s is projected at 6.8 billion. Of particular concern are some 40 developing countries, many of them in the least developed country group, which as a whole may lose some 10 to 20% of their cereal production potential by the 2080s due to climate change.

Climate Change and Agricultural GDP and Trade: By 2080s, climate change will reduce Asia's agricultural GDP by 4%, Africa's by around 8%; meanwhile, North America's agricultural GDP will increase by up to 13%. The results also reveal a growing dependence of developing countries on net annual cereal imports – between 170 and 430 million tons. Climate change will add to this dependence, increasing imports of developing countries by up to 40%.

Climate Change and Number of People at Risk of Hunger: Some fairly robust conclusions emerge from the analysis of climate-change impacts on the number of people at risk of hunger. First, climate change will most

likely increase the number of people at risk of hunger. Second, the importance and significance of the climate-change impact on the level of undernourishment depends entirely on the level of economic development assumed in the SRES scenarios. For the wealthy societies of IPCC development path scenario A1 – where even the currently poor regions are assumed to reach economic levels exceeding in per capita terms the current average OECD income – hunger is a marginal issue and remains so even with climate change. This is a desirable vision, but perhaps overly optimistic in view of the actual achievements of the last 30 years. Figure 4 summarizes the simulation results, showing the additional number of people at risk of hunger in 2080 plotted against different levels of atmospheric CO_2 concentrations and associated climate changes and Figure 5 (see page 415) shows the additional number of undernourished people by selected regions, climate modes, and IPCC future development scenarios.

Global environmental change raises the issue of fairness and equity. Developing countries have thus far contributed relatively little to the causes of climate change. Yet many of the poorest countries and those in food deficit will suffer substantial losses in domestic food production, which will increase food insecurity and hunger. The climate-change-induced ecological changes will also further exacerbate the water, health, and infectious disease challenges in many of the same countries.

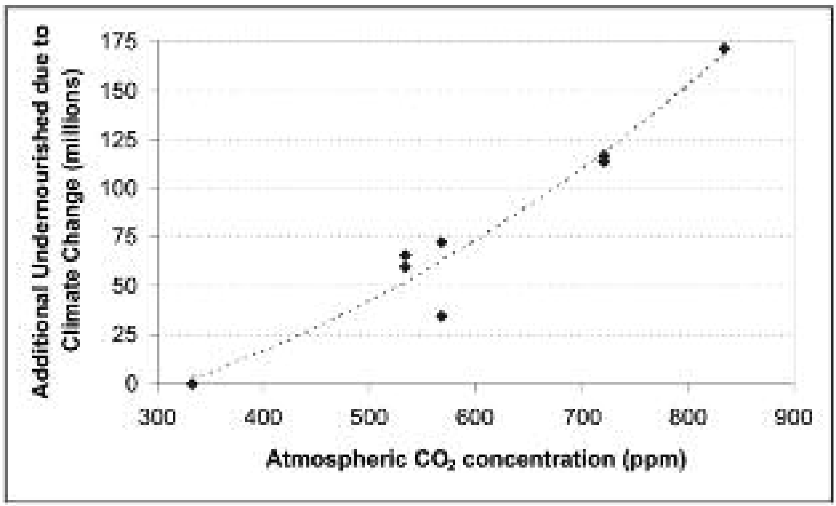


Figure 4. Increase in undernourished population due to climate change (Fisher *et al.*, 2002b).

2.3. *Food and Hunger: Examples of Spatial Extensions of CEED Methodology and Resource Database*

– Incorporation of regional and global climate change models;
– Formulation of CEED future agricultural path scenarios;
– Assessment of surface and groundwater resources and irrigation potentials;
– Assessment of bio-energy crop cultivation and production potentials;
– Assessment of natural vegetation and analysis of interaction with crop agriculture;
– Assessment of livestock production potential – pastoral and intensive;
– Assessment of sustainable marine fisheries and potentials for aquaculture;
– Analysis of national and international trade, subsidies, tariffs, and quotas;
– Assessment of nutrition, hunger, rural poverty, and livelihood options;
– Assessment of food security and food safety;
– Analysis of differential vulnerability of ecosystems and populations;
– Risk and uncertainty analysis;
– Initiation of detailed CEED national and regional food system studies.

3. Water and Health

The 1972 United Nations Conference on the Human Environment and Development in Stockholm endorsed the statement 'that all people have the right to have access to drinking water'. Some twenty years later the United Nations General Assembly declared the International Drinking Water and Sanitation Decade to achieve universal access to water supply and sanitation.

In 2002 the United Nations Committee on Economic, Social, and Cultural rights, interpreting the provision of the International Covenant on Economic, Social, and cultural rights, proclaimed, 'Water is fundamental for life and health. The human right to water is indispensable for leading a healthy life in human dignity. It is pre-requisite to the realization of all other human rights'. Some 145 countries that have ratified the covenant are now compelled to progressively ensure that everyone has access to safe and secure drinking water, equitably and without discrimination.

Only 2.5% of the world's water is fresh water. Present global water consumption accounts for 5% of total renewable water resources and this use

is projected to double in the next two decades. The worldwide inequity in water consumption – an average of 20 liters per capita per day in the developing countries compared to over 400 liters in the developed countries – is an example of the global disparity between the rich and the poor.

Globally, some 170 million people in urban areas and about 1 billion in rural areas currently lack access to safe water. On average about 80% of the global population has access to safe water; however, there are significant regional differences. For example, in Africa only 60% of the population has access to safe water. To achieve the Millennium Development Goals (UN, 2005) of halving the number of people without access to water by 2015, over 1.5 billion additional people will need to be provided access to safe drinking water. This task, which will involve increasing water access capacity by over 30% of current water infrastructure, will require substantial investments. It will also require the local availability of water resources, as the transport of water over long distances is prohibitively expensive.

Although there are sufficient water supplies at the global level, these do not coincide with regional distribution of the global population. Water conflicts within countries and across countries are increasing. Water-stressed and water-scarce countries are defined, respectively, as those with less than 1700 and 1000 cubic meters of water available per capita. Currently, more than 30 countries with a total population of over 500 million are regarded as water scarce (Meinzen-Dick and Rosengrant, 2001). By 2025, some 50 countries with a total population of about 3 billion may be in this category. The arid and semi-arid zones of the world, which constitute 40 percent of the global landmass, have only 2 percent of global run-off.

Pollution and inefficient use of water resources have resulted in dropping water tables, drying rivers, the extinction of aquatic species, and the disappearance of invaluable wetland ecosystems. In many countries water supplies are drawn from underground aquifers faster than they are recharged with rainfall. The result is shrinking ground water tables, which has serious consequences for the rural poor who often cannot afford the cost of deeper drilling. Some of most populous countries, such as China, India, Pakistan, and Mexico, have substantially depleted their groundwater resources during the last three decades. The proximate causes of groundwater depletion, as well as groundwater pollution, are rooted in population growth, economic expansion, the distorting impacts of subsidies and financial incentives, and the spread of energized pumping technologies.

The pollution and depletion of aquatic ecosystems has serious consequences for biodiversity. During the last fifty years, half the world's wet-

lands have been destroyed, and pollution in coastal areas has resulted in declining fisheries and the destruction of biologically diverse habitats. Aquatic ecosystems, both freshwater and marine, are sensitive to pollution from agriculture, human waste, and industry.

In many tropical regions, annual rainfall occurs during a short rainy season, and most of this rainfall, unless it can be stored, is lost in runoff and river flows to oceans. The large majority of the 45,000 dams in the world were built between 1960 and 1990, but this construction has slowed with increasing awareness of the environmental disruption, displacement of populations, loss of agricultural lands, silting, and impacts on downstream areas created by these dams.

About 70% of the world's freshwater goes to agriculture. It is expected that over the next two decades the world will need 17 percent more water to grow food for increasing populations in developing countries, and that total water use will increase by 40 percent. This is projected to result in more than half of the world population facing moderate to high water scarcity (FAO, 2002).

With increasing water stress on the one hand and the increasing demand for water to meet the food needs and industrial and municipal needs of a growing population on the other, the challenge of water policy lies in finding integrated water resource management solutions that take into account multiple development needs while also protecting the environment. At the levels of watersheds, basins, and nations, water productivity must be understood from multi-stakeholder perspective in the broadest sense. For example, water productivity at the basin level must be defined to include: crop, livestock, and fishery yields; wider ecosystem services and social impacts, such as health; and the systems of resource governance that ensure equitable and cost effective distribution of benefits.

Irrigated agriculture with the heavy use of chemicals is a major cause of agricultural runoff of fertilizers and pesticides. This runoff in turn contaminates rivers, lakes, and underground aquifers, polluting drinking water supplies. These chemicals, even in low concentrations, can build up over time and eventually lead to chronic diseases. Agricultural pesticides such as DDT and heptachlor often wash off in irrigation water and as a result are found in water and food products. This has implications for human health because they are known carcinogens and also may cause low sperm counts and neurological disease.

Health problems caused by nitrates in water supplies are becoming a serious problem; in over 150 countries, nitrates from fertilizers have

seeped into water wells. Excessive concentrations of nitrates cause 'blue baby syndrome', an acute and life-threatening disease among infants and young children. High levels of nitrates and phosphates in water also encourage the growth of blue-green algae, which leads to eutrophication.

It is estimated that more then a third of the global burden of disease is attributable to environmental factors. Inadequate water and sanitation are the primary causes of diseases such as diarrhea (4 billion episodes annually resulting in over 2 million deaths per year), malaria (400 million episodes with 1.5 million deaths annually), schistosomiasis (200 million episodes with 200,000 deaths and 20 million suffering severe consequences a year), intestinal helminthes (1.5 billion infections a year with 100,000 deaths annually), and trachoma (500 million people at risk and 6 million blinded every year) (WHO, 2001). Some 60% of all infant mortality is linked to infectious and parasitic diseases, most of them due to consumption of contaminated water.

Population growth combined the with impact of climate change in the 21st century will significantly affect the availability of water resources as well risks of waterborne diseases (Vörösmarty, 2000). Climate change affecting temperature, precipitation, and variability is likely to affect the geographic distribution of vector borne diseases, especially malaria, cholera, dengue, and schistomiasis. For example, higher temperatures would result in an increased prevalence of malaria in higher altitudes and latitudes. Heavier rainfall would cause the increased transport of microbial agents from soil leaching, such as cryptosporidiosis, giariasis, amsebiasis, typhoid, and promote the growth of mucilaginous blue-green algae and coastal/marine copepod zooplankton that provides hosts for cholera pathogens.

Climate variability events such as El Niño have been found to extend malaria and dengue epidemics and increase incidences of diarrhea. The frequency and intensity of extreme climate events is likely to increase as climate changes. In poor and vulnerable countries lacking early warning systems and security support, such events already result in significant losses of life, population displacement, the destruction of individual assets, physical infrastructure, and services, and outbreaks of infectious diseases in the aftermath of the event.

3.1. *Water and Health: Examples of Spatial Extensions of CEED Methodology and Resource Database*

– Compilation of geographical distribution of water bodies, waterborne vectors and diseases, and populations at risk;

– Analysis of CEED driving forces and cofactors of water borne diseases;

– Analysis of capacity of health care systems, including prevention, detection, and treatment;

– Analysis of safe water supplies, access, and affordability;

– Assessment of pollution and contamination of surface and groundwater resources;

– Identification of water stress and scarcity areas and evaluation of water policy options;

– Spatial differential vulnerability of ecosystems and populations to waterborne diseases;

– Risk and uncertainty analysis;

– Initiation of detailed CEED national and regional water and health studies.

4. Ecology and Infectious Diseases

The biological diversity of nature lies in the variety of life and its processes. It includes the vast array of organisms, the genetic differences among them, the communities and ecosystems in which they occur, and the ecological and evolutionary processes that keep them functioning yet ever changing and adapting. The organisms that constitute the world's biodiversity, a major share of which are parasitic and infectious organisms, live interdependently in complex ecological networks, where each relies for nutrients and energy on those that share its habitat (Shah, 2004).

Healthy Ecosystems support the energy needs of all species and also perform essential services such as purification of air and water, binding of toxins, decomposition of wastes, watershed and flood management, stabilization of landscapes, and regulation of parasites and pathogens.

Biological species live and evolve in environments and biotic cycles that encompass physical, chemical, and behavioral relationships that promote continued evolution. The symbiotic relationships between species are not only invaluable for the future well-being of the human race but

also ensure the sustainability of the Earth's life-supporting capacity. More than half of the world's biodiversity is found in tropical rainforests, which cover just 7% of the earth's land area. It is well documented (Stevens, 1992; Chown and Gaston, 2000) that marine and terrestrial biodiversity decline as one moves farther from the equator, a phenomenon known as the latitudinal species diversity gradient. A number of studies have analyzed the diversity gradient and possible causal factors such as area, water, energy, geographical patterns, and habitat diversity (Rhode, 1992; Hawkins, 2003). However, parasitic and infectious organisms that account for a major share of the world's biodiversity have not been considered in most previous studies (Guernier *et al*., 2004)

Infectious agents include viruses, bacteria, protozoa, and multi-cellular parasites. The microbes that cause 'anthroponoses' have adapted, via evolution, to the human species as their primary host. In contrast, non-human species are the natural reservoir for those infectious diseases agents that cause 'zoonoses'. Zoonotic pathogens are the most significant cause of emerging infectious diseases, and account for over 70% of pathogens. Vectors, pathogens, and hosts survive and reproduce within a range of climatic and ecological conditions. It is important to identify geographically how hosts, vectors, and parasites interact with each other and with the environment. The key questions in this identification relate to the potential carrying capacity, extent, and transmission of parasites and pathogens, as well as insects and other intermediate hosts. Particular ecological systems are associated, to a large extent, with particular infectious diseases; the exception is malaria, which is found in dry land areas, forests, and wetlands. Ecological change plays an important role in the emergence or resurgence of infectious diseases. Hence understanding the relationship between ecological changes and the nature of epidemic and endemic diseases and emerging pathogens is critical.

Guernier *et al*. (2004) have investigated the influence of temperature and moisture, as well as certain demographic and economic characteristics, on the latitudinal diversity and distribution of parasitic and infectious diseases. The analysis was based on compilations of epidemiological data on 332 different human pathogens across 224 countries and allowed for cofactors such as size of country, demography, economy, and environment. The results showed that a large number of parasitic and infectious disease (PID) species follow the same diversity latitudinal gradient as other biological species. This study also found significant positive correlations between pathogen species richness and the maximum range of precipitation, but lit-

tle significant relationship between pathogen species richness and monthly and annual mean temperatures.

A number of previous studies have analyzed the influence of climate on specific infectious diseases:

– Regional cholera epidemics occur seasonally and are associated with periods of excessive rainfall, warm temperatures, and increases in plankton populations (Lipp *et al.*, 2002). Monthly and annual cholera deaths have been found to be positively correlated with sea surface temperature (an ENSO correlate) and air temperature (Speelmon *et al.*, 2000);

– Malaria outbreaks often occur following periods of increased rainfall and temperature, due to positive effects on vector breeding (Kilian *et al.*, 1999), development rates, parasite sporogony, and, ultimately, entomological inoculation rates;

– Meningococcal meningitis, an airborne bacterial disease, shows a highly seasonal and epidemic pattern in sub-Saharan Africa where outbreaks occur during the hot, dry season and decline when the rainy season begins (Molesworth *et al.*, 2002);

– Dengue epidemics are characteristically associated with high rainfall as well as elevated temperatures and humidity due to direct and indirect effects on pathogen and vector biology (Gubler *et al.*, 2001);

– Yellow fever, a zoonotic viral disease, has been found to be dependent on temperature but the importance of temperature fluctuations in the inter-annual variation of disease is unclear (Reiter, 2001);

– Rhodesiense African trypanosomiasis studies have suggested a link between temperature and vegetation and the distribution of tsetse in Africa (Fischer *et al.*, 1985; Robinson *et al.*, 1997);

– Japanese encephalitis epidemics are highly seasonal, occurring during the monsoon season when temperatures reach 30°C or above (Mellor and Leake, 2000);

– Rift Valley fever outbreaks are positively associated with warm ENSO events and above-normal precipitation (Anyamba *et al.*, 2002);

– West Nile virus epidemics occur during unusually hot and dry periods (Epstein, 2001);

– Schistosomiasis is related to environmental factors such as rainfall, temperature, and water body composition (Brooker and Michael, 2000);

– Lymphatic filariasis epidemics are related to temperature and precipitation (Lindsay and Thomas, 2000);

– Chagas disease is associated with high temperatures and low humidity (Carcavallo, 1999) as well as particular types of vegetation (Dumonteil *et al.*, 2002);

– Lyme disease incidences peak during high temperature summer months (Estrada-Peña, 2002).

Many of the disease-causing parasites and pathogen populations have flourished as environmental changes impact predator species. The wealth of biodiversity and its intricate connections are increasingly threatened as growing human populations and their ever-increasing consumption ravage the environment. The damage is evident: partial loss of the ozone layer; global warming and climate change; air, water, and land pollution; land degradation and erosion; salinity and desertification; wetland destruction; disappearing forests; extinction of species; and depletion of mineral resources. The cumulative consequences of unsustainable human activities will, in the long run, threaten nature's life-supporting capacity and resources.

Over the past half century, more than a quarter of the world's 8.7 billion hectares of crop-lands, pastures, forests, and woodlands have been degraded through misuse or overuse. Agriculture is by far the largest use of land and has the greatest impact on the environment and its biodiversity. Agricultural expansion has already resulted in the loss and fragmentation of the world's forests, modification of wetlands, streams, estuaries, lakes, and coastal and marine ecosystems. Many ecosystems are also threatened by unchecked expansion of urban areas and road infrastructure (Patz *et al.*, 2004).

Evolutionary ecologists are trying to identify the drivers that regulate species diversity. Among various hypotheses, factors such as area and energy, geographic constraints, and habitat diversity have been considered. Intuitively, one can hypothesize that plant species diversity results from an interaction between temperature, precipitation, topography, and soils. In turn, the herbivore species diversity results from the diversity of vegetation, and carnivore species diversity depending on the herbivores. The diversity of domesticated livestock has evolved in tandem with the variability and range of food supplies and in the context of social and economic development around the world. Next in this chain, the evolution and spread of parasite and pathogen diversity depends on the availability of habitat and breeding environments and hosts, including wildlife, livestock, and humans.

In recent years there has been a growing emphasis on Integrated Vector Management (IVM) strategies, as they combine the most effective epidemiological and ecological chrematistics. IVM, which integrates environmental management, biological control methods, and chemical control methods, aims to 'improve the efficacy, cost-effectiveness, ecological soundness

and sustainability of disease vector control'. Furthermore IVM encourages a multi-disease control approach, integration with other disease control measures and the considered and systematic application of a range of interventions, often in combination and synergistically (WHO, 2004).

All biological species need access to dispersal areas as climatic and environmental conditions, and food and water supplies undergo seasonal changes. The movement of zoonotic and wildlife hosts combined with the existence of areas of urban population concentration further add to risks and prevalence of infectious diseases.

The Earth's latitudinal diversity of biological species including parasites and pathogens, and the relevance of and inter-relationships with cofactors such as mean and variability of climate parameters, environmental degradation and pollution, vegetation changes (crops, pastures, natural vegetation), and water resources remain to be analyzed. These factors, as well as economic and demographic development, are relevant to assessing the risk and prevalence of infectious diseases.

4.1. *Ecology and Infectious Diseases: Examples of Spatial Extensions of the CEED Methodology and Resource Data Base*

– Compilation of geographical distribution of parasites, pathogens and infectious diseases and populations at risk;

– Analysis of CEED driving forces and cofactors of infectious diseases;

– Analysis of capacity of health care systems: prevention, detection, and treatment;

– Spatial differential vulnerability of ecosystems and populations to infectious diseases;

– Risk and uncertainty analysis;

– Initiation of detailed CEED national and regional ecology and infectious diseases studies.

5. Concluding Remarks

Climate change is global, long term, and involves complex interactions among demographic, climatic, environmental, economic, health, political, institutional, social, and technological processes. It has significant international and intergenerational implications in the context of equity and sustainable development (UN, 1992). Climate change will

impact social, economic, and environmental systems, and it will shape prospects for food, water, and health security. Quantitative information for geographically specific areas provides important knowledge that can underpin sub-national, national, and regional adaptive policies to mitigate the consequences of global environmental change. It may also facilitate international negotiations on climate change and sustainable development that take into account global inequities.

REFERENCES

Anyamba A.K., *et al*. (2002), Mapping potential risk of Rift Valley fever outbreaks in African savannas using vegetation index time series data, *Photogrammetric Engineering and Remote Sensing*, 68(2): 137-145.

Brooker S., Michael E. (2000), The potential of geographic information systems and remote sensing in the epidemiology and control of human helminth infections, *Advances in Parasitology*, 47: 245-288.

Carcavallo R.U. (1999), Climatic factors related to Chagas disease transmission, *Memorias do Instituto Oswaldo Cruz*, 94: 367-369.

Chown, S.L., Gaston, K.J. (2000), Areas, cradles and museums: The latitudinal gradient in species richness, *Trends Ecol. Evol.* 15, pp. 311-315.

Dumonteil E., *et al*. (2002), Geographic distribution of Triatoma dimidiata and transmission dynamics of Trypanosoma cruzi in the Yucatan Peninsula of Mexico, *American Journal of Tropical Medicine and Hygiene*, 67(2): 176-183.

Epstein P.R. (2001), West Nile virus and the climate, *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, 78(2): 367-371.

Estrada- Peña A. (2002), Increasing habitat suitability in the United States for the tick that transmits Lyme disease: A remote sensing approach, *Environmental Health Perspectives*, 110(7): 635-640.

FAO (2002), *Crops and Drops – Making the Best Use of Land and Water*, Rome.

Fischer G., Shah M.M. and Rollinson, D.H. (1985), Agricultural Perspectives in the Tsetse Infested Areas in Africa, *IIASA/FAO Collaborative Paper*, CP-85-11, IIASA, Laxenburg.

Fischer G., Vetlthuizen H., Shah M.M., and Nachtergale F. (2002a), *Global Agro-ecological Assessment for Agriculture in the 21st century – Methodology and Results*, RR-02-02, IIASA, Laxenburg.

Fischer G., Shah M.M., and Velthuizen H. (2002b), Climate Change and Agricultural Vulnerability, *WSSD*, Johannesburg.

Gubler D.J., *et al*. (2001), Climate variability and change in the United States: Potential impacts on vector- and rodent-borne diseases, *Environmental Health Perspectives*, 109: 223-233.

Guernier, V., Hochberg, M.E., Guégan, J.F., (2004), Ecology drives the worldwide distribution of human diseases, *PLoS Biol.* 2 pp. 740-746.

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guegan, J.-F. (2003), Energy, water, and broad-scale geographic patterns of species richness, *Ecology* 84, pp. 3105-3117.

Higgins G., Kassam A.H., Naiken L., Fischer G., Shah M.M. (1983), Potential Population Supporting Capacities of the Lands in the Developing World, *Final Report of the Land Resources for Populations of the Future project*, FAO/IIASA/UNFPA, Rome.

IPCC (2001), *Climate Change 2001*, Cambridge University Press, Cambridge.

Kilian AHD, *et al*. (1999), Rainfall pattern, El Niño and malaria in Uganda, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93(1): 22-23.

Lutz, W., Global demographic trends, education and health, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 252-268, The Pontifical Academy of Sciences, Vatican City, 2006.

Lutz W. and Shah, M.M., (2002), Population should be on the Johannesburg Agenda, *Nature*, 418: 17.

Meinzen Dick R.S., Rosengrant M.W., *Overcoming water scarcity and quantity constraints*, Focus 9, Brief 1, IFPRI, Washington DC.

Mellor P.S., Leake C.J. (2000), Climatic and geographic influences on arboviral infections and vectors, *Revue Scientifique et Technique de l'Office International des Epizooties*, 19(1): 41-54.

Molesworth A.M., *et al*. (2002), Where is the meningitis belt? Defining an area at risk of epidemic meningitis in Africa, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 96(3): 242-249.

Reiter P. (2001), Climate change and mosquito-borne disease, *Environmental Health Perspectives*, 109: 141-161.

Robinson T.P., Rogers D.J., Williams B. (1997), Mapping tsetse habitat suitability in the common fly belt of southern Africa using multivariate analysis of climate and remotely sensed data, *Medical and Veterinary Entomology*, 11: 223-234.

Rohde, K. (1992), Latitudinal gradients in species diversity: The search for the primary cause, *Oikos* 65, pp. 514-527.

Shah M.M. and Strong M. (2000), *Food in the 21st Century – from Science to Sustainable Agriculture*, ISBN 0-8213-4757-8, World Bank, Washington DC.

Shah M.M. (2001), Food in the 21st Century – a Global climate of Disparities, *GAIA* 4, pp. 273-278.

Shah M.M. (2003), Population, Water and Food – Securing a Sustainable Future, Contribution to UNFPA Third Water Forum Report *Global Population and Water – Access and Sustainability*, Kyoto.

Shah M.M. (2004), Dimensions of Diversity – Humanity, Nature, and Sustainable Development, *UNESCO/International Association of Universities Conference*, Sao Paulo.

Speelmon E.C., *et al*. (2000), Cholera incidence and El Niño-related higher ambient temperature, *Journal of the American Medical Association*, 283(23): 3072-3074.

Stevens, G.C. (1992), The elevation gradients in altitudinal range: An extension of Rapoport's latitudinal rule to altitude, *Am. Nat.* 140, pp. 893-911.

U.N. (1992), *The Global Partnership for Environment and Development – A Guide to Agenda 21*, United Nations Press, New York.

U.N. (2005), Investing in Development, A Practical Plan to Achieve Millennium Development Goals, *Millennium Development Project*, New York.

Vörösmarty C.J., Green P., Salisbury, J., Lammers, R.B. (2000), Global water resources: Vulnerability from climate change and population growth, *Science* 289(5477), pp. 284-288.

WHO (2000), *WHO report on global surveillance of epidemic-prone infectious diseases*, Geneva.

WHO (2001), *Water for Health: Taking charge*, WHO World Water Day Report, Geneva.

WHO (2002), *The World Health Report, Reducing Risks, Promoting Healthy life*, Geneva.

GLOBAL DEMOGRAPHIC TRENDS, EDUCATION, AND HEALTH

WOLFGANG LUTZ

A Demographically Divided World

Current global demographic trends and the challenges associated with these trends are somewhat confusing to many observers. On the one hand, in many developing countries birth rates well above the replacement level (of two surviving children per woman) and a very young population age structure keep population growth rates very high. For these reasons, the population in a number of countries is likely to double over the coming decades. At the global level, we expect that the world population will increase from its current 6.4 billion to somewhat below 9 billion by the middle of the century. On the other hand, in an increasing number of countries the birth rate has fallen well below replacement level and the population is aging rapidly. For these countries we expect even more rapid population aging in the future, and in many cases, a shrinking of total population size. Because demographic trends differ significantly in different parts of the world, concerns about the negative consequences of rapid population growth exist simultaneous to concerns about the negative implications of rapid population aging.

The fact that this demographic divide does not always follow the traditional divide between industrialized and developing countries further complicates this picture. Some developing countries have recently seen very rapid fertility declines, and the number of 'poor' countries with sub-replacement fertility is increasing. China is the most prominent example, where fertility has fallen to an (uncertain) level between 1.4 and 1.8. Over the coming two decades, China will have significant further growth and significant population aging. The momentum created by a very young present age structure will cause the population to grow by around 200

million people, with the consequence that more women will enter reproductive age. At the same time, the one-child family policy will continue to cause serious problems in terms of the support of the rapidly increasing number of elderly. The United States will also be among the countries that simultaneously experience growth and aging, because – unlike Europe – the United States is expected to grow significantly due to high levels of immigration and higher birth rates than in Europe.

Figure 1 illustrates these trends in population growth rates for different world regions from 1950 to 2050, based on UN data estimates and projections. It shows that Europe consistently has the lowest population growth rate of all continents, falling from 1% per year in 1950 to zero growth today and an expected 0.5% per year decline by 2050. The figure also shows that all world regions have passed their peak growth rate and entered a declining trend that is expected to continue over the next half century. The only continent to depart from this general pattern is North America, which saw mostly stable population growth of around 1% from 1965 to present. That growth is expected to decline only moderately in the future. Actually, the UN expects North America to have a higher population growth than Latin America, and than the world average, by 2050.

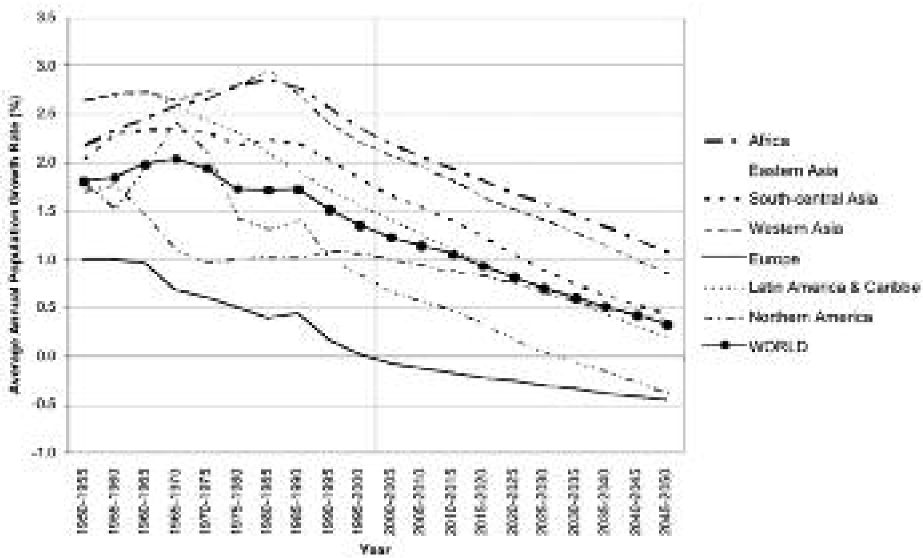


Figure 1. Average annual population growth rates of selected world regions, 1950-2050. Source: United Nations (2003) (medium variant).

Table 1 presents the trends in the two main drivers of population growth, namely mortality and fertility. As it shows, over the past half century, life expectancy has increased considerably in all parts of the world. Only in Africa over the past decade has there been a moderate decline at the continental level, as a result of HIV/AIDS, with life expectancy having declined considerably in some of the hardest hit countries. Projections assume a recovery in Africa, along with a continued increase in life expectancy in all parts of the world. Fertility rates have declined considerably around the world over the past decades. With less than 1.4 children per woman, Europe has the lowest fertility rate; the other extreme is Africa, where the average is around 4.9 children per woman. For the coming decades, the UN assumes continued declines in fertility around the world with the exception of Europe, where a recovery is assumed. Although the assumed continuation of the fertility transition in developing countries is uncontroversial, the assumption of substantial fertility increases in Europe is more disputed.

Demographic Transition as the Main Driver

Explanations and projections of fertility trends in different parts of the world generally have been guided by the paradigm of demographic transition. This paradigm assumes that after an initial decline in death rates,

	Life Expectancy at Birth (both sexes)					**Total Fertility Rate**				
Region	1950-1955	1975-1980	2000-2005	2025-2030	2045-2050	1950-1955	1975-1980	2000-2005	2025-2030	2045-2050
Africa	37.8	48.2	48.9	57.1	64.9	6.74	6.59	4.91	3.23	2.40
Eastern Asia	42.9	66.4	72.1	75.0	77.7	5.68	3.13	1.78	1.83	1.85
South-central Asia	39.4	52.6	63.2	69.1	74.0	6.08	5.09	3.25	2.18	1.91
Western Asia	45.2	60.6	69.1	75.2	78.0	6.46	5.30	3.45	2.57	2.19
Europe	65.6	71.5	74.2	78.1	80.5	2.66	1.97	1.38	1.63	1.84
Latin America & Caribbean	51.4	63.0	70.4	75.5	78.5	5.89	4.48	2.53	1.98	1.86
Northern America	68.8	73.3	77.4	79.7	81.8	3.47	1.78	2.05	1.96	1.85
WORLD	**46.5**	**59.8**	**65.4**	**70.2**	**74.3**	**5.02**	**3.90**	**2.69**	**2.25**	**2.02**

Table 1. Life expectancy at birth and total fertility rates by selected regions (1950-2050). Source: United Nations (2003) (medium variant).

birth rates also start to fall after a certain lag. In this general form, the model has received overwhelming empirical support in its ability to capture the remarkable fertility changes that happened over the 20th century.

The demographic transition began in today's more developed countries (MDCs) in the late 18th and 19th centuries, and spread to today's less developed countries (LDCs) in the last half of the 20th century (Notestein, 1945; Davis, 1954; Davis, 1991; Coale, 1973). The conventional 'theory' of demographic transition predicts that, as living standards rise and health conditions improve, mortality rates decline first, followed somewhat later by fertility rates. During the transition, population growth accelerates because the decline in death rates precedes the decline in birth rates. The demographic transition 'theory' is a generalization of the typical sequence of events that occurred in what are now MDCs, where mortality rates declined comparatively gradually, beginning in the late 1700s and then more rapidly in the late 1800s, and where, after a varying lag of up to 100 years, fertility rates declined as well. Different societies experienced this transition in different ways and various regions of the world are now following distinctive paths (Tabah, 1989). Nonetheless, the broad result has been a gradual transition from a small, slowly growing population with high mortality and high fertility to a large, slowly growing or even slowly shrinking population with low mortality and low fertility.

Figure 2 illustrates the demographic transition for two distinct world regions. It plots the crude birth and death rates (births and deaths per 1,000 of the population) for Europe and India. In the 1950s, the birth rates were almost twice as high as the death rates, which resulted in significant population growth. In Europe, the gap between birth and death rates closed during the 1990s. Looking ahead, Europe is expected to have a birth deficit, resulting in negative natural growth (with actual growth still depending on net migration). By comparison, India is in a much earlier phase of its demographic transition. Death rates have declined significantly, and with some lag, the birth rates also have started to decline, although from a much higher level than in Europe due to universal and very early marriage in India. India is expected to have completed its demographic transition by the middle of the 21st century.

This demographic transition paradigm, which has been useful for explaining global demographic trends during the 20th century and has strong predictive power for projecting future trends in countries that still have high fertility, has nothing to say about the future of fertility in Europe. The recently popular notion of a 'second demographic transition'

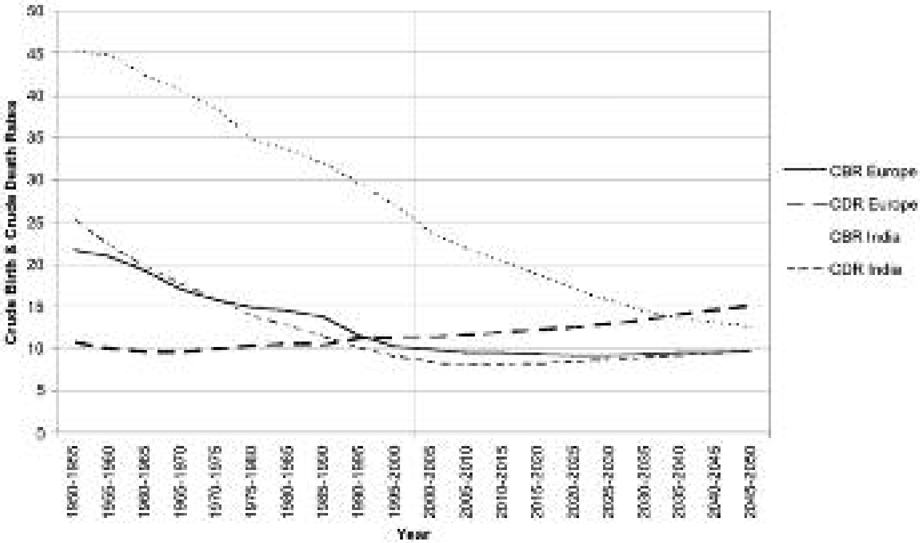


Figure 2. Crude birth and crude death rates for Europe and India, 1950-2050. Source: United Nations (2003) (medium variant).

is a useful way to describe a bundle of behavioral and normative changes that have emerged in Europe, but the concept has no predictive power. The social sciences have not yet developed a useful theory to predict the future fertility level of post-demographic transition societies. All that forecasters can do is try to define a likely range of uncertainty.

Mapping the Uncertainty Range of Demographic Trends in the 21st Century

The future trends of all three components of demographic change – fertility, mortality, and migration – are uncertain. The UN medium variant presented above is based on assumptions about what is most likely from today's perspective. But we already know that there is a high probability that the actual future trends will be either above or below the medium assumption. How should we deal with this significant uncertainty in population forecasting? This same question is asked by a recent special issue of the *International Statistical Review* (Lutz and Goldstein, 2004). The state-of-the-art report shows that the field of population fore-

casting is currently seeing a paradigm change to move from scenarios to probabilistic forecasting. Scenarios, as used in many fields of forecasting and as described by Huynen and Martens (2006, this volume), are descriptions of possible future paths without any statement of their likelihood. Particularly in cases of deep uncertainty, i.e., when there is not only parameter uncertainty but the entire model is uncertain, scenarios have become a standard tool for thinking about the future. Because in population forecasting we know the model, as described by the cohort component model of population projections, only the parameters are uncertain. For this reason, forecasting agencies around the world some decades ago followed the example of the United Nations Population Division and produced high and low variants in addition to the medium variant. This high-low range is supposed to indicate a ‘plausible range’ of future population trends. But a high-low range can only be defined in terms of one of the three components of change and as a result is mostly based on alternative fertility assumptions, while uncertainty in mortality and migration is disregarded.

To remedy such shortcomings, IIASA (Lutz *et al.*, 1997) produced the first fully probabilistic projections of the world population. These were based essentially on subjective probability distributions for future fertility, mortality, and migration, as defined by a group of experts. In 2001, IIASA performed new probabilistic projections based on a synthesis of three alternative approaches (time series analysis, ex post error analysis, and argument-based expert views) (Lutz *et al.*, 2001). The examples below are taken from this most recent forecast. Such probabilistic projections go beyond the traditional scenario analysis in several important dimensions: they are able to simultaneously consider the uncertainty in all three components of change; they can define in more precise quantitative terms what uncertainty intervals the given ranges cover; and, based on the assumption of certain correlations, they can aggregate from the regional to global level in a probabilistically consistent way. These important advantages of a probabilistic approach over a scenario approach suggest that other fields such as environmental change or future health should consider moving beyond scenarios.

The key findings of Lutz *et al.* (2001), with a high probability of above 80%, are that world population will peak over the course of this century and then start to decline. The findings also show that the 21st century will bring significant population aging in all parts of the world. In short, where the 20th century was the century of population growth, with the

world growing from 1.6 to 6.1 billion people, the 21st century will be the century of population aging, with the global proportion above age 60 increasing from currently 10% to between 24 and 44% (80% uncertainty interval). Even more significant, the proportion of the world population above age 80 will increase from currently 1% to between 4 and 20%, depending largely on the future course of life expectancy.

Figure 3 (see page 416) shows that in Western Europe the proportion of the population above age 80 might increase much more dramatically than it will at the global level. As the figure shows, currently around 3% of the population in Western Europe is above age 80, and this proportion will not change much over the coming decade. After the year 2030, however, the uncertainty range opens up very quickly. In 2050, the 95% interval already ranges from around 4% at the low end to more than 20% at the high end, with the median at around 10%. In other words, in 2050 the proportion above age 80 is likely to be three times as high as today, but it could even be six times as high. Its actual level will depend mostly on future old-age mortality – whether life expectancy will level off toward a maximum or whether it will continue to increase unabated. This difference becomes much more significant during the second half of the century. By the end of the century, the 95% interval is extremely wide, ranging from essentially the current level of 3% to an incredible 43% of the population above age 80. Even the median shows a proportion of about 20%. Societies with significant proportions of the population above age 80 will clearly be very different from today's societies, although it is likely that an average 80-year-old person during the second half of the century will be in much better physical health than an average 80-year-old person today. Clearly, the future course of old-age mortality and disability provides us with many difficult but highly important research questions.

Figure 3 also shows two blue lines for the year 2100. These give the proportions of the population above age 80 as provided by the 'high' and 'low' variants of the most recent UN long-range projections (United Nations, 1999). Because the UN does not use alternative mortality assumptions in their variants, it is not surprising that the range is quite narrow. Also, the UN projections seem to anticipate much lower improvements in life expectancy in Western Europe. This difference illustrates that the traditional variants approach, which only varies the fertility assumptions, is a highly problematic way of dealing with uncertainty and should not be interpreted as giving a 'plausible range', as is often done.

The Changing Global Distribution of Population and Human Capital

The demographic trends of the past decade, together with those projected, have and will continue to result in major changes in regional population distribution on our planet. Figure 4 plots the continents' changing shares of the world population. Asia, which holds the giant share of the world population and stays at a rather stable 55-60% over the entire 100-year period, is not included in the graph. The shares of North America, Latin America, and Oceania are also surprisingly stable over time. Big changes affect only Europe and Africa, where over the course of 100 years they fully exchange their positions. In 1950, Europe (including Russia) was home to some 550 million people constituting 22% of the world population. At present, Europe has increased to 725 million, but because the world population has increased much more rapidly, Europe's share has declined to only 12%. By 2050 Europe is expected to shrink to some 630 million, which at that point will only be 7% of the world population. Africa, on the other hand, which started at 8% in 1950, is likely to grow to around 1.8 billion by 2050, almost three times the expected population of Europe at that time.

Although these changes in relative population size are significant, it is not clear what they imply for a region's geopolitical standing. The strength and influence of a nation or a continent is not directly a function of its population size. If this were the case, then Africa today should have a similar standing in international politics, economics, or military strength to that of Europe; this is not remotely the case. What seems to count more than solely the population size of a country is its level of human capital, which can be defined in a simplified way by stratifying the people of working age by their level of education. The global distribution of human capital is changing, but the pattern looks rather different from that of population numbers alone.

The first global projections of human capital have been produced recently by IIASA (Lutz *et al.*, 2004b). Table 2 lists the persons of working age that have at least some secondary or tertiary education in 2000, and provides two alternative scenarios to 2030. The scenario 'constant' assumes that current school enrollment rates stay unchanged over time, which will result in significant human-capital improvements in many countries because of past improvements in education and the process by which the less educated older cohorts are replaced by better educated younger cohorts. The other scenario, called 'ICPD', assumes that the ambitious edu-

	Secondary and Tertiary					
	Base year		Constant		ICPD	
	2000	2000	2030	2030	2030	2030
Regions	Male	Female	Male	Female	Male	Female
North Africa	19	11	47	38	49	41
Sub-Saharan Africa	32	17	79	61	106	90
North America	88	89	100	99	100	99
Latin America	66	65	140	143	143	147
Central Asia	13	13	25	25	25	25
Middle East	17	12	50	40	53	46
South Asia	134	57	250	116	288	195
China region	238	153	416	354	406	346
Pacific Asia	53	41	99	90	106	99
Pacific OECD[a]	40	40	40	39	39	40
Western Europe	106	95	124	122	125	122
Eastern Europe	26	23	31	30	31	31
FSU Europe[b]	54	57	58	61	59	62
World	887	673	1,459	1,219	1,531	1,343

Table 2. Population (in millions) aged 20-65 by education and sex in 2000 and in 2030 according to the 'constant' and the 'ICPD' scenarios. Source: Lutz *et al*. (2004b: 149).

cation goals defined at the International Conference on Population and Development (ICPD) 1994 in Cairo will be achieved. These include a closing of the gender gap in education and universal primary education.

Table 2 shows that in terms of human capital, Europe (including Russia) is still a world power, with well over 350 million working age people who have received higher education – many more than in Africa, and even more than the huge South Asian subcontinent. This puts the pure population numbers into perspective. The table also shows that significant changes in the global distribution of human capital are to be expected, even under the constant enrollment scenario. Under this scenario, every world region will see some improvement of its overall human capital. On a relative scale, gains in today's least developed regions will be strongest, partly because the recent improvements in educating the younger generation are a significant gain in comparison to the virtual absence of education among the older cohorts. In absolute terms, even under this constant enrollment scenario, huge gains in the number of working age people with secondary or tertiary education are expected in

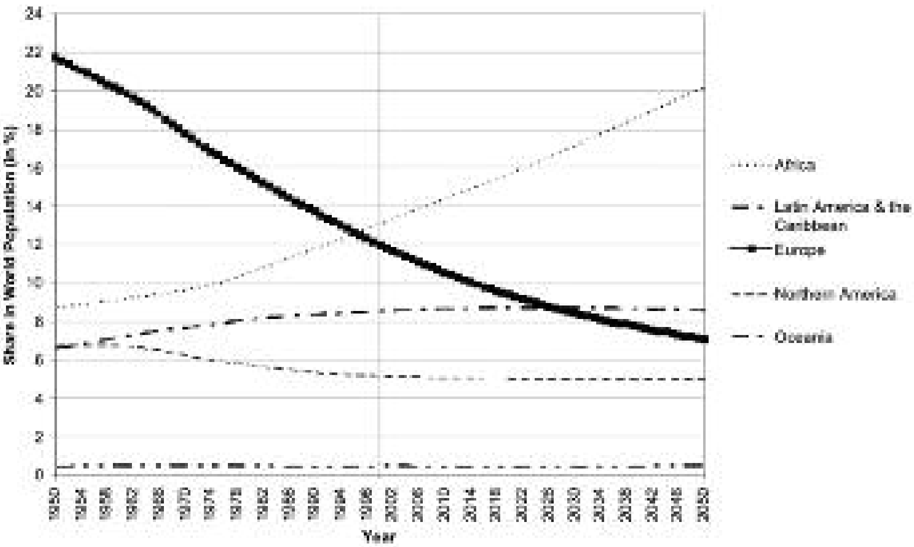


Figure 4. Share (in percent) of selected regions in world population, 1950-2050. Source: United Nations (2003) (medium variant).

Latin America, South Asia, and the China region. In today's industrialized countries, only moderate gains are expected. Comparing these results to the most optimistic scenario, which assumes the education goals of the ICPD, there is surprisingly little difference to the constant enrollment scenario. This is due to the great momentum of educational improvement. Increases in school enrollment today and over the coming decade will only very slowly affect the average educational attainment of the whole working age population. The difference between the scenarios is worth noting in Sub-Saharan Africa, where the current school enrollment rates are still far below the Cairo targets. Because the ICPD also implies lower fertility in some regions, the absolute numbers for human capital are smaller under the ICPD than under the constant rate scenario.

Figure 5 presents the information of Table 2 in graphical form. In comparing the four 'mega regions', it shows that currently Europe and North America together dominate the world in terms of human capital, although South Asia and the China region are already bigger in terms of their working age populations. The figure also shows the different pathways of China and South Asia (India). Unlike South Asia, China has

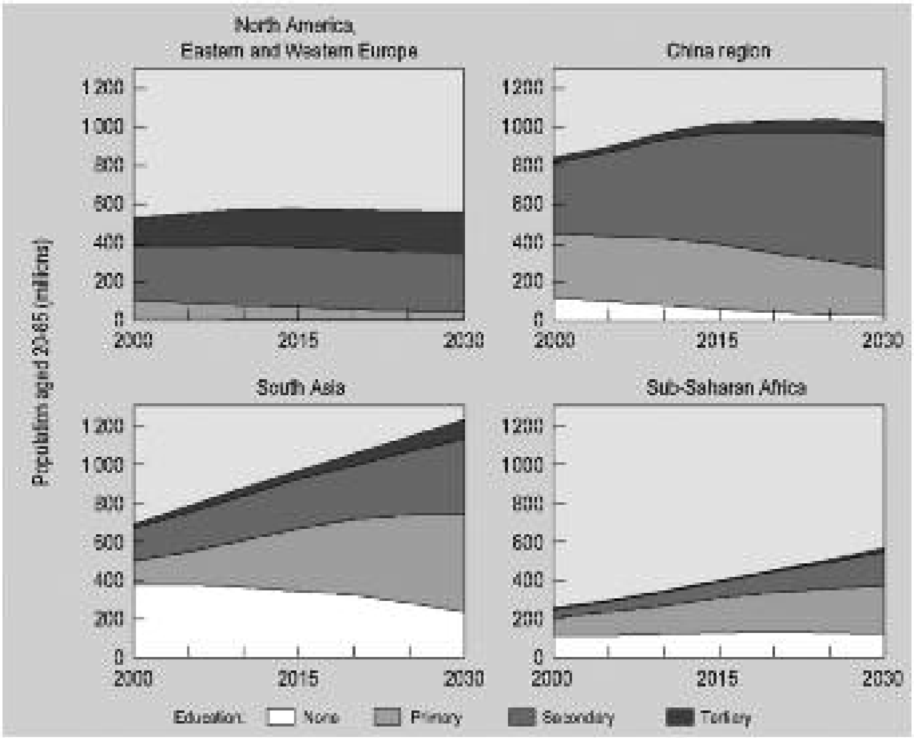


Figure 5. Population (in millions) aged 20-65 by level of education, according to the 'ICPD' scenario in four mega-regions, 2000-2030. Source: Lutz *et al*. (2004b: 138).

invested heavily over the past decades in primary and secondary education and will see a peaking of its population size over the coming decades. As a result, South Asia will soon surpass the China region in terms of population size, but will fall behind in terms of human capital. Even under the most optimistic scenario, Africa will see only very moderate increases in human capital. An interesting point is that China's human capital is increasing so rapidly that by around 2015 China will have more people of working age with secondary or tertiary education than Europe and North America together. These global shifts in human capital are likely to result in changing geopolitical and economic status and also have significant implications for global health and well-being.

Implications on Health

As outlined above, the major changes in the global population distribution – both in terms of age distribution and regional distribution – affect health in many dimensions, ranging from the implications of international mobility for infectious diseases, to the consequences of rapid population aging for the prevalence of disabilities, to the financial viability of health care systems in the face of major shifts in the ratio of contributions to entitlements. On top of these demographic changes, the outlined change in the educational composition of the global population is expected to have significant positive effects on health. At least on the individual level, education is probably the single most important determinant of health differentials. Almost universally, people with better education live longer, have better a health status, and have healthier children than less educated people of the same age in the same place of residence. The mechanisms by which education leads to better health are manifold and operate both in developing and developed countries. There are many implications of demographic and education trends on health, and we discuss only two interesting questions, one for the rapidly aging European society and one for some of the world's poorest countries.

To What Extent Will Population Aging in Europe Lead to an Increase in the Number of Disabled People?

It is evident as a feature of human life that the number of functional disabilities of all kinds increases with age. Surveys conducted in all EU member countries have collected systematic information on this progression. These surveys usually distinguish to some extent between categories of severely disabled and disabled. The data show, for instance, that on average only 2% of the women aged 25-34 are severely disabled. This proportion already reaches 10% for the age group 55-64 and further increases to almost 30% for women above age 85. The proportion of moderately disabled increases more rapidly, from 7.5% in the age group 25-34 to 36% for women above age 85.

It is equally evident and uncontested that the age structure of the European population will become significantly older over the coming decades. The proportion of the population above age 65 is expected to increase from its current 16% to possibly more than 30%. The proportion above age 80 is likely to increase even more rapidly. Combining these

observations about disability and aging, one would expect the number of disabled people in Europe to increase rapidly over the coming decades as more and more elderly people enter the ages of high disability rates. However, the calculations presented here show that this is not necessarily the case if the pattern of age-specific risk of disability continues to shift to higher ages, i.e., if at any given age the risk declines.

Figure 6 (see page 416) shows the population of the EU-15 in 2000 by age and sex and disability status as measured by the most recent health surveys. Here the red area includes both categories of the disabled, severe and moderate. The age pyramid shows a particularly large number of disabled women aged 55 to 80 at present.

Figures 7a and 7b (see page 417) present two alternative scenarios for the year 2030. Both are based on an identical projection of the total population of the EU-15, which includes the assumption of a two-year gain in total life expectancy per decade (as well as medium migration and fertility assumptions). Superimposed on this population forecast are two different assumptions about future trends in age-specific disability rates. Figure 7a presents the projected age pyramid under the assumption that currently observed age-specific proportions disabled do not change over time. This somewhat implausibly assumes that while mortality rates at higher ages decline due to improving life expectancy, the probability of becoming disabled at a given age does not change. The result of this first scenario shows that the number of disabled persons in the EU will indeed increase from currently around 60 million to around 75 million in 2030.

The second scenario shown in Figure 7b assumes that while life expectancy increases by two years per decade, the schedule of age-specific proportions disabled is also shifted to higher ages by two years per decade. This means, for example, that by the year 2030 the risk of being disabled at age 70 is equal to that the risk at age 64 in 2000. For this scenario, the results show almost no increase in the number of disabled people in Europe, with the total number increasing only from 60 to 62 million by 2030. Extended to 2050, this scenario results in slight declines in the disabled population.

In conclusion, these scenario calculations of Europe's future physically disabled population indicate that the number of elderly people in need of care and assistance will not necessarily increase as a consequence of population aging. The key factor will be the future trend in age-specific risks to become disabled, an area where preventive medicine and public health measures may be able to make a big difference.

The Effect of Improving Education on Adult and Child Mortality

In a recent paper, Lutz *et al*. (2004a) study the effects of changing educational structures on adult and child survival in Guinea, Zambia, and Nicaragua. Table 3 illustrates the considerable differences in the survival chances of children depending on the education of their mothers. In Nicaragua, the children of women without formal education die almost four times more often than the children of women with higher education. The reasons for this remarkable difference are not difficult to find. Women with a better education generally have a better social and economic standing, are better informed about healthy practices, and typically have better access to the health care system.

The same effects are likely to operate when it comes to adult mortality and total life expectancy. Table 4 shows that the difference in life expectancy among Nicaraguan men who have higher education and those who have no education is about 12 years. In Zambia, which is very hard hit by HIV/AIDS, the absolute difference is less because AIDS tends to affect all educational groups, but, even there, the more educated have a life expectancy that is 20% higher than that of the uneducated.

It is interesting to note that this important influence of education on health and survival does not disappear as countries become more developed. For example, a study using data from Austria in the 1980s shows that the mortality risk for the age group 35-64 is twice as high for men with basic education as it is for men with tertiary education. For women, the same ratio is about 1.5 (Lutz *et al.*, 1999). Translated into life expectancy, this educational difference is 6-7 years, which is comparable to the differential between the sexes. In the countries of Eastern Europe, the educational mortality and health differential is even bigger.

	Guinea	Zambia	Nicaragua
No education	204	198	72
Primary	162	177	43
Secondary	104	124	26
Higher			19
Total	195	NA	45

Table 3. Under-five mortality rates (5q0) per 1,000 live births by mother's level of education in Guinea, Zambia, and Nicaragua. Sources: DHS Guinea 1999; DHS Nicaragua 2001; DHS Zambia 2002.

	Guinea	Zambia	Nicaragua
Males			
No education	46.7	30.7	62.7
Primary	51.5	32.7	68.2
Secondary	55.5	36.0	72.8
Higher	55.5	36.0	74.5
Total	48.8	32.7	67.2
Females			
No education	48.3	30.6	67.5
Primary	53.3	32.3	72.7
Secondary	59.0	36.0	76.7
Higher	59.0	36.0	78.3
Total	49.5	32.1	71.9

Table 4. Estimated life expectancy at birth by sex, education, and country in 2000-2005. Sources: DHS Guinea 1999; DHS Nicaragua 2001; DHS Zambia 2002.

Conclusions

Over the course of the 21st century the world will likely see large demographic changes quite different from those observed during the 20th century. The population will continue to grow substantially in some of the poorest countries, while it will age massively in the richer countries. For all countries, these trends will pose major new health threats. The 21st century will also see an increasingly urbanized population. It is estimated that more than half of the world population will live in urban areas within a few years. This will bring new health problems, such as those associated with a sedentary lifestyle, air pollution, and possibly higher transmission rates of infectious diseases. These well-established demographic changes have not yet been studied sufficiently well with respect to their consequences for human health and well-being.

This paper adds a potentially very important new aspect to the analysis – namely, global change in the composition of the population by education. Given the very strong link between education and health, the fact that we are likely to see significant improvements in the educational composition of populations around the world is good news. In addition, a better-educated population can be assumed to be less vulnerable to new health threats

resulting from environmental change or other global influences, and to show a greater adaptive capacity. Unfortunately, some countries are clearly falling behind in the improvement of their human capital.

REFERENCES

Coale, A.J. (1973), The demographic transition. In: *Proceedings of the International Population Conference*, Vol. 1, Liege, Belgium: International Union for the Scientific Study of Population.

Davis, K. (1991), Population and resources: fact and interpretation, pp. 1-21. In: K. Davis and M.S. Bernstam (eds.), *Resources, Environment and Population: Present Knowledge*, Oxford: Oxford University Press.

Davis, K. (1954), The world demographic transition, *Annals of the American Academy of Political and Social Science*, 237: 1-11.

Huynen, M., and P. Martens, Globalisation and human health: towards scenarios for the 21st century. In: *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 269-289, The Pontifical Academy of Sciences, Vatican City, 2006.

Lutz, W. and J. Goldstein (Guest Eds.) (2004), How to deal with uncertainty in population forecasting?, *International Statistical Review 72* (1 & 2): 1-106, 157-208.

Lutz, W., A. Goujon, and A. Wils. 2004a, *Forecasting Human Capital: Using Demographic Multi-State Projection Methods to Show the Long-Term Effects of Near-Term Investments in Education*, World Bank Working Paper, Washington, D.C.: The World Bank (forthcoming).

Lutz, W., W.C. Sanderson, and S. Scherbov (Eds.) (2004b), *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation and Sustainable Development*, London: Earthscan.

Lutz, W., W. Sanderson, and S. Scherbov (2001), The end of world population growth, *Nature*, 412: 543-545.

Lutz, W., A. Goujon, and G. Doblhammer-Reiter (1999), Demographic dimensions in forecasting: Adding education to age and sex, pp. 42-58. In: W. Lutz, J.W. Vaupel, and D.A. Ahlburg (eds.), *Frontiers of Population Forecasting*. A Supplement to Vol. 24, 1998, Population and Development Review, New York: The Population Council.

Lutz, W., W. Sanderson, and S. Scherbov (1997), Doubling of world population unlikely, *Nature*, 387: 803-805.

Notestein, F.W. (1945), Population: The long view, pp. 36-57. In: T.W. Schultz (ed.), *Food for the World*, Chicago: University of Chicago Press.

Tabah, L. (1989), From one demographic transition to another, *Population Bulletin of the United Nations*, 28: 1-24.

United Nations (2003), *World Population Prospects: The 2002 Revision*, New York: United Nations.

United Nations (1999), *Long-Range World Population Projections: Based on the 1998 Revision*, New York, United Nations, ESA/P/WP.153.

GLOBALIZATION AND HUMAN HEALTH: TOWARD SCENARIOS FOR THE 21st CENTURY

MAUD HUYNEN & PIM MARTENS

1. Introduction

Good health for all populations has become an accepted international goal. Looking at past and contemporary developments in human health, we can state that there have been broad gains in life expectancy over the past century. However, health inequalities between rich and poor persist, and the future prospects for health depend increasingly on the relatively new processes of global change and globalization. In the past, globalization was often seen as a more or less economic process characterized by trade liberalization and capital mobility. Nowadays, globalization is increasingly perceived as a more comprehensive phenomenon that is rapidly reshaping society.

Due to the processes of globalization, the geographical scale of important health issues is progressively increasing. This was, for example, clearly demonstrated by the rapid spread of the Severe Acute Respiratory Syndrome (SARS) epidemic in 2003. Additionally, the intergenerational equity implied by sustainable development also forces us to consider the right of future generations to a healthy life.

The pathways from globalization to health are often complex and mediated by a multitude of factors, such as economic development, lifestyle and ecological changes. Therefore, exploring the health effects of globalization requires a more holistic approach than has previously been taken. This paper describes a first attempt to add a health dimension to existing global scenarios in order to explore the future health effects of globalization.

2. Population Health: A Conceptual Framework

As the world around us becomes more interconnected and complex, human health is increasingly perceived as the integrated outcome of its ecological, socio-cultural, economic and institutional determinants. Therefore, it can be seen as an important high-level integrating index that reflects the state – and, in the long term, the sustainability – of the natural and socio-economic environment [1, 2]. The conceptual framework of population health described in this section is based primarily on a comprehensive analysis of a diverse selection of existing models of population health (see for more details [3]). Although the selected models vary in complexity, purpose and content, their strengths and weaknesses reveal the following guidelines for an ideal-type model of population health, which:

– makes a distinction between determinants of different natures;

– makes a distinction between determinants of different levels of causality;

– is as comprehensive as possible without becoming too complex; and

– includes response variables/determinants.

The nature and level of causality of the determinants can be combined into a basic framework that conceptualizes the complex multi-causal aspects of population health. In referring to the nature of the determinants, we make the traditional distinction between institutional, economic, socio-cultural, and environmental factors. These factors have different positions in the causal chain, and so operate at different hierarchical levels of causality. The chain of events leading to a certain health outcome includes both proximal and distal causes: proximal factors act directly to cause disease or promote health, and distal determinants affect health via (a number of) intermediary causes [4]. We also distinguish contextual factors. These are the macro-level conditions that form the context in which the distal and proximal factors operate and develop. Determinants with different positions in the causal chain probably also differ in their temporal dimensions. Individual-level proximal health risks can be altered relatively quickly, for example by a change in personal behavior; for disease rates in whole populations to change requires slower and more sedimentary changes in contextual factors, often over the course of a few decades.

Further analysis of the selected health models and an intensive literature study resulted in a wide-ranging overview of the health determinants that can be fitted within this framework (Table 1). Figure 1 (see page 418),

TABLE 1. Determinants of population health [3, 5].

Level/Nature	General determinants	More detailed determinants
Contextual		
Institutional	Institutional infrastructure	Governance structure
		Political environment
		System of law
		Regulation
Economic	Economic infrastructure	Occupational structure
		Tax system
		Markets
Social-cultural	Culture	Religion
		Ideology
		Customs
	Population	Population size
		Structure
		Geographical distribution
	Social infrastructure	Social organization
		Knowledge development (incl. technology)
		Social security
		Insurance system
		Mobility and communication
Environmental	Ecological settings	Ecosystems
		Climate
Distal		
Institutional	Health policy	Effective public health policy
		Sufficient public health budget
	Health-related policies	Effective food policy
		Effective water policy
		Effective social policy
		Effective environmental policy
Economic	Economic development	Income/wealth
		Economic equity
	Trade	Trade in goods and services
		Marketing
Social-cultural	Knowledge	Education and literacy
		Health education
		Technology
	Social interactions	Social equity
		Conflicts
		Travel and migration
Environmental	Ecosystem goods and services	Habitat
		Information
		Production
		Regulation
Proximal		
Institutional	Health services	Provision of and access to health care services
Economic	-	-
Social-cultural	Lifestyle	Healthy food consumption patterns
		Alcohol and tobacco use
		Drug abuse
		Unsafe sexual behavior
		Physical activity
		Lifestyle related endogen factors (blood pressure, obesity, cholesterol levels)
		Stress coping
		Child care
	Social environment	Social support and informal care
		Intended injuries and abuse/violence
Environmental	Food and water	Adequate water quality and quantity
		Adequate food quality and quantity
		Sanitation
	Physical living environment	Quality of the living environment: biotic factors (e.g., infectious disease pathogens), physical factors (e.g., temperature, radiation) and chemical factors (e.g., pollution).
		Unintended injuries (e.g., disasters, traffic accidents, work-related accidents)

which draws on this analysis, shows a manageable number of general determinants and includes important response variables such as health policies and health-related policies.

We must keep in mind, however, that determinants within and between different domains and levels interact in complex and dynamic ways to 'produce' health at a population level. In addition, the pathways between these determinants and population health are not unidirectional; for example, ill health can have a negative impact on economic development.

3. The Globalization Process as a Determinant of Population Health

More and more scholars agree that globalization is an extremely complex phenomenon. Rennen and Martens [6] define contemporary globalization as an intensification of cross-national cultural, economic, political, social and technological interactions that lead to the establishment of transnational structures and the global integration of cultural, economic, environmental, political and social processes on global, supranational, national, regional and local levels. This definition aligns with the view of globalization as deterritorialization, and it explicitly acknowledges the multiple dimensions involved in the process.

To focus our study, we identify global governance structures, global markets, global communication, global mobility, cross-cultural interaction and global environmental changes as important features of the globalization process (Table 2). These features all operate at the contextual level of health determination and influence distal factors such as health (related) policy, economic development, trade, social interactions, knowledge, and ecosystem goods and services. In turn, these changes in distal factors have the potential to affect proximal health determinants and, consequently, health [3, 5].

4. Health in Existing Global Scenarios

The value of scenario studies to explore possible future events and provide sound policy-relevant guidance for decision-makers is increasingly and widely recognized (Box 1). Globalization is often included as an important driver in existing global scenarios and sometimes even as a distinguishing factor between different storylines (see e.g., the Special Report on Emission

TABLE 2. Features of globalization [3, 5].

Features of globalization	
Global governance structures	Globalization influences the interdependence among nations as well as nation-states' sovereignty, leading to (a need for new) global governance structures.
Global markets	Globalization is characterized by worldwide changes in economic infrastructures and the emergence of global markets and a global trading system.
Global communication	Globalization makes the sharing of information and the exchange of experiences with common problems possible.
Cross-cultural interaction	Globalizing cultural flows result in interactions between global and local cultural elements.
Global mobility	Global mobility is characterized by a major increase in the extensity, intensity and velocity of movement and by a wide variety in 'types' of mobility.
Global environmental changes	Global environmental threats to ecosystems include global climate change, loss of biodiversity, global ozone depletion and a global deterioration of natural areas.

Scenarios [7]). However, a set of integrated global scenarios on future health has not been generated to date.

With the following criteria in mind – integration, long-range outlook and global scope – we considered eight scenario studies (with a total of 31 scenarios) developed since 1995: Global Environmental Outlook 3 (GEO-3) [8], the Global Scenario Group [9], Global Trends 2015 [10], the Millennium Project [11], Which World [12], the Special Report on Emission Scenarios (SRES) [7], the World Business Council on Sustainable Development Global Scenarios [13], and the World Water Scenarios (WWS) [14] (for more details see [15, 16]). Only fourteen out of the 31 selected scenarios give a reasonable description of future developments in health. Eight scenarios completely neglect the health dimension. Only four scenarios explicitly discuss several socio-cultural, economic and ecological developments as determinants of health [15, 16].

A mere 15% of the selected scenarios describe health adequately and in an integrated way, which indicates that health is not consistently handled within current global scenarios. However, other developments that possibly affect future health (e.g., food, water, environment, social change, equity, economic growth, technology) are well addressed in most scenarios [15, 16]. Therefore, it would have been possible to describe future developments in health as an outcome of these multiple drivers and pressures.

Box 1. Scenarios

Scenarios are descriptions of journeys to possible futures that reflect different assumptions about how current trends will unfold, how critical uncertainties will play out and what new factors will come into play [8]. They describe hypothetical future pathways that consist of states, events, actions and consequences that are causally linked.

Scenarios were first used primarily as planning and forecasting tools, displaying a mechanistic and deterministic view of the world. Today, it is generally accepted that scenarios do not predict but paint pictures of possible futures by exploring different outcomes that might result from changing basic assumptions [8]. The relevant question that scenarios can address is not whether any particular development will happen in the future, but rather what might happen and how we act to encourage, discourage, prepare for, and/or respond to such an event or development. In this way scenarios can go beyond the conventional paradigm and may result in surprising and innovative insights.

5. Linking Scenarios to Future Health

So what 'health future' lies ahead? We explore this question by looking at two recently developed sets of scenarios: the SRES-scenarios [7] and the GEO3-scenarios [8]. The socio-cultural, institutional, economic and environmental developments described in these scenarios are linked to three potential health futures (Box 2): the 'Age of emerging infectious diseases', the 'Age of medical technology' and the 'Age of sustained health'. Although these futures are hypothetical, they are based on views in current literature and possible 'early signs' observed within society. They also build upon past and current developments described by the health transition (Figure 2, see page 418). The projected picture of future health in a particular scenario evolves from (our interpretation of) a combination of the described developments in relevant health determinants. We describe the results of this exercise in the next sections.

BOX 2. THREE POTENTIAL FUTURE STAGES IN THE HEALTH TRANSITION

Past changes in population health encompass several related developments such as the increase in life expectancy, declining mortality and fertility, shifting causes of death, the changing character of morbidity and continuing developments in the provision of health services. These long-term changes in the patterns of health and disease and their multiple determinants can be described and explained within a conceptual framework known as the health transition (sometimes also referred to as the epidemiologic transition [17, 18]). Although it has limitations [19], the health transition is a useful tool for understanding current health trends and exploring future developments.

Although the future of human health cannot be predicted with certainty, there are patterns of change and signs that can be anticipated. Below, three possible, but hypothetical, health futures are sketched; these build on past and current transition 'stages'.[1] These health futures are based on views in current literature and possible 'early signs' observed within society. They could follow from stages in the health transition we have seen in the past and are facing at present (Figure 2). There is also the possibility that economic, political, social, or environmental crises will cause the process of transition to stagnate, or to go into reverse. Additionally, these 'futures' are not sharply delineated but reflect a continuum of possible outcomes.

Age of Emerging Infectious Diseases

Current outbreaks of SARS and other (re)emerging diseases are a reminder that sudden disease emergence is a permanent part of the world and should be anticipated [20]. It is recognized that communicable diseases are possible threats to the future of mankind [18]. According to Olshansky *et al.* [21], for example, the next stage in the health transition could possibly be characterized by emerging-disease outbreaks.

In this picture of future health [22], the emergence of new infectious diseases or the re-emergence of 'old' ones will have a significant impact on health. A number of factors will influence this development: travel and trade, microbiological resistance, human behavior, breakdowns in health systems, and increased pressure on the environment [23]. Social, political and eco- ↘

[1] This first stage of the health transition (the 'Age of pestilence and famine') is characterised by the kind of mortality that has prevailed throughout most of human history. Most developing countries are now in the second stage: the 'Age of receding pandemics'. It involves a reduction in the prevalence of infectious diseases, and a fall in mortality rates. In the third stage (the 'Age of chronic diseases'), the elimination of infectious diseases makes way for chronic diseases among the elderly. Currently, most developed countries are in this stage. Adopted from Omran (1983, 1998).

nomic factors that cause the movement of people will increase contact between people and microbes; environmental changes caused by human activity (e.g., dam and road building, deforestation, irrigation, and climate change) will contribute to the further spread of disease. The overuse of antibiotics and insecticides, combined with inadequate or deteriorating public health infrastructures will hamper or delay responses to increasing disease threats. Control of infectious diseases will be hampered by political and financial obstacles, and by an inability to use existing technologies. As a result, infectious diseases will increase drastically, and life expectancy will fall (as is currently the case in many developing countries due to the AIDS pandemic). Ill health will lead to lower levels of economic activity, and countries will be caught in a downward spiral of environmental degradation, depressed incomes and ill health.

Age of Medical Technology

Past shifts in health patterns and risk factors have been driven mainly by economic development (and associated modernization processes) and improvements in (medical) technology and health care [17, 18]. Vice versa, shifts in health and disease patterns have influenced the organized response to the changing needs of the global population, particularly in the provision of health services [24]. In the developed world, for example, the emergence of chronic health problems and unhealthy lifestyles changed the focus of health systems dramatically. In the developing world, policies concentrate on the widespread implementation of modern health care and development programs.

A continuation of these trends could possibly be described as the 'Age of medical technology' [22]. Such a future is in line with Omran's futuristic stage called 'aspired quality of life with paradoxical longevity and persistent inequities' [18]. There will be continued achievements in disease control, health promotion, and prolongation of life. To a large extent, increased health risks caused by unhealthy lifestyles and environmental changes, among other things, will be offset by increased economic growth and technological improvements. Still, some health problems will, at least for a while, challenge existing diagnostic and therapeutic abilities (just as with the evolution of HIV/AIDS). Additionally, longevity is a mixed-blessing, as it is accompanied by increasing chronic morbidity and mounting medical costs. There will also be continued socio-economic inequities.

Without long-term, sustainable economic development, increased environmental pressure and social imbalance may eventually propel poor societies into the 'Age of emerging infectious diseases'. On the other hand, if environmental and social resources eventually are balanced with economic growth, then sustained health may be achieved.

↘

Age of Sustained Health

The Earth Charter Initiative [25] is a good example of a present-day movement promoting a global ethic for sustainability. It is based on the participation of thousands of organizations, groups and individuals worldwide. The Earth Charter envisions a future characterized by a societal transformation toward sustainability, which the document itself calls 'a change of mind and heart'.

From a health perspective, such a future can possibly be described as an 'Age of sustained health' [22]. Economic growth will stay within social and ecological limits. In order to enhance physical, mental, spiritual and social well-being, policies will focus on the wide-range of health determinants, social participation, social justice, and the sustainable use of the environment. Investments in social services will lead to a sharp reduction in lifestyle related diseases, and most environment-related infectious diseases will be eradicated. Health policies will be designed to improve the health status of a population in such a way that the health of future generations is not compromised, for example, by the depletion of resources needed by future generations. Although there is only a minimal chance that infectious diseases will emerge, improved worldwide surveillance and monitoring systems will properly manage any outbreak. Health systems will be well adjusted to the ageing world population. Furthermore, disparities in health between rich and poor countries will eventually disappear. This picture of future health is in line with Omran's vision of future health described as 'quality of life, equity, development, and social justice for all' [18], which takes a holistic view of health in the context of human well-being and human rights.

6. Linking the SRES-Scenarios to Health

The most recent scenario efforts of the Intergovernmental Panel on Climate Change (IPCC) resulted in four scenarios that extend through 2100 [7]: A1, A2, B1, B2. The distinctions among these scenarios are broadly structured by defining them ex ante along two dimensions. The first dimension relates to the extent of cross-regional economic convergence and social and cultural interactions; the second relates to the balance between economic objectives and environmental and equity objectives. See also Table 3.

TABLE 3. The SRES-scenarios [7].

Scenario study	Scenarios	Brief description of storyline
Special Report on Emission Scenarios (SRES)	A1	Rapid market-driven growth, with convergence in incomes and culture.
	A2	Self-reliance and preservation of local identities; fragmented economic and technological development.
	B1	Convergent world with rapid changes in economic structures and emphasize on global solutions to sustainability.
	B2	Local solutions to economic, social, and environmental sustainability.

6.1. *Future Health in A1*

In scenario A1, economic growth, technological developments and globalization play a central role. This scenario describes decreasing mortality and increasing life expectancy due to economic growth. Global population will peak in mid-century and decline thereafter due to a rapid worldwide demographic transition. Societies will emphasize the health needs of an ageing population. Although economic development will contribute to improvements in social conditions, the focus on economic growth may lead to the 'social exclusion' of some communities. Relative income disparities will decrease, but absolute differences will remain large. Additionally, income growth will put pressure on (global) resources, leading to ecological degradation. Market-based and technological approaches will be the common response to environmental problems.

From a health perspective, this scenario might see a divergence between the developed world and parts of the developing world. In developed countries, increasing wealth, technology, and improvements in healthcare will offset most of the emerging health risks. At least in the short-to-medium term, material advances, allied with improving social conditions, will lead to gains in overall population health. As a result, the richest populations may experience particularly pronounced health improvements as they advance to the 'Age of medical technology'. Although the poorer countries will experience economic growth and subsequent health improvements, leading to increased life expectancy and increasing prevalence of chronic diseases, absolute income differences will remain. Poorer countries will not advance to the 'Age of medical technology' because they will not have sufficient means to finance wide-scale use of newly developed technologies (despite the diffusion of technological knowledge). As a result they will not be able to achieve the same level of health care as the developed countries, and they will experience more difficulties averting the neg-

ative health impacts of increasing environmental pressures (e.g., water scarcity). Consequently, there may also be resurgence of old diseases and an increase in new infections. The developing world will likely experience an increase in both chronic and infectious diseases ('Age of chronic disease'/'Age of emerging infectious diseases').

6.2. *Future Health in A2*

In scenario A2, health to a large extent is left to individual choice and not public policy. In comparison to scenario A1, economic development is moderate. The greatest economic growth will take place in the developed regions, and technological advances will benefit only rich countries due to limited diffusion of knowledge. Developed countries will increasingly invest in better welfare. Globally, however, the gains in health brought about by economic development and technology will be partly offset by environmental problems and the exacerbation of the income gap between and within countries. Although most developed countries will be able to partially counteract the threat of emerging infectious diseases by increasing investment in public health and medical care (slowly advancing toward the 'Age of medical technology'), the proportion of the total burden of disease that is due to infectious disease will increase. This will be the result of population growth and ecological degradation combined with only moderate economic growth and 'leaner' governments. The situation will be fragile, and in some developed countries the risk of infectious disease may rise considerably, creating the potential for these countries to fall back into the 'Age of emerging infectious diseases'.

In developing countries, levels of health and welfare spending will either remain the same or decline. In poor countries, current barriers to the control of major diseases such as malaria will likely persist, and the importance of adequate water and food supplies will increase, as population growth remains high and environmental degradation increases. This combination of limited economic resources, high population growth, and increasing pressure on the local and global environments will increase the prevalence of infectious diseases, leading to the 'Age of emerging infectious diseases'.

6.3. *Future Health in B1*

A central element of scenario B1 is a high level of environmental and social consciousness, combined with a global approach to sustainable development. In the developed world, mortality will decline and life expectancy

will increase as a result of improved social infrastructure and institutions, economic growth, dematerialization, and investments that decrease pressure on ecological systems via the sustained management of resources. An extensive welfare net will prevent poverty-based social exclusion. Although the average age of the population will increase due to the rapid worldwide demographic transition, healthcare systems will probably be well adjusted to an older population. Under this scenario, developed countries may well complete the transition toward the 'Age of sustained health'.

Thanks to transfers of knowledge and technology, declining national debts, low population growth, increasing education levels, and decreasing social and environmental pressures, the developing world will pass through the 'Age of receding pandemics'. Although some countries will arrive at the 'Age of chronic diseases' (i.e., the stage at which the developed world finds itself today), the global approach toward sustainability will enable most of them to skip this stage and move toward the 'Age of sustained health'.

6.4. *Future Health in B2*

The scenario B2 is characterized by an increasing concern for environmental and social sustainability in a heterogeneous world. Governments primarily concentrate on community- and policy-based solutions to environmental and health problems. Most governments will increase public spending, including public health spending. Environmentally aware citizens will exercise a growing influence on national and local policy. There will be a shift to regional and local decision-making, with a high priority given to human welfare, equality and environmental protection. Education and welfare programs will be widely pursued, reducing mortality and fertility. Nonetheless, in this differentiated world, social and environmental progress will be relatively slow and will vary across regions and countries. Increased expenditure on 'health' and 'environment' will be implemented first in richer countries, and it will take time for developing countries to follow.

In this scenario, developing countries may experience an increase in life expectancy and chronic diseases (moving slowly to the 'Age of chronic diseases') and some may eventually also achieve technological progress on their own. However, due to the slow pace of change, the developing world will not make any significant progress toward true sustainable societies within the given timeframe. For developed countries, the situation will be more robust than in A2; they will slowly start to advance toward

the 'Age of sustained health', possibly via the 'Age of medical technology'. But the transition toward sustainability will be far from complete, and whether developed countries will ever be able to achieve sustained health beyond the timeframe of the scenario will probably depend on further health developments in current developing countries. It is important to note that this scenario incorporates a lack of global governance, which might cause difficulties in solving global problems. If, for example, severe global environmental changes were to occur, the improvements in health might be adversely affected, or even be reversed.

7. Linking the GEO3-Scenarios to Health

The GEO-3 scenario exercise developed four archetype views of the future up to the year 2032 [8]: Markets First (MF), Policy First (PF), Security First (SeF) and Sustainability First (SuF). These scenarios describe possible futures based on anticipated developments in demography, economic development, human development, science and technology, governance, culture, and environment. See also Table 4.

7.1. *Future Health in Markets First*

Markets First describes the continuation of economic growth and globalization. Environmental and social issues are valued as important, but do not have the highest priority, and governments primarily rely on market-based and technological solutions. In this scenario, however, technological

Table 4. THE GEO3-scenarios [8].

Scenario study	Scenarios	Brief description of storyline
Global Environmental Outlook-3 (GEO3)	Markets First (MF)	Market-driven developments converge on the values and expectations that prevail in industrial countries.
	Policy First (PF)	Strong actions are undertaken by governments in an attempt to reach specific goals.
	Security First (SeF)	A world of great disparities, where inequality and conflict prevail, brought about by socio-economic and environmental stresses.
	Sustainability First (SuF)	A new development paradigm emerges in response to the challenge of sustainability, and is supported by new, more equitable values and institutions.

innovation will not be able to keep pace with economic development and population growth. As a result, increases in social problems and environmental degradation will continue. Human health will be negatively affected by ongoing population growth (especially in the developing world), high migration pressures, regional conflicts (e.g., in Africa), the ongoing AIDS pandemic, pressures on food and water, losses in biodiversity, pollution and climate change. There will be, however, improvements in medical technology and health care. Although the developing world will participate in the global market, inequity and poverty will persist.

Growing environmental and social health pressures combined with serious economic troubles will cause developing countries to have a difficult time reaching the 'Age of chronic diseases'. They will slowly be overwhelmed by the accumulation of social, environmental and economic problems and gradually shift into the 'Age of emerging infectious diseases'. The developed countries, on the other hand, will continue as they are now; using economic and technological means to avert negative health impacts. They will advance to the 'Age of medical technology'. However, as pressures on health continue to increase and the migration from South to North facilitates the spread of infectious diseases, developed countries will have to keep in mind that there is a considerable risk of falling into the 'Age of emerging infectious diseases'.

7.2. *Future Health in Policy First*

In the Policy First scenario, sustainable development becomes the cornerstone of political agendas. This future is, however, also characterized by slow progress and mixed results of policy measures. There will be advances in education, reduction of extreme poverty, improvement in environmental quality and slowed population growth. However, progress in food and water availability will not be able to keep pace with the increasing demand, especially in developing countries. Other problems will remain or possibly increase: inequity (although efforts will be made to lower foreign debts and stimulate development in developing countries), regional conflicts, and climate change. In this scenario, there will be some progress toward sustainability, but a lot of work will still have to be done. The scenario itself describes improvements in infant and child mortality.

In this future, the developed countries will be in the process of shifting toward the 'Age of sustained health', but within the timeframe of the

scenario they will have not reached the completion of this transition to a considerable degree. Whether they will ever achieve sustained health will strongly depend on the health developments in the current developing world. The developing countries will not benefit as much from the described improvements as will the developed world, due to persistent inequity. They will probably advance to the 'Age of chronic diseases', but it is unclear whether they will ever be able to progress toward the 'Age of sustained health'.

7.3. *Future Health in Security First*

The main characteristic of the Security First scenario is the enormous increase in the income gap. In this 'future of inequity', the poor will inevitably be the first victims of the adverse effects of the numerous and growing pressures on population health. These pressures will include increasing resource problems (e.g., food and water scarcity), environmental degradation, (political) conflicts and tensions, migration, population growth, lack of education, inadequate healthcare, the continuing AIDS pandemic and climate change. The scenario also describes the resurgence of old diseases and the emergence of new diseases, relatively slow technological progress that only benefits the rich, low priority for social problems, and stagnant economies.

Society will find itself in a downward spiral and the poorest countries will not be able to advance to 'Age of chronic diseases'. Social, environmental and economic pressures will lead them to the 'Age of emerging infectious diseases' very rapidly. The rich will be able to avert negative health impacts, at least in the short-to-medium term. They will live separately from the poor in (metaphorical) fortresses, where they will be (temporarily) protected against environmental and social problems and where they will have access to proper health care and medical technology. At first, the developed world will be able to continue in the 'Age of chronic diseases' or even advance to the 'Age of medical technology'. But because the situation for the rich is less robust than in the Markets First scenario, the proportion of the total burden of disease comprising communicable diseases will grow. It is only a question of how long it will take for the rich countries to eventually shift completely into the 'Age of emerging infectious diseases' as the social, environmental and economic pressures from the outside increase.

7.4. *Future Health in Sustainability First*

In the Sustainability First scenario, people embrace a new sustainability paradigm. Social issues (including health) and environmental quality have high priority, policy measures have strong results, and gradual economic growth occurs within the limits of sustainable development. This scenario describes a successful transition toward sustainability that results in great reductions in the pressures on population health, stabilization of population at moderate levels, increasing education levels, reductions in conflicts and tensions, increasing environmental quality, sufficient water, and sufficient food. It is also characterized by a closing gap between rich and poor, and deliberate efforts to reduce child mortality and to increase life expectancy.

In the future described by the Sustainability First scenario, conditions will become favorable for both the developed countries and the developing world to reach the 'Age of sustained health'. It is even possible that the current developing countries will skip the 'Age of chronic diseases' and advance directly to the 'Age of sustained health'.

8. Future Health in a Globalizing World

The futures presented by the SRES, and GEO3 scenarios are diverse, and we must keep in mind that the timeframes of these scenarios differ [15]. However, beneath the diversity in the choice of scenario names and the narrative motivation for each lies a common set of globalization pathways: a globalizing world with an economic focus, a globalizing world with a focus on sustainability, and a fragmented world resulting from the retreat of globalization. In addition, each pathway has two main variants (see also Table 5).

- In a globalizing world with an economic focus, the scenarios present the following options. In the future of GEO3-MF, developing countries are likely to move slowly toward the 'Age of emerging infectious diseases', while the developed world manages to advance to the 'Age of medical technology'. SRES-A1, on the other hand, is more optimistic about the mitigation of social and environmental problems through global economic and technological developments. These developments make it possible for developing countries to experience improvements in health and increased life expectancy, while at the same time experiencing emerging infectious diseases.

- In a globalizing world with a focus on sustainability, as described by SRES-B1 and GEO3-SuF, both developing and developed countries are likely to advance to the 'Age of sustained health'. However, in the future described by the GEO3-PF scenario, global advances toward sustainability are slow and the developing countries are not likely to advance beyond the 'Age of chronic diseases'. The developed countries progress toward the 'Age of sustained health', but are not able to complete the transition to a sustainable society.
- The scenarios that unfold a fragmented world, SRES-A2 and GEO3-SeF, can be related to a future where the developed world is likely to advance to the 'Age of medical technology', but may also experience an increased risk of infectious disease. The developing countries shift into the 'Age of emerging infectious diseases'. In the alternative fragmented future presented in SRES-B2 there is some local and slow progress in achieving sustainability in the developed world, but the transition is not complete. In developing countries, life expectancy increases but the pace of health improvements is too slow for a shift beyond the 'Age of chronic diseases'. Some developing countries might achieve modest technological progress by themselves.

TABLE 5. Future health in a globalizing world: linking the SRES-scenarios and the GEO3-scenarios to future images of health (adopted from [15, 16]).

Globalization pathway	Variant (scenarios*)	Future health image	
		developed world	developing world
Globalization with an economic focus	Lower mitigation capacity (GEO3-MF)	'Age of medical technology'	Gradual shift into the 'Age of emerging infectious diseases'
	Higher mitigation capacity (SRES-A1)	'Age of medical technology'	'Age of chronic diseases'/ 'Age of emerging infectious diseases'
Globalization with a focus on sustainability	Rapid progress (SRES-B1 & GEO3-SuF)	'Age of sustained health'	'Age of sustained health'
	Slower progress (GEO3-PF)	Progress toward the 'Age of sustained health', but transition is not complete	'Age of chronic diseases'
Fragmentation (retreat of globalization)	Economic focus (SRES-A2 & GEO3-SeF)	'Age of medical technology'/ 'Age of emerging infectious diseases'	'Age of emerging infectious diseases'
	Some local focus on sustainability (SRES-B2)	Progress toward the 'Age of sustained health' (possibly via the 'Age of medical technology'), but transition is not complete	'Age of chronic diseases' (some local progress toward the 'Age of medical technology').

9. Discussion

The world around us is becoming more interconnected and complex, and human health is increasingly perceived as the integrated outcome of its ecological, socio-cultural, economic and institutional determinants. The effects of globalization are causing a growing concern for human health, and the intergenerational equity implied by 'sustainable development' forces us to consider the right of future generations to a healthy environment and healthy lives.

Scenario analyses are useful tools for the exploration of possible health impacts of different globalization pathways, and can be used to gain insights with regard to future global health and to support the decision-making process. An integrated set of global health scenarios could make a significant contribution to ongoing discussions on the health effects of globalization, and could stimulate a more integrated approach toward global health among scientists, governments and other stakeholders.

Recent research shows that the human health dimension is largely missing in existing global scenarios [15, 16]. Given that health is widely regarded as one of the most important aspects of human well-being and an important component of human security, one might ask why there has been so little effort to explicitly address human health in the past decade of scenario development. From the point of view of the global scenario community, exploring the potential health impacts of global changes poses a difficult challenge. Health is an integrated bottom-line outcome, and scenario builders might hesitate to include such a complex and multi-causal issue into their studies. From a public health point of view, exploration of these global, long term and complex risks to human health is far removed from the tidy examples that abound in textbooks of epidemiology and public health research. It is difficult to engage epidemiologists and other population health scientists in this unfamiliar domain. As a result, health is only beginning to play a role in global scenario assessments.

There are two main approaches to the development of global health scenarios [26]. First, one could develop new integrated health scenarios from scratch. This would be, of course, very challenging, but it would be possible to make use of the expertise already available in the scenario community. The second approach would build on the outcomes of earlier studies and would enrich existing global scenarios with a health component.

This paper describes an initial attempt to follow the second approach, adding a health dimension to existing global scenarios to explore the

health effects of future globalization. It provides useful insights in how to incorporate health in scenarios and shows that a comprehensive picture of future health evolves when all relevant socio-cultural, institutional, economic and environmental developments are taken into account. In order to connect current scenarios to a more robust analysis of changes in health outcomes, supplementary analysis is required. For example, an additional step would be the quantification of narrative storylines through modeled scenarios and quantitative estimates of relevant indicators such as life expectancy, healthy life expectancy or disease specific morbidity and mortality rates.

To conclude, the integration of health into global scenario development has the potential to be both instructive and exciting. In today's era of globalization, global environmental change and the subsequent increasing concern for present and future human health, the call for good global health governance becomes stronger and stronger. International agreements and conventions regarding environment, energy and many other sustainability issues need to be informed by the most comprehensive information regarding future scenarios and associated model outcomes – and health should be an integral part of this information.

REFERENCES

1. McMichael A.J., Integrated assessment of potential health impact of global environmental change: prospects and limitations, *Environmental Modeling and Assessment*, 1997, 2: 129-137.
2. Martens P., McMichael A.J., Patz J., Globalization, environmental change and health, *Global Change and Human Health*, 2000, 1: 4-8.
3. Huynen M.M.T.E., Martens P., Hilderink H., *The health impacts of globalization: a conceptual framework*, Bilthoven: Netherlands Environmental Assessment Agency (MNP-RIVM), 2005.
4. WHO, *The world health report 2002: reducing risks, promoting healthy life*, Geneva: World Health Organization, 2002.
5. Huynen M.M.T.E., Martens P., Hilderink H., The health impacts of globalization: a conceptual framework, *Globalization and Health*, 2005, 1: article number 14 (12 pages).
6. Rennen W., Martens P., The globalization timeline, *Integrated Assessment*, 2003, 4: 137-144.
7. IPCC, *Special Report on Emission Scenarios*, Cambridge: Cambridge University Press, 2000.

8. UNEP, *Global Environmental Outlook 3*, London: Earthscan, 2002.
9. Gallopin G., Hammond A., Raskin P., Swart R., *Branch points: global scenarios and human choice. A resource paper of the Global Scenario Group*, Stockholm: Stockholm Environment Institute, 1997.
10. NIC, *Global trends 2015: a dialogue about the future with non-governmental experts*, National Intelligence Council, 2000.
11. Glenn J.C., Gordon T.J. (eds.), *1998 State of the future: issues and opportunities*, Washington: American Council for the United Nations University, 1998.
12. Hammond A., *Which world? Scenarios for the 21st century: global destinies, regional choices*, London: Earthscan, 1998.
13. WBCSD, *Exploring sustainable development: global scenarios 2000-2050*, London: World Business Council for Sustainable Development, 1998.
14. Gallopin G., Rijsberman F., Three global water scenarios, *International Journal of Water*, 2000, 1: 16-40.
15. Huynen M.M.T.E., Martens P., *Future health: the health dimension in global scenarios*, Maastricht: ICIS, 2002.
16. Martens P., Huynen M.M.T.E., A future without health? Health dimension in global scenario studies, *Bulletin of the World Health Organization*, 2003, 81: 896-901.
17. Omran A.R., The epidemiological transition: a preliminary update, *Journal of Tropical Pediatrics*, 1983, 29: 305-316.
18. Omran A.R., The epidemiological transition theory revisited thirty years later, *World Health Statistics Quarterly*, 1998, 51: 99-199.
19. Mackenbach J.P., The epidemiological transition theory, *Journal of Epidemiology and Community Health*, 1994, 48: 329-331.
20. Newcomb J., *Biology and borders: SARS and the new economics of biosecurity*, Cambridge: Bio Economic Research Associates, 2003.
21. Olshansky S.J., Carnes B.A., Rogers R.G., Smith L., Emerging infectious diseases: the fifth stage of the epidemiological transition?, *World Health Statistics Quarterly*, 1998, 51: 207-217.
22. Martens P., Health transitions in a globalizing world: towards more disease or sustained health?, *Futures*, 2002, 37: 635-648.
23. Barrett R., Kuzawa C.W., McDade T., Armelagos G.J., Emerging and re-emerging infectious diseases: the third epidemiologic transition, *Annual Review of Anthropology*, 1998, 27: 247-271.
24. Bobadilla J.L., Possas C., Health policy issues in three Latin American countries: implications of the epidemiological transition. In: *The epidemiological transition: policy and planning. Implications for develop-*

ing countries, Gribble J., Preston S. (eds.), Washington, DC: National Academy Press, 1993.

25. *The Earth Charter Initiative*, www.earthcharter.org. Accessed January 24, 2006.
26. Huynen M., Scenarios and global health: the road ahead, *IHDP Newsletter*, issue 3/2003, 2003: 14.

EFFECTS OF SOCIAL, ENVIRONMENTAL AND ECONOMIC FACTORS ON CURRENT AND FUTURE PATTERNS OF INFECTIOUS DISEASES

DAVID L. HEYMANN

1. Introduction: Microbial Resilience

Social, environmental and economic factors, linked to a host of human activities, often accelerate and amplify the natural phenomena that modify infectious disease patterns in humans. They influence the ease at which microbes adapt to new environments and hosts, the rapidity with which they develop resistance to the antimicrobial agents used for treatment, and the limits to which they spread geographically. Social factors include the impact of urbanization on sanitation and the water supply; and behaviour – of health workers as they perform their routine functions – and of the general public as they strive to prevent and control the daily threat of infectious diseases. Environmental factors include naturally occurring variations in temperature and rainfall, and the impact of economic development on rivers, forests and agricultural lands; while economic factors include the level of investments in public health, and patterns in international trade and travel as globalization of commerce and markets continues to accelerate.

These factors in turn are influenced by the tools with which public health systems defend against the infectious disease threat – currently available vaccines and antibiotics that were ushered in through intensified research and development that began immediately following World War II; and technological advances in hygiene and sanitation. Industrialized countries have greatly benefited from these public health tools. Malaria, for example, endemic in many industrialized countries in both North America and Europe, disappeared with insecticides and antimalarial drugs. Tuberculosis hospitals in Europe and North America emptied as living conditions improved and effective drugs became available that could be used

to treat those with acute tuberculosis and prevent infection in their contacts. Influenza deaths could be prevented in industrialized countries by vaccinating elderly populations, and epidemics of diarrhoeal disease became rare events, limited to foodborne outbreaks when temporary breaches in sanitation occurred. And one infectious disease, smallpox, completely disappeared from industrialized countries through routine use of the smallpox vaccine, and then in the rest of the world after a global eradication effort [1, 2].

Progress has been more difficult and variable in developing countries, and has slowed as public health investments made by governments have decreased. At the same time, the microbes that cause infectious diseases remain complex, dynamic, and constantly evolving. They reproduce rapidly, mutate frequently, and adapt with relative ease to new environments and hosts; and they develop resistance to the drugs used to treat them. With increasing use of antimicrobial drugs, warning signs of microbial resilience began to appear. By the end of the 1940s resistance of hospital strains of staphylococcus to penicillin in the United Kingdom were as high as 14%, and by the end of the 1990s had risen to levels of 95% or greater [3]. In New York City in the 1990s multidrug-resistant strains of tuberculosis gained their hold in hospitals, prisons, and homeless populations [4]. At the same time multidrug-resistant tuberculosis merged in the Russian Federation and more than doubled in less than seven years, with over 20% of tuberculosis patients in prison settings infected with multidrug-resistant strains [5]. Antimicrobial drugs developed to treat AIDS and other sexually transmitted infections such as gonorrhoea likewise began to lose their efficacy because of the rapid development of resistance. During this same period, vaccine development lagged – with no effective vaccines to prevent infection of many of the major mortality causing infections such as tuberculosis, malaria, and AIDS. And newly emerging infectious diseases such as severe acute respiratory syndrome (SARS) and avian influenza (H5N1) have presented new global infectious disease challenges as they spread from country to country, and continent to continent [6].

2. Social and Behavioural Factors that Influence Infectious Diseases

The world's population more than doubled in the second half of the 20th century, accelerating most rapidly in the developing countries of the tropics and sub-tropics, where infectious diseases continued to have a hold

[7]. Rural-urban migration resulted in inadequacy of sanitation, crowded living conditions and other basic infrastructure associated with population growth. It thus contributed to the resurgence of many diseases, such as tuberculosis, cholera, typhoid, and plague, that are transmitted when living conditions and hygiene are sub-standard, and when overcrowding occurs. Cholera caused epidemics during the 1990s in parts of Latin America where it had previously been quiescent for over 100 years. By the 1980s crowded major urban areas in Africa and South America had experienced a dramatic re-emergence of yellow fever epidemics as the yellow fever virus was introduced by mosquitoes from rain forests into new and densely populated urban areas where bednets to protect from mosquito bites were no longer being used [8,9].

Behaviours such as over or under-prescribing of antibiotics by health workers, and excessive demand for antibiotics by the general population, have had a remarkable impact on the selection and survival of resistant microbes, rapidly increasing levels of microbial resistance. Drug-resistant microbes have then spread from person to person and geographically, raising the prospect that common infectious diseases could become prohibitively expensive or impossible to treat [10]. The bacterial infections which contribute most to human disease are also those in which emerging resistance is of most concern: diarrhoeal diseases such as dysentery; respiratory tract infections, including pneumococcal pneumonia and multidrug-resistant tuberculosis; sexually transmitted infections such as gonorrhoea; and a host of hospital-acquired infections that are notoriously difficult and expensive to treat. Among the major infectious diseases, the development of resistance to drugs commonly used to treat malaria is of particular concern, as is the emerging resistance to anti-HIV drugs. Most alarming of all are microbes that have now accumulated resistance genes to virtually all currently available antimicrobial drugs, such as *Staphylococcus aureus* and *Salmonella typhi*, that now have the potential to cause untreatable infections.

Trends in tourism, with tourists penetrating deep into tropical forests, often without appropriate protection against insect bites or vaccination, result in importations of malaria and yellow fever to industrialized countries [11]. At the same time weak infection control procedures by health workers have caused the amplification of transmission in outbreaks such as Ebola to health workers and their contacts in sub-Saharan Africa, and hepatitis and SARS to health workers and those with whom they have contact in both developing and industrialized countries [12,13].

3. Environmental Factors that Influence Infectious Diseases

Human disturbance and alternation of ecological zones throughout the world has increased the frequency with which microbes, usually confined to animals, cross the species barrier to infect humans. Deforestation disrupts natural habitats of animals, and can force animals, searching for food, into closer contact with humans. Outbreaks of Lassa fever in West Africa and of hantavirus in North America have been linked to such phenomena [14, 15]. In Latin America, Chagas disease emerged as an important human disease after deforestation caused the insect that transmits the infection to move from its wild natural hosts to involve humans and domestic animals in the transmission cycle, eventually transforming the disease into an urban infection that can be now also transmitted by blood transfusion [16].

Climate extremes, whether involving excessive rainfall or drought, can likewise displace animal species and bring them into closer contact with human settlements, or increase vector breeding sites. A 1998 outbreak of Japanese encephalitis in Papua New Guinea has been linked to extensive drought, which led to increased mosquito breeding as rivers dried into stagnant pools [17]. The Japanese encephalitis virus is now widespread in Papua New Guinea and threatening to move farther east.

An outbreak of Rift Valley Fever in Eastern Kenya resulted from flooding related to El Niño. Humans and cattle, forced to live in close proximity on islands of dry land surrounded by water, facilitated the transfer or the Rift Valley Fever virus from unvaccinated animals to humans by mosquitoes that had increased in numbers because of the abundance of pooled-water breeding sites [18].

Other examples of how insects that carry infectious diseases have exploited new opportunities created by environmental degradation and human behavioural change include epidemics of dengue and yellow fever that have been fuelled by the adoption of modern consumer habits in urban areas where discarded household appliances, tyres, plastic food containers and jars have created abundant artificial mosquito breeding sites. The *Aedes aegypti* mosquito is now well established in most, if not all, large African cities, increasing the risk of explosive urban outbreaks of Dengue [19]. Similar examples are occurring in Asia where Dengue and Dengue Haemorrhagic Fever have caused major outbreaks during 2004 in Indonesia and India [20]. In countries of the former Soviet Union, large amounts of stagnant water, created by ineffective irrigation schemes, provided mosquito breeding sites that permitted the re-emergence of malaria

in the most southern states, where a few incidental and probably imported cases in Tajikistan in the early 1990s multiplied to almost 20,000 reported cases in 1998 [21]. Such problems are compounded by the very small number of new cost-effective chemical pesticides, suitable for public health, that have been developed in recent years.

Though intensive research has failed to disclose the origins of Marburg and Ebola haemorrhagic fever outbreaks, microbes causing both diseases are also thought to be transmitted to humans as they encounter animal sources somewhere in the transmission cycle [22]. An outbreak of Ebola haemorrhagic fever in humans in 1995 was linked to a woodsman, who worked deep within the tropical rainforest making charcoal, and who is somehow thought to have become infected with the Ebola virus which he then carried back to his home village and family members, while a Swiss researcher is thought to have become infected with the Ebola virus while searching for the cause of a major die out of chimpanzees in a forest reserve in West Africa [23,24].

The consequences of the environment and interspecies transmission of microbes are most clearly demonstrated in the case of the influenza virus. It is thought to be only matter of time until an animal influenza virus circulating in domestic animals, recombines with a human influenza virus, and causes the next highly lethal influenza pandemic [25]. Intensive farming practices have placed humans in Asia in close proximity to domestic animals in densely populated areas. In 1997 in the Hong Kong Special Administrative Region of China, crowded conditions, and live poultry markets adjacent to residential areas, facilitated the transmission of a new avian influenza A virus (subtype H5N1), previously thought confined to birds. At least 18 humans were infected and six died, raising considerable alarm [26]. Although human-to-human transmission of the virus was documented, it was found to be relatively inefficient and uncommon [27]. A re-emergence of this same virus throughout Asia in late 2003 and 2004 has resulted in 42 human infections with 30 deaths in Thailand and Vietnam by 1 October 2004, and the continued threat of a global human pandemic [28].

3.1. *Economic Factors that Influence Infectious Diseases*

With the control of infectious diseases in industrialized counties during the 20th century came great optimism that infectious diseases were no longer a health problem. As a result of this optimism financial resources once used to combat infectious diseases were shifted to non-communicable disease

problems, and there was decreased and insufficient investment during the last quarter of the 20th century to adapt and use newly developed technologies for detection, monitoring and responding to infectious diseases [29].

In parallel, investment in research and development for new vaccines and antimicrobial drugs for infectious diseases was shifted to non-communicable diseases, those that still affected industrialized countries and were related to lifestyle, aging and environmental hazard. The resultant 10/90 research gap, with less than 10% of public and private funds being placed into research for infectious diseases such as tuberculosis, diarrhoeal diseases, malaria and AIDS, created a slow down in the development of new antimicrobial drugs and vaccines [30].

With decreased investment came the resurgence of known infectious diseases in industrialized countries, such as tuberculosis and common sexually transmitted infections such as syphilis and gonorrhoea; and a host of new infectious diseases such as hepatitis C and AIDS that escaped detection until they were firmly implanted in human populations [31].

Developing countries, on the other hand, had never accomplished the same decrease in infectious diseases. In these countries infectious diseases continued to remain an important cause of sickness, disability and death throughout the 20th century because of a lack of access to vaccines and drugs with which to prevent and treat them, and because of weak health systems that failed to consistently reach populations in need. The resources that these countries were able to invest in infectious disease control in the first half of the 20th century, often with the assistance of the colonial powers, decreased as the millennium progressed, with some of the greatest decreases occurring in sub-Saharan Africa [32].

Common infectious diseases that are spread directly from person to person such as tuberculosis continued to cause significant suffering and death in developing countries because of inequitable distribution of and curative drugs within developing country health systems, and because of the continued lack of development of preventive vaccines [33]. Treatment with partial regimes of tuberculosis drugs because of inconsistent drug supply to developing country health systems caused an increase in drug-resistant strains that resulted in an increase in cost for treatment because of the need to use more expensive second line tuberculosis drugs [34].

Infections such as diarrhoeal diseases of children, caused by lack of adequate sanitation and safe water, increased as sanitation systems in major metropolitan areas failed with increases in urban populations, and those in rural areas failed to develop. Vaccine preventable diseases such as measles,

for which access to vaccines was likewise not sustained, remained major public health problems. Decreased investment in childhood immunization programmes in Russia and the Ukraine in the early 1990s resulted in epidemics of diphtheria in Russia and the Ukraine in the early 1990s [35].

With decreased financial investment in programmes to control infection-carrying mosquitoes, a resurgence of malaria, dengue and yellow fever occurred, and mosquitoes then spread to new geographic areas. Following the deterioration of *Aedes aegypti* control campaigns during the 1970s, dengue resurged dramatically, with unprecedented numbers of its haemorrhagic form [36]. Prior to 1970, only nine countries had experienced epidemics of dengue. By 1998 a dengue pandemic occurred in which 1.2 million cases were reported from 56 countries. Since then dengue has continued to cause major epidemics with the most recent epidemic in 2004 resulting in over 60,000 cases and 700 deaths in Indonesia alone [37].

Other infectious diseases such as African trypanosomiasis or sleeping sickness began to resurge in the 1980s with a decline of most surveillance and tsetse fly control activities [38]. By the late 1990s, approximately 26% of mortality in low income countries was directly caused by infectious diseases that continued to circulate among large urban and rural populations. These low-income countries today represent over 80% of the estimated 14 million deaths caused each year by infectious diseases [39].

Other economic factors are also playing a role in infectious disease patterns throughout the world. Globalization, with a phenomenal growth in international travel and trade since the 1950s, has greatly increased the speed with which microbes, incubating in unsuspecting humans, can cross continents and invade new geographic territories. At the same time microbes living in insects concealed in cargoes or in the luggage holds and cabins of jets; in animals traded internationally, or in improperly or non-processed food and food products can also travel across continents and internationally. As a result, the threat of epidemic diseases with origins in one country and spread to others has become a real and constant threat.

Nothing more clearly demonstrates this global threat than the spread of AIDS in humans throughout the world during the latter half of the 20th century. Spreading throughout the world and amplified by unsafe sexual behaviour, AIDS has had a negative impact on economic development and healthy population growth. In 1999, the lower figure in the world life expectancy range, which had seen a steady increase in previous decades, declined to 33.2 years, just above the 33 years seen in 1949, largely due to the emergence and spread of HIV [40]. In recent years, every continent has

experienced an unexpected outbreak of some infectious disease directly related to increased travel and trade, the most recent having been severe acute respiratory syndrome (SARS) [41].

Advances in food production and storage technology, coupled with the globalization of markets, have resulted in a food chain that is unprecedented in its length and complexity, thus creating an efficient vehicle for microbes to spread to new areas and susceptible hosts. Tracing the origin of all ingredients in a meal has become virtually impossible, constituting an enormous challenge for the control of foodborne diseases [42]. Medical advances in such areas as blood transfusion, organ transplantation and other sophisticated surgical procedures, and the development of intensive care units have likewise opened new opportunities for the microbial world, creating ideal conditions for in-hospital transmission of infectious agents to new, atypical hosts [43].

In the late 1990s infections such as West Nile fever, that arrived in North America through the introduction of a single virus, and Rift Valley fever that arrived in the Arabian peninsula in infected livestock, have become endemic in these new geographic areas, adding to the infectious disease burden [44,45,46]. Once established on new continents, emerging or re-emerging infectious diseases change the dynamics in local infectious disease patterns change.

The universal nature of the microbial threat, with agents of disease, including drug resistant forms, passing undetected across increasingly porous borders, has placed all nations on an equally vulnerable footing. Economic prosperity has produced a world that is interconnected in matters of economics and trade, with the result that health has become both a domestic issue, and an issue with foreign policy considerations as well [47].

4. Future Patterns of Infectious Diseases

Social, environmental and economic factors; the availability of antimicrobial drugs and vaccines; and the resilience and natural selection and evolution of the microbial world will continue to have an impact on future patterns of infectious diseases. Water and sanitation systems will be challenged as populations continue to move to urban areas in search of work and economic betterment, with continued endemic and epidemic transmission of intestinal infections and diarrhoeal disease. Behaviour of health workers and the general population will continue to play an important role

in transmission of infectious diseases, and in development of antimicrobial resistance. Continued lack of vaccines for many of the major infectious diseases will dampen progress in prevention, while continued alterations in temperature and rainfall, and human impact on agricultural lands, forests and rivers will in some instances increase the number of insect vectors and alter the geographic distribution of animal hosts, leading to the emergence of new human infections and/or re-emergence of those that are known.

Underlying all these factors is the current acceleration in globalization, increasing the risk that infections in one country spread internationally in humans, insects, animals or food – thus raising infectious diseases higher on the agenda of human security. And finally, with continued inequitable distribution of the vaccines, medicines and goods available now to prevent, treat and control infectious diseases, coupled with weak public health delivery systems in developing and low income countries a disproportionate burden of human suffering and death from infections diseases will continue to occur in developing countries.

Public-private partnerships are currently addressing some of the needs to modify these future patterns of infectious diseases. The global Alliance for Vaccines and Immunizations provides financial support to accelerate the development of new vaccines needed primarily in developing countries, and has strategic and financial mechanisms for ensuring that new vaccines, when available, are supplied to developing countries in the highest quality and at the lowest possible price. Another public-private partnership, the Global Fund to Fight AIDS, Tuberculosis and Malaria, helps countries address complex and pressing problems that include supplies of medicines and other goods, weak health infrastructure, poorly trained personnel and weak delivery systems. Other public/private partnerships are addressing diseases such as leprosy, lymphatic filariasis and polio, the latter a disease that is targeted for eradication sometime during 2005 [48].

At the same time, with the continued evolution of antimicrobial resistance, our armamentarium for infectious disease control requires continued and increased investment for new antimicrobial drugs and for vaccines. For new antimicrobial agents and vaccines, even if the pharmaceutical industry were to step up efforts to develop new drugs immediately, current trends suggest that some diseases may have very few and, in some instances, no effective therapies within the next ten years [49]. Moreover, if current trends continue, many important medical and surgical procedures, including cancer chemotherapy, bone marrow and organ transplantation, and hip and other joint replacements, could no longer be undertaken out of fear that the

associated compromise of immune function might place patients at risk of acquiring a difficult to treat and ultimately fatal infection. Opportunistic infections in AIDS patients would likewise become an especially difficult challenge, and the choice of antimicrobial drugs for most infectious diseases would be severely limited. In research for new drugs, as well as that for vaccines, the challenge is to apply the genomic information and knowledge that has become available to tailor make the medicines and vaccines needed.

To minimize the impact of emerging and re-emerging infectious diseases, increased investment in national pubic health infrastructure is required in order to detect early and rapidly respond to infectious diseases outbreaks, as well as long-term programmes to modify or remove the factors that facilitate emergence. Likewise there is a need for an international safety net of global surveillance with a response mechanism should diseases begin to travel internationally. Recently a global partnership, supported by several new mechanisms and a computer-driven tool for real time gathering of disease intelligence, has been developed to detect and respond to infectious diseases of international importance [50].

Related to the emerging and re-emerging infectious diseases is a new infectious disease threat that dominates public health thinking and policies in some industrialized countries – that of deliberately-caused infectious disease outbreaks. Following the deliberate dissemination of anthrax spores through the US postal system in 2001, questions concerning the deliberate use of biological or chemical weapons have been raised with great urgency. The prospect of introduction of an infectious disease to non-immune populations that could cause severe illness and death has now become a stark reality.

Infectious diseases have caused human suffering, illness and death since Biblical times, and before. The threats posed by infectious diseases are today being amplified by social, economic and environmental factors that accelerate the natural phenomena that modify infectious disease patterns. Global recognition and partnership are required to keep them at bay.

REFERENCES

1. Fenner, F., Henderson D.A., Jezek Z., Ladnyi I.D., *Smallpox and its eradication*, Geneva, World Health Organization, 1988.
2. Lederberg J., Shope R.E., Oaks S.C. Jr (eds.), *Emerging infections: microbial threats to health in the United States*, Washington DC, National Academy Press, 1992.

3. Levy S.B., *The Antibiotic Paradox: how miracle drugs are destroying the miracle*, Plenum Press, 1992.
4. Frieden T.F., Fujiwara P.I., Washko R.M., Hamburg M.A., Tuberculosis in New York City – turning the tide, *N. Engl. J. Med.* 1995, 333: 229-33.
5. Small M.A., A view from the ground: tuberculosis as an example of reemerging infectious disease in the former Soviet Union, in: Davis J.R., Lederberg J. (eds.), *Emerging infectious diseases from the global to the local perspective*, Washington D.C., National Academy Press, 2001.
6. Heymann D.L., The Fall and Rise of Infectious Diseases. Bioalert: Disease Knows no Borders, *Georgetown Journal of International Affairs*, Summer/Fall 2001, 7-14.
7. *World population prospects, the 2000 revision, volume 1: comprehensive tables*, New York, United Nations, 2001 (ST/ESA/SER.A/198).
8. Cholera, 2000, *Wkly Epidemiol. Rec.* 2000; 75: 233-40.
9. Vasconcelos P.F.C., Rosa A.P.A.T., Pinheiro F.P., Rodrigues S.G., *et al. Aedes aegypti*, dengue and re-urbanization of yellow fever in Brazil and other South American countries – past and present situation and future prospects, *Dengue Bull.* 1999; 23: 55-66.
10. *WHO global strategy for containment of antimicrobial resistance*, Geneva, World Health Organization, 2001 (WHO/CDS/CSR/DRS/ 2001.2). Further information available at: www.who.int/emc/ amr.html
11. World Health Organization, *Airport Malaria*, 1996. www.who. int/docstore/wer/pdf/1996/wer7147.pdf
12. Kahn A.S., Kweteminga, T., Heymann D.L., The reemergence of Ebola haemorrhagic fever, Democratic Republic of the Congo, 1995, *The Journal of Infectious Diseases*, 1999, 179 (suppl. 1):S76-86.
13. Heymann D.L., Rodier G., SARS: A global response to an international threat, *The Brown Journal of World Affairs*, Winter/Spring 2004, X(2):185.197.
14. Birmingham K., Kenyon G., Lassa fever is unheralded problem in West Africa, [News] *Nat. Med.* 2001; 7: 878, 2001
15. 12 Update: Hantavirus pulmonary syndrome – United States, 1993, *MMWR* 1993; 42: 816-20.
16. Control of Chagas disease, *World Health Organ. Tech. Rep. Ser.* 1991; 811.
17. Mackenzie J.S., Emerging diseases in the Australasian region, in: Davis J.R., Lederberg J. (eds.), *Emerging infectious diseases from the global to the local perspective*, Washington DC, National Academy Press, 2001.

18. Outbreak of Rift Valley fever, Yemen, August-October 2000, *Wkly Epidemiol. Rec.* 2000; 75: 392-94.
19. Rodier G.R., Ryan M.J., Heymann D.L., Global epidemiology of infectious diseases, in Strickland G.T. (ed.), *Hunter's tropical medicine and emerging infectious diseases*, eighth edition, Philadephia, WB Saunders Company, 2000.
20. World Health Organization, *Dengue Indonesia update 4*, 2004. www.who.int/csr/don/2004_05_11a/en/
21. Small M.A., A view from the ground: tuberculosis as an example of a reemerging infectious disease in the former Soviet Union, in: Davis J.R., Lederberg J. (eds.), *Emerging infectious diseases from the global to the local perspective*, Washington DC, National Academy Press, 2001.
22. Klenk H.D., Slenczka W., Feldmann H., Marburg and Ebola viruses, in: Webster T.G., Granoff A. (eds.), *Encyclopedia of Virology*, New York, Academic Press, 1994.
23. Formenty P., Boesch C., Wyers M., Ebola outbreak among wild chimpanzees living in a rain forest of Côte d'Ivoire, *The Journal of Infectious Diseases*, 1999.
24. Kahn A.S., Kweteminga T., Heymann D.L., The reemergence of Ebola haemorrhagic fever, Democratic Republic of the Congo, 1995, *The Journal of Infectious Diseases*, 1999, 179 (suppl. 1):S76-86.9; 179 (suppl. 1):S120-6. 1999.
25. Influenza pandemic preparedness plan, Geneva, World Health Organization, 1999 (WHO/CDS/CSR/EDC/99.1).
26. Isolation of avian influenza A(H5N1) viruses from humans – Hong Kong, May-December 1997, *MMWR* 1997; 46:1205-7.
27. Class E.C.J., Osterhaus A.D.M.E., van Beek R. *et al.*, Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus, *Lancet* 1998; 351: 472-77.
28. Avian influenza update Http://www.who.int/csr/disease/avian_influenza/country/cases_table_2004_09_28/en
29. Lederberg J., Shope R.E., Oaks S.C. Jr (eds.), *Emerging infections: microbial threats to health in the United States*, Washington DC, National Academy Press, 1992.
30. World Health Organization, *Removing obstacles to health development 1999*, .int/infectious-disease-report/index-rpt99.html
31. Lederberg J., Shope R.E., Oaks S.C. Jr (eds.), *Emerging infections: microbial threats to health in the United States*, Washington DC, National Academy Press, 1992.

32. World Health Organization, *Removing obstacles to health development 1999*, www.who.int/infectious-disease-report/index-rpt99.html
33. World Health Organization, *Removing obstacles to health development 1999*, www.who.int/infectious-disease-report/index-rpt99.html
34. World Health Organization, *Overcoming antimicrobial resistance*, www.who.int/infectious-disease-report-2000/
35. Weinberg J., European responses to emerging infections and their policy implications, in: Davis J.R., Lederberg J. (eds.), *Emerging infectious diseases from the global to the local perspective*, Washington DC, National Academy Press, 2001.
36. Dengue/dengue haemorrhagic fever: situation in 2000, *Wkly Epidemiol. Rec.* 2000; 75: 193-96.
37. World Health Organization, *Dengue Indonesia update* 4, 2004. www.who.int/csr/don/2004_05_11a/en/
38. Molyneux D.H., Vector-borne infections in the tropics and health policy issues in the twenty-first century, *Trans. R. Soc. Trop. Med. Hyg.* 2001; 95: 1-6.
39. World Health Organization, *Removing obstacles to health development 1999*, www.who.int/infectious-disease-report/index-rpt99.html
40. *The World Health Report 2000*, Health systems: improving performance, Geneva, World Health Organization, 2000.
41. Heymann D.L., Rodier G., SARS: A global response to an international threat, *The Brown Journal of World Affairs*, Winter/Spring 2004, X(2):185.197.
42. Rodier G.R., Ryan M.J., Heymann D.L., Global epidemiology of infectious diseases, in: Strickland G.T. (ed.), *Hunter's tropical medicine and emerging infectious diseases*, eighth edition, Philadelphia, WB Saunders Company, 2000.
43. Rodier G.R., Ryan M.J., Heymann D.L., Global epidemiology of infectious diseases, in: Strickland G.T. (ed.), *Hunter's tropical medicine and emerging infectious diseases*, eighth edition, Philadelphia, WB Saunders Company, 2000.
44. Petersen L.R., Toehrig J.T., West Nile virus: a reemerging global pathogen, *Emerg. Infect. Dis.* 2001; 7: 611-614.
45. Outbreak of Rift Valley fever, Saudi Arabia, August-October 2000, *Wkly Epidemiol. Rec.* 2000; 75: 370-71.
46. Outbreak of Rift Valley fever, Yemen, August-October 2000, *Wkly Epidemiol. Rec.* 2000; 75: 392-94.

47. Heymann D.L., Evolving infectious disease threats to national and global security, *Global Health Challenges for Human Security,* Chen, Leaning, Narasimhani (eds.), pp. 105-124, Harvard University Press, 2003.
48. World Health Organization, *Global defence against the infectious disease threat*, 2002, www.who.int/infectious-disease-news/cds2002.
49. WHO, *Global strategy for containment of antimicrobial resistance*, Geneva, World Health Organization, 2001 (WHO/CDS/CSR/DRS/ 2001.2). Further information available at: www.who.int/ emc/amr.html
50. Heymann D.L., Rodier F., Hotspots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases, *The Lancet Infectious Diseases*, vol. 1, December 2001, 345-353.

WHO Web Addresses for Further Information

World Health Organization: www.who.int
Disease outbreak news: www.who.int/disease-outbreak-news/
Weekly Epidemiological Record: www.who.int/wer/
Infectious diseases: www.who.int/health-topics/idindex.htm
Communicable disease surveillance and response: www.who.int/emc/index.html
WHO training office in Lyon, France: www.who.int/emc/lyon/ index.html
HealthMap: www.who.int/emc/healthmap/healthmap.html

EPIDEMICS AND ECONOMICS[1]

DAVID E. BLOOM & DAVID CANNING

Introduction

Principiis obsta; sero medicina paratur cum mala per longas convaluere moras.
'Stop it at the start; it is late for medicine to be prepared when disease has grown strong through long delays'.
(Ovid, around 17 AD)

The association between health and wealth is well established. Rich countries tend to have healthier populations than poor countries. Figure 1 (see page 419) shows that life expectancy, a commonly used summary measure of a population's health, increases sharply as income per capita rises among poorer countries and continues to rise, though at a much slower rate, as countries become wealthier. Between 1995 and 2000, life expectancy at birth in the world's least developed countries was 51 years. In other low-income countries it was 59 years, in middle-income countries 70 years, and in high-income countries 78 years.[2] For decades, it was thought that the causality ran in only one direction – as incomes rose, health improved. Recent years, however, have seen increased attention paid to assessing the reverse effects, as health has come to be seen as a key driver of economic development.

[1] Larry Rosenberg and Mark Weston provided extremely helpful assistance in the preparation of this paper. Marcella Alsan, Till Baernighausen, Elizabeth Brainerd, David Heymann, Maureen Lewis, Marc Lipsitch, Tony McMichael, Darren Morris, Andrew Noymer, Josh Salomon, Kate Taylor, and John Tharakan provided very useful comments. The authors are also grateful to the workshop participants for helpful comments.

[2] Commission on Macroeconomics and Health (2001), *Macroeconomics and Health: Investing in Health for Economic Development*, Geneva: World Health Organisation.

This paper discusses the links between infectious disease epidemics and income and asks how these links are affected by changing global circumstances.[3] At the outset we must recognize that there is considerable heterogeneity among infectious diseases that have the potential to become epidemics – in terms of transmission; latency; prevention, treatment, and care; and their short- and long-term economic effects. Regarding transmission, for example, poverty has a greater influence in some diseases than in others. Because of the wide range of characteristics that distinguish some epidemic diseases from others, detailed consideration of specific diseases is beyond the scope of the paper.

Part 1 of the paper looks at the mechanisms through which health affects wealth and vice versa. It then looks at how epidemics and pandemics in particular interact with economic development. Part 2 discusses the challenges facing those attempting to respond to epidemics. Part 3 offers some ideas for overcoming the challenges, while the final section provides conclusions.

In brief, this paper argues that epidemics of infectious diseases can have sizable economic impacts – both in the short and long term – and that their management and control require investment in national and international health systems. When they threaten to erupt, control of epidemics should be a priority, given the significant human and economic tolls they inflict. Globally, the long-term vulnerability to epidemics should decrease as development standards rise, but a more highly interconnected world may actually promote the occurrence of infectious disease epidemics.[4]

[3] The definition of 'epidemic' does not technically include malaria, because it is globally endemic. However, malaria is tempo-spatially epidemic and is therefore included in the scope of this paper. On the other hand, we do not address 'deliberate' epidemics (e.g., ones that may be caused by bioterrorism), although many of the arguments made herein apply to those as well. We discuss some of the differences between epidemic, pandemic, and endemic diseases below.

[4] The spread of SARS and HIV are cases in point. Still, not every new disease is likely to spread wildly, and those arising in very isolated areas do not necessarily pose near-term threats to the industrialized world.

Part 1: Health and Wealth

Wealth as a Determinant of Health

'Health is the first wealth'
(Ralph Waldo Emerson, *The Conduct of Life*, 1860).

The view that higher incomes lead to better health stems from several mechanisms. Within a population, wealthier individuals can afford the essential products necessary for a healthy life: sufficient, nutritious food, adequate housing, clean water, and quality health care. They also enjoy better access to information and therefore have more knowledge of how to foster and maintain one's health. The wealthy have a stronger influence with policymakers than the poor, so their demands for improvements in health services are more likely to be acted upon. Wealthier countries and their residents, moreover, have greater strengths in education, ability to develop and buy medicines and other health technologies, training and paying doctors and nurses, building and maintaining hospitals and clinics, minimizing nosocomial infections (although recent reports emphasize the extent to which these are still a problem in developed countries), providing water and sanitation facilities, (often) using less noxious energy sources, and implementing comprehensive public health campaigns.[5]

The empirical literature, however, is mixed with respect to the link that runs from average income to population health. Lant Pritchett and Lawrence Summers have found that differences in income growth account for 40% of the difference in mortality rate reductions between countries.[6] Angus Deaton, however, finds that 'gains in life expectancy have been only weakly correlated with growth rates'.[7] Mortality has declined as incomes have risen, but it has also fallen (at least until the advent of HIV/AIDS) in countries that have remained poor. Some poor countries and regions, notably Cuba and the state of Kerala in India, have first-rate health systems and healthy populations. Some rich countries, on the other hand, have pockets of ill health (poor communities in the US

[5] All of this said, we note that population aging in wealthy countries is one factor that will make them more vulnerable to epidemics.

[6] Pritchett, Lant, and Lawrence H. Summers (1996), Wealthier is Healthier, *Journal of Human Resources*, 31 (4): 842-68.

[7] Deaton, Angus (2004), Health in an Age of Globalization, *NBER Working Paper No. w10669*, August.

and the OECD, and indigenous populations in the US and Australia, for example).[8] Improving economic conditions can lead to better health, if governments and societies mobilize to bring it about, but the link is not assured.[9] Other factors that may affect population health at a given level of income are the distribution of income[10] and the level and distribution of health spending and the organization of the health system.[11]

Health as a Determinant of Wealth

There are a number of mechanisms through which population health can increase or decrease per capita income. First, health can affect income via its intermediate effect on education. Healthy children are less likely to suffer from impaired cognitive development and better able to attend school regularly.[12] Their parents are more likely to invest in their education if they are confident their offspring will live long enough to benefit from it. Better-educated children, in turn, have greater potential to contribute to a country's economic development when they reach working age.

Second, good health boosts labor productivity. As with schoolchildren, healthier workers attend work more often and are more mentally alert and physically energetic. This effect is particularly important in poor countries, where, unlike in developed countries, people whose sickness

[8] Sen, Amartya (1993), The Economics of Life and Death, *Scientific American*, 268(5): 40-47.

[9] Moreover, developed countries often experience poor health and even epidemics, such as obesity, that are based on behavior.

[10] Kennedy, Bruce P., Ichiro Kawachi, and Deborah Prothrow-Stith (1996), Income Distribution and Mortality: Cross-Sectional Ecological Study of the Robin Hood Index in the United States, *British Medical Journal*, 312: 1004-1007; Wilkinson R.G. (1996), *Unhealthy Societies. The Afflictions of Inequality*, London: Routledge.

[11] Anand, Sudhir, and Martin Ravallion (1993), Human Development in Poor Countries: On the Role of Private Incomes and Public Services, *Journal of Economic Perspectives*, 7:133-50; Anand, Sudhir, and Till Bärnighausen (2004), Human Resources and Health Outcomes: Cross-Country Econometric Study, *The Lancet*, 364 (30 October) 1603-09.

[12] Leslie, J. and Dean Jamison (1990), Health and Nutrition Considerations in Educational Planning: The Cost and Effectiveness of School-Based Interventions, *Food and Nutrition Bulletin*, 12: 204-215; Bhargava, Alok (2001), Nutrition, Health, and Economic Development: Some Policy Priorities, World Health Organization, Commission on Macroeconomics and Health, *CMH Working Paper Series*, Paper No. WG1: 14. Available at http://www.cmhealth.org/docs/wg1_paper14.pdf

impedes their ability to do manual labor, often lose what may be their only major resource. Healthier communities, too, mean that workers take less time off to care for sick relatives.

Third, good health promotes saving and investment, which are important drivers of economic growth. As life expectancy increases, saving for retirement becomes a more rational decision for individuals.[13] Retirement savings translate into funds available for investment. In addition, healthy, productive workforces are more likely to attract foreign investment than sick populations.

Fourth, health improvements will have both transitional and permanent effects on a country's population age structure, with potentially huge impacts on economic development. As health advances begin to reduce infant and child mortality, the number of children surviving to adulthood increases. Either rapidly or over a longer period of time, parents realize they need to bear fewer offspring in order to attain their desired family size, thus prompting a decline in fertility rates. This effect is typically buttressed by rising levels of female education and labor market opportunities that reduce desired fertility. Fertility levels also fall in response to the provision of family planning services,[14] and knowledge about family planning options may increase with access to cultural influences from outside the local community, such as relevant public service announcements[15] and docudramas[16] on television. The cohort of children born after mortality rates fall and before fertility rates decline is larger than both the preceding and succeeding cohorts – often known as a 'baby boom' generation. A baby boom cohort has the potential, when it reaches adulthood, to swell the workforce, and provide a boost to a nation's economic produc-

[13] Bloom, David, David Canning, and Bryan Graham (2003), Longevity and Life Cycle Savings, *Scandinavian Journal of Economics*, 2003 (Vol. 105, No. 3): 319-338.

[14] Schultz, T. Paul (1997), The Demand for Children in Low Income Countries, *Handbook of Population and Family Economics*, Vol. 1A, Chapter 8, (eds.) M.R. Rosenzweig and O. Stark, Amsterdam: North-Holland Publishing.

[15] Hegazi, Sahar, Suzan Kelini, and Gihan Rashti (2002), Family planning TV spots in Egypt: Did they change knowledge, attitudes and practices of women in slum areas in Cairo?, Abstract #38946 of November 11 presentation at annual meeting of American Public Health Association. Available at: http://apha.confex.com/apha/130am/techprogram/paper_38946.htm

[16] Relevant experience in Bangladesh is documented at http://www.jhuccp.org/pubs/ci/7/index.shtml

tivity.[17] The aforementioned fertility decline implies these economic gains will not be encumbered by having to support a large number of children.

East Asia appears to have benefited greatly from the effects of its baby boom. Mortality in the region declined sharply during the period of the baby boom, likely the result of improvements in sanitation and the introduction of antibiotics and DDT.[18] Eventually, fertility rates followed and when the baby boomers reached working age, aided by strong education systems and flexible labor market policies that encouraged job creation, they gave a huge boost to East Asian economies. It has been estimated that this 'demographic dividend' accounted for one third of the region's spectacular economic growth between 1965 and 1990.[19]

The links between health and wealth are often mutually reinforcing. In East Asia for example, a virtuous spiral was created whereby health improvements catalyzed the baby boom generation and boosted economic

[17] The economic boost that demographic change can thus provide depends on a country's ability to employ its large cohort of working-age people. Policies that lead to such employment are therefore crucial. Numerous countries have failed to capitalize on the potential dividend that demographic change made available. And since this dividend most often arises from declines in mortality and fertility, it creates opportunities during a specific period of time for economic growth in a given population. Whether these opportunities spill over into an increased economic growth rate in the longer term is unknown.

[18] In the West, however, the decline in mortality from infectious diseases predated the introduction of antibiotics and vaccines. See T. McKeown, *The Role of Medicine: Dream, Mirage, or Nemesis*, Princeton: Princeton University Press, 1980, or Armstrong G.L., Conn L.A., Pinner R.W., Trends in infectious disease mortality in the United States during the 20th century, *Journal of the American Medical Association*, 1999 Jan 6;281(1):61-6. For further discussion of this issue, see David M. Cutler, Angus S. Deaton, Adriana Lleras-Muney, The Determinants of Mortality, *NBER Working Paper No. 11963*, January 2006, which finds that the application of scientific advances is 'the ultimate determinant of health'. In explaining mortality differences across countries, Samuel Preston stresses the importance of 'medical and public health factors' in The Changing Relation between Mortality and Level of Economic Development, *Population Studies*, Vol. 29, No. 2. (Jul., 1975), pp. 231-248. See also, Easterlin, R. (1999), How beneficent is the market? A look at the modern history of mortality, *European Review of Economic History*, 3(3): 257-294, for research that finds that the determinants of mortality decline involve the advent of new institutions and policy initiatives that build on an understanding of disease.

[19] Bloom, David E. and Jeffrey Williamson (1998), Demographic Transitions and Economic Miracles in Emerging Asia, *World Bank Economic Review*; Bloom, David E., David Canning, and Pia Malaney (2000), Demographic Change and Economic Growth in Asia, *Population and Development Review*, 26: 257-90; Bloom, David E., David Canning, and Jaypee Sevilla (2002), *The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change*, Santa Monica, California: RAND, MR-1274.

growth. Some of the fruits of this growth were then reinvested in health care, family planning, and education, which further reduced fertility rates and lightened the burden the workforce had to support.[20] Vicious spirals are also possible. In Sub-Saharan Africa, the infectious disease burden has continued to undermine children's prospects of surviving to adulthood. Fertility rates have therefore remained high and, instead of concentrating their resources on one or two children, parents have had to spread their investments in nutrition, health, and education thinly, thus limiting their potential for lifting the family out of poverty. The loss of workdays and selling of assets to pay for health care not only impoverishes poor families further, but also makes it more difficult to cope with future health shocks. Ruger et al.[21] detail, for example, the high cost of outpatient visits by the poor in China, and cite Liu *et al*.[22] to the effect that 'high health expenditures were a major cause of poverty in rural areas [in China]'.

Finally, and turning specifically to recent epidemics: Health scares may deter both investors and tourists from a country, as the recent SARS epidemic in China and Southeast Asia demonstrated. Foreign direct investment (FDI) in Hong Kong at the height of the crisis fell by 62% in one quarter.[23] These observations jibe with the finding of Alsan, Bloom, and Canning (2006) that FDI into low- and middle-income countries grows by 9% for every one-year increase in life expectancy.[24]

Health, then, can have strong effects on economic growth. Bloom, Canning, and Sevilla (2004) have found that each additional year of life

[20] Bloom, David E. and David Canning (2001), Demographic Change and Economic Growth: The Role of Cumulative Causality. In: Nancy Birdsall, Allen C. Kelley, and Stephen Sinding (eds.), *Population Matters: Demographic Change, Economic Growth, and Poverty in the Developing World*, New York: Oxford University Press, 2001.

[21] Ruger, Jennifer Prah, Dean Jamison, David Bloom, and David Canning (2006), Health and the Economy. In: Michael H. Merson, Robert E. Black, and Anne J. Mills (eds.), *International Public Health: Diseases, Programs, Systems and Policies*, Sudbury, MA: Jones and Bartlett Publishers, 601-647.

[22] Liu, Y., S. Hu, W. Fu, and W.C. Hsiao (1998), Is community financing necessary and feasible for rural China?. In: M.L. Barer, T.E. Getzen, and G.L.Stoddart (eds.), *Health, health care and health economics: Perspectives on distribution*, New York: John Wiley & Sons.

[23] Tam, J. (2003), SARS slashes FDI inflows by 62 per cent, *The Standard: Greater China's Business Newspaper*, 1 October.

[24] Alsan, Marcella, David E. Bloom, and David Canning (2006), The Effect of Population Health on Foreign Direct Investment Inflows to Low- and Middle-Income Countries, *World Development*, April, 2006 (Vol. 34, No. 4), forthcoming.

expectancy raises per capita GDP by 4%.[25] GDP per capita does not provide a full picture of the economic impacts of improved health, however, since health improvements, by reducing mortality, boost population size as well as productivity. Health reversals, conversely, reduce population size. When the value of the extra lives that result from health improvements is taken into account, the effect of health improvements on economies will be much greater than the effect on per capita GDP alone.[26] In addition, GDP per capita, since it is an average, does not account for the greater economic benefits that may accrue to the poor, as compared to the rich, when population health improves.

The foregoing arguments are germane to showing a causal link from health to prosperity, but much research remains to be done. Efforts to assess and quantify these links are severely hampered by lack of data (in the cases of plague and influenza) or by reverse causality between GDP and health (in the case of HIV/AIDS).[27] Many studies of the economic impacts of epidemics, such as that of Brainerd and Siegler (2003) and Bloom and Mahal (1997a), focus on GDP per capita effects, ignoring the potential for duplicating the longer-term but very significant changes in education, fertility, and savings rates that may have resulted from the influenza epidemic.[28] In addition, as the SARS case vividly illustrates, expectations have a very significant role (more than ever), which can lead to economic effects vastly out of proportion to the number of infected people. This suggests that one take care not to extrapolate economic effects on the basis of earlier epidemics, which arose in a far less integrated world. Expectations, however, are only one factor in retrospectively evaluating the response to a budding epidemic. In the face of uncertainty about an agent's contagiousness, how fatal it is likely to be, and how it is transmitted, caution makes sense.

[25] In principle, this effect is net of the reverse causal effect that runs from higher income to life expectancy. See Bloom, David E., David Canning and Jaypee Sevilla (2004), The Effect of Health on Economic Growth: A Production Function Approach, *World Development*, January, (Vol. 32), pp. 1-24.

[26] Bloom, David E., David Canning and Dean T. Jamison (2004), Health, Wealth, and Welfare, *Finance and Development*, March 2004, Vol. 41, No. 1, pp. 10-15.

[27] Bloom and Mahal (1997a), *op. cit*.

[28] See, for example, Andrew Noymer and Michel Garenne, The 1918 Influenza Epidemic's Effects on Sex Differentials in Mortality in the United States, *Population and Development Review*, Volume 26, No. 3, 2000, for a discussion of the demographic effects of the epidemic.

In short, the links between health and wealth change over time and vary in different contexts – the effects of one on the other are not inevitable and are often unpredictable. Health is, however, clearly an intrinsic part of the development process. A country's wealth (and indeed its health) can often be enhanced by traditional measures such as opening up to trade, promoting exports, restructuring or eliminating inefficient state-owned enterprises, improving infrastructure, and investing in education. It can also be enhanced by investments in health. Specific, cost-effective entry points such as efforts to reduce infant mortality via vaccines or to improve maternal health via family planning programs can, by changing a country's age structure and increasing labor productivity, have resounding long-term economic effects. Neglect of such issues, on the other hand, can contribute to economic stagnation or impoverishment.

Epidemics, Pandemics, and Economic Prosperity

Epidemic: 'The occurrence in a community or region of cases of an illness, specific health-related behaviour, or other health-related events clearly in excess of normal expectancy. The community or region and the period in which the cases occur are specified precisely. The number of cases indicating the presence of an epidemic varies according to the agent, size, and type of population exposed; previous experience or lack of exposure to the disease; and time and place of occurrence … Generally, a disease that exhibits large inter-annual variability can be considered as epidemic' (World Health Organization).[29]

Pandemic: 'An epidemic occurring over a very wide area (several countries or continents) and usually affecting a large proportion of the population' (US Centers for Disease Control and Prevention).[30]

Endemic: 'The constant presence of a disease or infectious agent within a given geographic area or population group; may also refer to the usual prevalence of a given disease within such area or group' (World Health Organization).[31]

The subject of this paper is epidemics and their relationship to economic outcomes. Since pandemics are simply epidemics spread over a

[29] Adapted by WHO from Last J.M., *A Dictionary of Epidemiology*, 2001.
[30] http://www.cdc.gov/excite/library/glossary.htm
[31] Adapted by WHO from Last J.M., *A Dictionary of Epidemiology*, 2001.

wide area, a discussion of the two naturally goes hand in hand. Endemicity, however, presents a slightly different situation. First, some diseases that are endemic (such as malaria) periodically cause epidemics, as they break out into, or become more prevalent in, a specific population. AIDS and tuberculosis are also endemic in many areas. Each of these diseases, in part by virtue of its endemicity, brings with it economic effects that are likely to be long term and that have been the subject of much consideration. Epidemics are of a more well-defined time and geographic scope. Their economic effects are more likely to be short term, although they can also have long-term economic consequences.

Assessing whether the linkages between epidemics and wealth are similar to those between general health and wealth is complex. This section will ask whether wealth protects against epidemics, and then whether epidemics affect wealth.

Epidemics – and we limit our discussion here to epidemics of communicable diseases spread by long-known pathogens – are most likely to arise and persist under conditions commonly created by poverty.[32,33] First, the spread of infectious diseases often requires close and frequent contact between individuals. This, in turn, requires either or both of crowded living conditions and a high level of (preferably rapid) mobility. (Recall Charles Dickens' graphic descriptions of the overcrowded and poorly ventilated housing of the poor and how those conditions abetted the spread of tuberculosis). Second, poor sanitation and hygiene allow bacteria, viruses, parasites (e.g., worms and amoebae), and the vectors of transmission to thrive. Third, weak bodies are more easily infected and less able to fight infection;[34]

[32] The spread of the SARS virus in Asia is an example.

[33] As with epidemic diseases, endemic diseases are also often abetted by poverty conditions. Examples include malaria, tuberculosis, schistosomiasis, dengue fever, and hepatitis B and C.

[34] There is evidence that better nutritional status helps to prevent infection and more evidence that it helps in fighting infections. See, for example, Scrimshaw N.S., SanGiovanni J.P., Synergism of nutrition, infection, and immunity: an overview, *Am. J. Clin. Nutr.*, 1997 Aug. 66(2):464S-477S; Bhutta Z.A., Bird S.M., Black R.E., Brown K.H., Gardner J.M., Hidayat A., Khatun F., Martorell R., Ninh N.X., Penny M.E., Rosado J.L., Roy S.K., Ruel M., Sazawal S., Shankar A., Therapeutic effects of oral zinc in acute and persistent diarrhea in children in developing countries: pooled analysis of randomized controlled trials, *Am. J. Clin. Nutr.*, 2000 Dec;72(6):1516-22; Fawzi W., Msamanga G., Spiegelman D., Hunter D.J., Studies of Vitamins and Minerals and HIV Transmission and Disease Progression, *J. Nutr.*, 2005 Apr. 135(4):938-44; Villamor E., Fawzi W.W,. Vitamin A supplementation: implications for morbidity and mortality in children, *J. Infect. Dis.*, 2000 Sep. 182 Suppl 1:S122-33.

where populations are malnourished, left weak by other health setbacks, or have a high proportion of very young or very old members, epidemics have the potential to thrive. Fourth, epidemics tend to occur where health systems are weakest and therefore incapable of detecting and responding to rapidly evolving health threats. Finally, poverty conditions can lead individuals to engage in behaviors that facilitate disease transmission (e.g., sex workers, or poor Chinese farmers who sell blood in Henan province, who contract HIV).

Wealth enables people to safeguard themselves against or mitigate the effects of many of these risk factors. The wealthy generally have less crowded living space than the poor, greater access to health care, drugs, and vaccines (and where state provision is weak, they can better afford private care), better sanitation, and better nutrition. Economic decline, moreover, often triggers the spread of disease. In Russia in the 1990s, for example, political and economic stress had severe effects on health systems. Diphtheria thrived, and there was a resurgence of tuberculosis and measles. As Figure 2 (see page 419) shows, the share of deaths from infectious diseases is much higher in poor countries than in rich ones.[35]

The issue of mobility, however, is more complicated. The rich may be more mobile than the poor, and tourism and business travel can lead to greater exposure to disease (trade has been a conduit for epidemics since ancient times, and globalization allows disease-causing microbes to spread faster and further than previously). The mobility of the poor, on the other hand, is likely to be more risky. Migration, for example, is a risk factor for HIV transmission, as male migrants tend to have social anonymity and cash, and may be lonely when separated from their families, leading them to seek commercial sex; female migrants are often abused.

More generally, poor people often move to escape catastrophic events such as war or environmental disasters, frequently ending up in crowded, makeshift camps or slums where poor sanitary conditions create an ideal environment for epidemics such as typhus and cholera. They may also move to find work. Much of this movement is from rural areas into cities,

[35] WHO supplies detailed data online on cause of death for each of 14 region/mortality strata. Infectious diseases are the cause of many deaths. To this number we have added respiratory infections, which are not included in the former category. For each region, this sum, divided by total deaths, gives the fraction that appears on the vertical axis. The source of the data for the horizontal axis is World Bank, *World Development Indicators 2004*, online.

and occurs when cities begin to prosper. Friedrich Engels observed the perils of urbanization in the 19th century: ‘Dirty habits’, he wrote, ‘do no great harm in the countryside where the population is scattered. On the other hand, the dangerous situation which develops when such habits are practiced among the crowded population of big cities must arouse feelings of apprehension and disgust’.[36] As well as offering some protection against epidemics, the changes in living habits (specifically, movement from rural to urban areas) that accompany increased incomes can also make epidemics more likely to occur.

The effect of an epidemic disease on the economy is complex and depends upon many factors. These include what group of individuals is most at risk for contracting the disease, the natural history of the illness (e.g., how long the epidemic lasts), and how the disease is transmitted (via airborne vs. blood-borne pathogen).

Annual epidemics often reserve their harshest impacts for youth and the elderly (although this is not necessarily so for pandemics).[37] By reducing dependency ratios, this pattern can have positive impacts on economies – with fewer non-earners and a reduced number of dependents for those who are earning, per capita income is likely to rise (although it would be difficult to argue that these income gains offset the welfare costs of premature deaths associated with the flu). Of course, when a depleted youth generation enters the workforce, the reduced number of workers may itself become an economic problem.

Given the vast trail of death and morbidity left by epidemics, they obviously diminish human well-being. Even where they affect people of working age, however, epidemics may not have negative economic impacts on income per capita. The Black Death, which wiped out an estimated quarter of Western Europe’s population in the 14th century, is thought by some to have had a positive effect on incomes,[38] (which could have resulted from a reduced number of individuals benefiting from and building on the same set of fixed assets, including, for example, land and

[36] Friedrich Engels (1845), *The Condition of the Working-Class in England*, Leipzig: new edition, Stuttgart, 1892.

[37] See Simonsen L., Clarke M.J., Schonberger L.B., Arden N.H., Cox N.J., Fukuda K., Pandemic versus epidemic influenza mortality: a pattern of changing age distribution, *Journal of Infectious Diseases*, 1998 Jul;178(1):53-60.

[38] Clive Bell and Maureen Lewis (2004), The economic implications of epidemics old and new, *World Economics*, Vol. 5, No. 4, October-December.

physical capital). Others have concluded that it had little or no effect on incomes (at least in England and France).[39] The 1918 flu epidemic killed over 40 million people worldwide, but its economic effect in the US was, if anything, positive.[40] By contrast, Bloom and Mahal found no evidence of a significant effect from that epidemic on output (i.e., acreage sown) per capita in India.[41] When there are deaths, of course, there are welfare losses. But these are cases where, although the economy may have been disrupted, there was no long-term negative effect on per-capita income. These results highlight the fact that many economies are flexible and resilient, and that people adapt to unforeseen changes. Finally, even though these epidemics caused many deaths, some of them came and went quickly, with the possible consequence that the economic dislocations that arise from morbidity, treatment, and transitional social arrangements did not last long enough to engender a long-term effect, at least with respect to the measures used in this literature.

Writers on this topic, however, are most often looking only at the economic well-being of those who survive. In a high unemployment situation, total output may not change, since new workers take the place of those who die (in some instances, however, as when skilled workers die, there may be clear and negative financial repercussions and output may fall). Perhaps most typically, GDP per capita may rise, at least initially, when some people die. Longer-term equilibration may minimize this effect, however, and overall it is not clear whether GDP per capita is greatly affected either positively or negatively.[42] Recent research by Mahal,

[39] Bloom, David E. and Ajay Mahal (1997a), AIDS, Flu, and the Black Death: Impacts on Economic Growth and Well-Being. In: David Bloom and Peter Godwin (eds.), *The Economics of HIV and AIDS: The Case of South and South East Asia*, New Delhi: Oxford University Press, 1997, 22-52.

[40] Brainerd E. and M. Siegler (2003), The Economic Effects of the 1918 Influenza Epidemic, No 3791, *CEPR Discussion Papers*. The authors note that a disproportionate share of those who died were of working age. They then 'examine the impact of this exogenous shock on subsequent economic growth using data on US states for the 1919-30 period. Controlling for numerous factors including initial income, density, urbanization, human capital, climate, the sectoral composition of output, geography, and the legacy of slavery, the results indicate a large and robust positive effect of the influenza epidemic on per capita income growth across states during the 1920s'.

[41] Bloom and Mahal (1997a), *op. cit.*

[42] For example, notwithstanding the mixed evidence on the long-term effect of the Black Death on wages, it is clear from the diaries of plantation owners that the very short run effect was a substantial elevation of wages. See Hirshleifer, Jack (1987), *Economic Behaviour in Adversity*, Chicago: The University of Chicago Press.

however, has uncovered signs that as the AIDS epidemic grows, there may be negative effects on a country's GDP per capita.[43]

If we take the economic value of human life into account, however, it is clear that epidemics have a high cost. How high, of course, depends on the economic value one assigns to life. People who die during epidemics, along with their families, obviously lose out economically. One could, moreover, add other losses to this if one places an economic value on intangible impacts such as the lost companionship and love felt by friends and relatives of the dead. A thorough discussion of valuing the cost of life appears in Viscusi and Aldy (2003).[44]

In recent years, some epidemics have had clear negative economic effects. The HIV/AIDS pandemic too, while not (yet) significantly affecting per capita GDP, has had major effects at the household level, particularly on the poor.[45,46] As well as suffering from lost earnings, poor families hit by AIDS have to draw down savings and sell assets to pay for treatment. Relatives of infected individuals may have to take time off work or withdraw from school to provide care. A family's future economic prospects are thereby further impaired; indeed, even the immediate spending can be catastrophic to the individual family. The long-term costs of HIV/AIDS, and in particular the vast scale on which social and economic effects are likely to be felt (because of decreased investment in human capital), are laid out in Bell, Devarajan, Gersbach (2003).[47] In addition, the long latency period of the HIV virus means that the long-

[43] Mahal, A. (2004), The economic implications of inertia on HIV/AIDS and the benefits of action, *Economic and Political Weekly*, 39(10): pp.1049-63.

[44] Viscusi, W. Kip, and Joseph E. Aldy (2003), The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World, *Journal of Risk and Uncertainty*, 27(1): 5-76.

[45] Bloom, David and Ajay Mahal (1997b), Does the AIDS Epidemic Threaten Economic Growth?, *Journal of Econometrics*, 105-124; Bloom, David E., Ajay Mahal, Larry Rosenberg, Jaypee Sevilla, David Steven, and Mark Weston (2004), *Asia's Economies and the Challenge of AIDS*, Asian Development Bank.

[46] Although the poor have in general been the most affected segment of society, GDP can still remain only minimally affected. This is because the total economic output of the poor is in many instances sufficiently small that decreased output by the poor has only a small effect on total GDP. In countries in which the poor contribute a sizable share of GDP, this explanation suffices only if a small percentage of the poor are affected.

[47] Bell, C., S. Devarajan, and H. Gersbach (2003), *Thinking about the Long-run Economic Costs of AIDS*, World Bank (October) (Draft). Available at http://emlab.berkeley.edu/users/webfac/bardhan/e271_f03/thinking.pdf

term course of the disease is difficult to predict, as are, therefore, the long-term economic effects.

AIDS may also affect businesses. In a global survey of 8719 firms conducted by the World Economic Forum in 2004,[48] business leaders expressed some concern over the current and future impacts of the pandemic on their workforces and other aspects of company operations.[49] 16% of respondents expected the virus to have serious impacts on their business. In low-income countries this figure rose to 35%, and in Sub-Saharan Africa to 45%. Although few firms reported current negative impacts on productivity, revenues, and medical expenses, in low-income countries over 1 in 10 respondents had already experienced serious impacts on each of these aspects of their business.

AIDS has, in recent years, begun to claim the lives of large numbers of people, especially in countries that were hit early by the virus. It is possible – some would say quite likely – that the economic effects will grow over time, as dwindling working-age populations (which are disproportionately affected by the disease) have to support burgeoning numbers of children and orphans and social breakdown becomes a reality. Education systems have already begun to suffer because of deaths among teachers. Orphans, in particular, are subject to becoming a particularly uneducated segment of the next generation. Weakened education systems that in many instances are already inadequate are likely to impair countries' efforts to strengthen their economies.

In the past decade, Severe Acute Respiratory Syndrome (SARS) and Bovine Spongiform Encephalitis (BSE)[50] have also had economic impacts. The panic caused by these diseases led to economic costs that were disproportionate to the number of cases. Although fewer than a thousand people died in the 2003 SARS outbreak, for example, tourism and business in affected areas were hit hard, with some estimates putting the cost of the virus at over $11 billion.[51] The European Association of

[48] World Economic Forum Global Health Initiative (2005), *Business and HIV/AIDS: Commitment and Action?*, http://www.weforum.org/pdf/Initiatives/GHI_Report_2005_Final.pdf

[49] For a review of the previous year's report, see David E. Bloom, Lakshmi Reddy Bloom, David Steven, and Mark Weston (2004), *Business and HIV/AIDS: Who Me?*, World Economic Forum, Geneva.

[50] BSE is epizootic (i.e., it affects many animals in a certain region and time period). But its much less prevalent human variant, Creutzfeldt-Jakob disease (CJD), has provoked widespread fear.

[51] Trish Saywell, Geoffrey A. Fowler, Shawn W. Crispin (2003), The Cost of SARS: $11 Billion and Rising, *Far Eastern Economic Review*, 24 April.

Animal Production, meanwhile, has estimated that BSE, which has caused fewer than 200 deaths, has so far cost the 15 European Union states at least $110 billion.[52,53]

It appears, then, that the relationship between epidemics and wealth has many similarities to that between overall health and wealth. Economic growth is likely to protect populations against health setbacks and against epidemics (although the examples of BSE and SARS show that even the richest countries cannot afford to be complacent). Both epidemics and other health problems, moreover, can impede economic development and set off vicious spirals whereby poor health leads to reduced wealth and makes the task of improving health ever harder. It is possible, however, that only in cases where epidemics are sustained over a long time period will the effects on survivors prove harmful at a macroeconomic level (although those who die sustain clear economic losses, along with their families). Short epidemics may have negative effects on individual households and firms, but the overall damage to economies will be less severe.

It is relevant to note here that, as with epidemics, endemic diseases can also have major economic consequences. On the individual and family level, for example, it is clear that malaria's effects can be economically devastating. And at the macro level, according to the report of the Commission on Macroeconomics and Health, high malaria prevalence in an area can reduce economic growth by over 1 percentage point per annum.[54]

Given time to accumulate, as in the cases of malaria and HIV/AIDS, micro-level effects may have macro-level consequences. The exceptions to this are short epidemics that cause large panics. SARS, for example, wreaked economic havoc not because of the number of cases of the disease but because of the alarm it provoked and uncertainty about its ease and routes of transmission. In addition, the chaos caused by SARS was as severe in Toronto as it was in Beijing and Bangkok, despite the fact that Toronto had significantly fewer cases.

[52] http://europa.eu.int/comm/research/press/2004/pr1805en.cfm

[53] $110 billion is roughly 1% of the GDP of the EU. But the $110 billion loss, which may not have all occurred during one year, is a *capital* loss, not an *income* loss. The share of income lost would be much tinier.

[54] *Commission on Macroeconomics and Health* (2001).

Part 2: The Challenge of Epidemics

Epidemics have afflicted human civilizations for ages. The challenge they present has not always been met, and many millions have died as epidemics have ravaged populations with no effective means of resistance. Although better sanitation, rapid response measures, and specific medical advances have given humanity new tools to resist epidemics, the forces of 'globalization' have, at the same time, abetted the spread of epidemics. For example, more frequent international travel facilitates the access of many pathogens to new populations. Similarly, increased international trade during the last couple of decades has meant that disease-causing agents from one area can more easily find their way to regions that were previously thought immune to a particular disease. Some have argued that competitive pressures and freer trade have led to export pressures that abet the spread of infections, as seems to have been true with Bovine Spongiform Encephalitis, the Nipah virus, and avian flu. Other twentieth-century developments are also relevant here: massive movement from rural areas to cities and the accompanying poverty in peri-urban areas makes epidemics more likely. Global climate change, too, may be affecting the spread of malaria, dengue fever, and yellow fever. Higher-temperature ocean currents have been linked to increased cholera rates.[55]

One of the sources of the panic created by some epidemics (but not all of them) is their unpredictability, which flows in part from some of the new factors cited above. This applies as much to their economic impacts as to their likely transmission speed and direction. As the examples of HIV/AIDS and SARS show, different epidemics can have widely differing economic effects.

This presents policymakers with two main challenges. First, their response needs to be swift. Rapid action taken by the global health community was instrumental in limiting the impacts of SARS. Had the response been slower and the disease allowed to spread further, its economic effects would have been even more catastrophic. In contrast, the sluggish speed of the reaction to HIV/AIDS in many parts of the world has greatly increased the virus's impacts. Second, the response needs to be flexible. Epidemics can change quickly, forcing decision-makers to act without complete information. The initial effects of HIV/AIDS, for example, were felt by gay communities and injecting drug users in the West.

[55] Colwell, R.R. (1996), Global Climate and Infectious Disease: The cholera paradigm, *Science*, 274:2025-2031.

Policy-makers were caught off guard when they began to realize that it was spreading rapidly among heterosexuals in Africa.

This need for flexibility in policy response applies, of course, to resource allocation. The costs of less-sudden health problems, such as chronic disease or mental illness, are relatively easy to forecast and budget for. The unpredictability of epidemics, however, makes planning for the effective allocation of human and physical resources difficult. Diverting funds to counter epidemics may, for instance, temporarily halt progress made against non-communicable diseases. It may also require health workers to take on new responsibilities for which they are not fully equipped. The costs of an epidemic, moreover, may change suddenly as it spreads or is contained. Avoiding waste while ensuring resources are sufficient to tackle a disease is a difficult balancing act, particularly for policymakers who have not previously been confronted with such a threat. This challenge is all the more difficult because anticipatory investment is required.

Another of the unique challenges posed by epidemics is the requirement they impose for people to limit their movement. As long ago as the 14th century, villages afflicted by the plague were cordoned off from the rest of the society to limit the disease's spread. In the more recent case of SARS, travelers were strongly advised by the World Health Organization not to visit high-risk areas such as southern China, Hong Kong, and even Toronto. Restricting movement, particularly in a globalizing world, has serious negative implications for tourism and sometimes for trade, as well. These impacts will likely be felt more strongly in industrializing or industrialized countries, where the requirement for movement is greater than in predominantly agricultural economies.

Epidemics that cross national borders and become pandemics pose further taxing challenges and mean that few countries can afford to drop their guard. Dengue fever, for example, had for long been confined to tropical regions until spreading into temperate zones in recent years, largely as a result of the increase in air travel and the shipping of used tires harboring dengue-carrying *Aedes* mosquitoes.[56] For epidemics not to

[56] It is sometimes suggested that the resurgence of tuberculosis in New York City in the early 1990s, was due to, among other factors, immigrants importing drug-resistant forms of the bacillus into the city. This, however, appears to be an urban myth, as the primary causes seem to have been the spread of HIV – TB is an opportunistic infection – and prison crowding stemming from the drug war of the1980s. New York's public health system had become unused to dealing with the disease and was thus at first unable to cope effectively with the new threat.

become pandemics, countries need to be constantly alert to their occurrence in neighboring and even far-off societies, and aware of the risks to which their links with affected countries expose them.

Perhaps the biggest test posed by epidemics is the propensity of the underlying disease agent to mutate. Success against one strain of a disease (either by natural resistance or medical intervention) can open a niche for the emergence of a new strain, and new strains are likely to be more virulent. The emergence of new epidemics – in the past 15 years alone, hantavirus, SARS, and avian flu have hit the international headlines – and the resurgence of old ones demands great agility and ingenuity from medical practitioners, drug developers, public health professionals, and policymakers.

The best way of countering drug-resistance, of course, is to prevent it, and the best way of preventing it is to halt the spread of disease in the first place through the use of vaccines, bed nets, elimination of vectors, educational programs, etc. Resistance occurs through the use of antimicrobials. Overuse is common in industrialized countries, where patients are inappropriately prescribed antimicrobials for viral infections (e.g., the common cold) or when antibiotics that are used by people are also employed in bulk as growth-promoters in livestock, which is common in OECD countries. Misuse may occur when patients lack access to appropriate therapy, but also when the general public is misinformed and individuals successfully demand unnecessary antibiotics. Finally, underuse is typically encountered among the poor, when inadequate financial resources impede the completion or continuation of treatment regimens. With diseases such as tuberculosis and HIV/AIDS, where drugs are now available, it is better – from a population health point of view – to have no treatment at all than 'bad' treatment.

While reacting speedily to epidemics is important, undue haste to push people onto underfunded treatment regimens may have dire unintended consequences. Recent research on effective strategies for using anti-retrovirals in treating HIV epidemics shows that determining the optimal use of such drugs is extremely complex.[57] The WHO has made

[57] Blower, Sally, Li Ma, Paul Farmer, Serena Koenig (2003), Predicting the Impact of Antiretrovirals in Resource-Poor Settings: Preventing HIV Infections whilst Controlling Drug Resistance, *Current Drug Targets – Infectious Disorders*, 3, 345-353.

significant efforts to summarize the world's response to increasing drug resistance and has highlighted the steps needed to counteract it.[58]

There are two forms of resistance: primary and secondary. Primary resistance occurs when a person is infected with a drug-resistant strain harbored by another individual. Thus, even when antimicrobials are not available in a given community, resistant strains can enter the population through trade and tourism. Secondary resistance develops when an individual's treatment is inadequate to achieve complete cure or suppression of the infectious agent and mutant strains develop that are not susceptible to the previous therapy. New models predict that drug-resistant strains of TB will soon come to dominate the drug-sensitive type.[59] This should alarm policymakers in wealthy and poor countries alike, given the ease of transmissibility of TB and the woeful lack of novel therapeutics to tackle this ancient microbe.

Thus, with international borders becoming ever more porous and the ability to treat infectious disease threats now within reach, it is essential for the international community to provide appropriate and sustainable therapy to those seeking treatment before resistance becomes unmanageable.

The recent emergence of Avian, H5N1 flu on the world scene has drawn renewed attention to epidemics and is an example of the challenges posed by new pathogens. Although flus in birds do not usually infect humans, they can do so on occasion, most likely by contact with infected poultry. The first observed case in humans took place in Hong Kong in 1997. During 2003 and 2004, humans contracted bird flu in Thailand and Vietnam. The death rate of infected people in the most recent set of reported cases is around 70%. If the virus mutates and human-to-human transmission becomes possible, a worldwide avian flu pandemic could occur.[60] Alternatively a trading of genetic material between the avian and human forms of the virus could lead to an epidemic. (Such trading is particularly likely to occur in pigs, which can host both the avian and human forms of the virus. The most virulent form of the avian flu virus, – (A) H5N1 – is particularly adept at obtaining genet-

[58] World Health Organization, 2001, *WHO Global Strategy for Containment of Antimicrobial Resistance*, Geneva. Available at http://www.who.int/emc/amrpdfs/WHO_Global_Strategy_English.pdf

[59] Cohen, Ted, and Megan Murray(2004), Modeling epidemics of multi-drug resistant M. tuberculosis of heterogeneous fitness, *Nature Medicine*, 10:10, October.

[60] Website of the United States Centers for Disease Control, http://www.cdc.gov/flu/avian/gen-info/facts.htm

ic material from other viruses, a process known as 'reassortment'. Such trading could lead to much easier human-to-human transmission).[61,62] A virus that resembles the one that caused the 1918 flu epidemic could 'kill 175 to 350 million people',[63] and some estimates go higher. (On the other hand, these very high estimates do not take into account the possibility that the death rates reported so far may be higher than the actual mortality rate, since people who contracted mild cases may not be included in the data).[64] A recent and careful estimate of the possible human and economic consequences suggests that, in a worst-case scenario, 142 million people could die and global economic losses could reach $4.4 trillion.[65] In any case, the economic and human consequences of such a pandemic are impossible to predict with any certainty, but the scale of the potential disaster warrants broad and serious international cooperation now. Measures to keep the disease from spreading from animals to humans are essential, as are preparations for containing it should it make further inroads among humans.

Part 3: Confronting the Challenges

Epidemic Prevention

Measures to confront epidemics can take the form of either prevention or treatment. Although it is impossible to prepare fully for diseases that do not yet exist or have been latent for many years, it is feasible to

[61] http://www.mayoclinic.com/invoke.cfm?id=DS00566

[62] A detailed explanation of the risks of an avian flu epidemic in humans appears in Ferguson N.M., Fraser C., Donnelly C.A., Ghani A.C., Anderson R.M., Public health risk from the avian H5N1 influenza epidemic, *Science*, 2004 May 14;304(5673):968-9

[63] Fedson, David S. (2005), Preparing for Pandemic Vaccination: An International Policy Agenda for Vaccine Development, *Journal of Public Health Policy*, 26, 429. This article reviews the many steps that will be needed to achieve international cooperation in development and testing of an effective vaccine for avian flu. It also points out the difficulties of ensuring that residents of countries that do not produce the vaccine will have access to it.

[64] Bradsher, Keith, and Lawrence K. Altman (2004), W.H.O. Official Says Deadly Pandemic Is Likely if the Asian Bird Flu Spreads Among People, *New York Times*, 30 Nov., page A12.

[65] Warwick J. McKibbin and Alexandra A. Sidorenko, *Global macroeconomic consequences of pandemic influenza*, Sydney: The Lowy Institute. 2006. Available at www.lowyinstitute.com//PublicationGet.asp?i=345

change behavior and strengthen health systems, sanitation, and people's bodies so that epidemics are both less likely and less damaging.

Similar to the great majority of microbes that affect human beings, many new viruses reaching humans (HIV/AIDS, hantavirus, and avian flu, for example) are thought to have originated in animals. Poor hygienic practices or behavior that brings humans into contact with animals can lead to epidemics. In other instances, a failure to adhere to drug regimens is believed to be at least partially responsible for the emergence of drug-resistant tuberculosis. Modifying human behavior to limit these risks is far from unachievable. The provision of high quality, convenient and *affordable* health services are essential aspects of an effective strategy to increase adherence at the population level. One such example is the successful delivery of antiretrovirals (ARVs) to treat HIV-infected persons in rural Haiti. Through a combination of community health workers, free antiretrovirals (ARVs) to treat HIV/AIDS, and an intensive information campaign, the Haiti project has served as a model for increasing adherence in resource-scarce settings.[66]

Closing the gap between health professionals and communities, so that relationships of trust are established and health messages are thereby more easily transmitted, is also important. A study by the Navrongo Health Research Centre in northern Ghana found that moving nurses into communities and mobilizing community volunteers to help them reduced mortality in an area by 30%. Part of this reduction was due to improved access to treatment, but much was also due to the behavior-related messages disseminated to communities by the nurses. The strategy has since been adopted as part of Ghana's national health care policy. Of course, such dissemination is only useful if the messages are adopted by the community.

Health systems' readiness for epidemics can also be improved. Reliable epidemiological surveillance is critical to identifying new health threats. Southern China's weak surveillance systems meant that the SARS virus had already begun to spread far beyond the original source area before policy-makers knew of its existence. The World Health Organization's global surveillance systems were effective in picking up new cases of the disease once it began to roam the world, but local monitoring capacity is more important if epidemics are to be nipped in the bud. It is in every country's interest for local disease surveillance in other countries to be effective – if your neighbor fails to spot a fledgling epidemic, you may well suffer

[66] Farmer, 2004.

(though not always, since some epidemics die out because of low transmissibility or other reasons). International efforts to improve surveillance are therefore needed, with countries that possess effective systems assisting those that lack them.

In the case of a new epidemic, especially when the transmission mechanism has not yet been identified, the first line of defense is isolation and quarantine. Even these measures, however, need preliminary research to establish the mode of transmission. Many of the most infectious and dangerous diseases are those that spread by airborne transmission. These may require stringent quarantine measures, whose effectiveness varies considerably from one situation to another.

Particularly important for readying health systems to manage epidemics is the focused allocation of resources. Poor countries, as we have seen, are most vulnerable to severe infectious disease outbreaks, and in these resource-constrained settings, it is vital to put epidemic preparedness at the center of health strategies. In societies where epidemics pose the main threat, directing resources towards preventing and coping with them is likely to have a greater impact on a population's health than dealing with chronic diseases or providing expensive surgical operations. Training health workers to respond to infectious diseases, for example, and ensuring immunization programs reach those – often the poor – who are most at risk, may prove a cost-effective approach to health care.[67] Simple, cheap, disease-prevention strategies such as the provision of insecticide-treated bed nets to combat malaria have also been proven to have impacts on health outcomes that go beyond the immediate users.[68] Where funds and human resources are limited, health strategists should not attempt to solve every problem, but should direct their efforts to areas where they can have the greatest effect on health outcomes. In some poor countries, this means epidemic prevention, although in most, endemic diseases will merit the greater share of attention.

[67] Even when some of the population cannot be reached by immunization programs, herd immunity has a positive spillover effect, as many non-immunized people are protected from disease because of those who are immunized.

[68] A study by the Navrongo Health Research Centre found that sleeping under bed nets soaked every six months in insecticides reduces the number of deaths in children below the age of five by 17%. It also found that even those who cannot afford nets may benefit, since the more insecticide-treated bed nets a community has, the fewer mosquitoes it harbors.

Health programs, such as for immunization, are also important for the other main strand of epidemic prevention: strengthening human bodies. Strong bodies are better able to fend off disease and better able to recover if an epidemic – whether old or new – does penetrate their defenses. The effectiveness of immunization programs depends critically on vaccines being available when they are needed – a condition that has not always been met. In the case of flu, for example, where the strains most prevalent in one year typically differ from those of previous years, manufacturing capacity is often not adequate to rapidly produce the required vaccines. Building up such capacity between epidemics is a crucial measure for preventing them. Such an investment in infrastructure (i.e., expanding and ensuring the reliability of the production line) can also be invaluable in containing an epidemic once it has started. For some diseases, however, such as malaria in Africa, the relatively low return on such investment elicits little vaccine supply. In such situations, it may be appropriate for the government to make or subsidize investments in vaccine infrastructure.

Other factors that help to strengthen bodies and thereby prevent epidemics will also likely require government intervention. Improving nutrition and providing a safe water supply are two examples, as is improving sanitation, which is important for depriving microbes of a favorable environment.

Treating Epidemics

Even if all these prevention methods are in place, however, some diseases will slip through the net. The effective global campaign to combat SARS provides several guidelines for those attempting to reverse epidemics that have evaded prevention efforts. The campaign, which was coordinated by the World Health Organization, attempted to disrupt the transmission pattern of the virus via a multi-pronged assault.[69] First, it limited movement of the virus by issuing a global travel warning against travel to and from infected areas. Second, it used the international media as a means of transmitting information about the virus and explaining preventive measures. Third, it mobilized the international scientific community to develop quarantine measures and identify the source of the

[69] Michael Merson (2003), SARS Proved Health is Global Public Good, *Yale Global*, 24 September.

virus – it took just one month to discover that a specific corona virus caused SARS, whereas the discovery that HIV caused AIDS took two years. Fourth, the WHO quickly strengthened its existing global surveillance system by working with immigration departments, airlines, and airports to track the spread of the virus.

The economic impacts of some of these measures offer an interesting lesson for future epidemic containment efforts. As Michael Merson notes, the extensive media coverage of SARS 'exaggerated public fear, feeding the growing stigma attached to the illness. Chinatowns were deserted, recovering SARS patients were turned down for jobs, and universities made it difficult for Asian students to attend commencement ceremonies. The economic impacts were immense'.[70] Although the travel warning and the use of the media were effective in highlighting awareness of the disease, their economic impact was disproportionate to its spread. Looking backwards, therefore, it may seem that the world overreacted. Nevertheless, when confronted with such situations, looking forward, and essentially always in the absence of complete information, it is difficult to argue that extreme precautionary measures should not be taken.

Modern technology, as the response to SARS showed, has great potential for combating epidemics. The WHO used a worldwide network of 11 laboratories to isolate the virus. The laboratories were connected by a shared website and held several telephone conferences each day. International cooperation to improve health is not new – the smallpox vaccine, developed in England in the late 18th century, was introduced to Latin America soon after by Spanish missionaries[71] – but new communication technologies have dramatically increased its scope. Developing countries now have expanded opportunities to learn from those with greater health care expertise, and international collaboration can more quickly spot emerging epidemics and more quickly mobilize to stifle them.

Mass mobilization is an important tool for reversing epidemics' spread. In Botswana, where over 100,000 people are in need of antiretroviral drug treatment for AIDS, the government committed to supplying treatment free of charge in January 2002. It has so far, however, been able to reach only a fraction of those in need, as the country lacks the medical staff to distribute the drugs and the stigma of HIV infection dissuades

[70] *Ibid*.

[71] William H. McNeill (1976), *Plagues and Peoples*, Monticello Editions, New York.

people from seeking treatment. In such settings, mobilizing non-medical personnel to provide treatment is likely to be the only way of reaching sufficient numbers of patients. Providing training in drug delivery to community health workers, traditional healers, and other prominent community figures, for example, may be a cost-effective way to deepen the pool of human resources. Many non-governmental organizations, too, have strong links with poor communities that governments struggle to reach, and businesses are often trusted sufficiently by their employees to transmit health messages and, perhaps, treatment.

Members of society should also be enlisted in the effort. Malcolm Gladwell, in *The Tipping Point*, writes of the role of 'connectors' in spreading social epidemics.[72] Connectors are community members whose large range of contacts puts them in a strong position to spark new trends or change behavior. With epidemics that require behavioral change, and particularly those such as HIV/AIDS whose containment depends on breaking stigma, enlisting connectors may be central to success. The stigma surrounding AIDS means many of its victims die rather than admit publicly to being infected. That stigma can be broken if influential individuals admit their infection and tell others it is nothing to be ashamed of. Nelson Mandela's acknowledgment that his son died of HIV/AIDS is an excellent example of this.

Once a disease has broken through prevention efforts and has become an epidemic, mass mobilization is essential to its containment. An epidemic's sudden, rapid spread is likely to be too much for resource-limited health systems to be able to cope with without wider societal support. As an epidemic crosses borders and becomes a pandemic, the international community is likely to awaken to the threat, but it is in the latter's interests to prevent pandemics emerging by helping vulnerable countries to prepare for and deal quickly with new infectious disease threats. International mobilization was instrumental in limiting the effects of SARS, but had that mobilization occurred while the disease was confined to southern China, its economic effects would have been drastically reduced.

[72] Malcolm Gladwell (2000), *The Tipping Point*, Little, Brown & Company.

Conclusion

> 'In view of globalization, high productivity is an essential ingredient for competitiveness. [Yet] a number of factors account for the loss of productivity, and one notable factor is the HIV/AIDS pandemic'.
>
> Levy Mwanawas (President of Zambia, March 31, 2004)

The links between epidemics and economics are broadly similar to those between health and wealth in general. Prosperous societies not only have better health; they are also at least somewhat protected against epidemics. Like other health problems, meanwhile, epidemics can hamper economic development and trigger vicious spirals whereby worsening health reduces wealth and diminishes the protection against further health threats. Recent epidemics of HIV/AIDS (and the linked epidemic of TB) and SARS have had strong impacts on the wealth of households, businesses and, in some cases, entire economies.

Epidemics, however, pose different challenges than other health problems. The speed and unpredictability of their spread mean that policymakers have to strike a fine balance between acting swiftly (and without complete information) and guarding against the undue haste that may make matters many times worse. Presenting facts in a clear and measured way and having the patience not to begin drug treatment programs until correct adherence is at least probable, for example, can help contain both the health and economic impacts of an epidemic.

Prevention of epidemics is also often feasible. Strengthening health systems and the populations they protect can help fend off disease outbreaks and limit the impacts of those that do emerge. Having strong quarantine measures ready to be employed to prevent an incipient epidemic, and preparation for the mass mobilization of civil society, business, and community members will both relieve the strain on health services caused by present epidemics and reinforce society's defenses against future outbreaks. However, a central obstacle to preventing epidemics is that spending on anti-epidemic and basic health infrastructure must be seen as an investment, with part of the return being the reduction in the future costs of epidemics. Policymakers are often influenced by short-term political considerations, causing them to pay too little heed to important longer-term realities.

International cooperation is also vital for preventing and containing epidemics. Epidemics can have economic impacts far beyond the borders of their source countries – and these potential economic impacts are

often amplified by those aspects of the modern world that facilitate disease spread, affect tourism and trade, disseminate awareness and fear, and exacerbate counter-productive forms of protective behavior and isolationism. Indeed, both the health and economic consequences of epidemics suggest the value of the international community shoring up the health systems of poor countries to strengthen their capacities to prevent and to respond to epidemics.[73] This, combined with their effects on health, makes combating epidemics a global public good. Globalization can facilitate the spread of new diseases, but it also offers opportunities for tackling them. International collaboration in epidemiological surveillance, scientific investigation, and public health and medical efforts to tackle and treat disease has already proved effective in tackling diseases such as SARS. It is likely to become increasingly important as new epidemics emerge and old ones reemerge.

[73] The Global Fund to Fight AIDS, *Tuberculosis and Malaria might serve as a starting point for such an effort*. See http://www.theglobalfund.org/en/

GLOBAL GOVERNANCE AND HUMAN HEALTH

CARLO JAEGER

1. Introduction: Epidemics as Social Challenges[1]

In the years between 1346 and 1350 an epidemic known as the 'Black Death' killed about one third of the population of Europe, i.e., more than 30 million people. Most likely, the epidemic had originated many centuries earlier in Mongolia, from where it had slowly traveled through China and the Middle East to Crimea and Turkey. In all these places it had created widespread suffering and killed huge numbers of people, without leading to far-reaching changes in the lifestyles of the surviving people.

The plague reached Sicily in 1346. One year later it was in France, two years later in England. In many cities, more than half of the inhabitants died within months. In Europe, the epidemic led to all sorts of reactions, ranging from the emergence of rituals of self-punishment to efforts at isolating the ill to block contagion. It also led to a massive shortage of manpower in many economic activities, and it lead to increased freedom to migrate for the surviving rural population. What it did not lead to was a general mood of despair or a period of economic stagnation. Quite the opposite, the plague was followed by the outburst of cultural creativity and economic expansion known as the Renaissance.

A remarkable cultural reaction to the experience of the Black Death is given by the Italian writer Giovanni Boccaccio. Immediately after the onset of the epidemic, around 1349, he started working on his masterpiece: the tale of a group of noble men and women fleeing from the

[1] I am grateful for helpful discussions to the research group on 'Globalization and the Financial Transition' at PIK and to O. Edenhofer, C. Engel, P. Epstein, A. Haas, E. Ostrom, O. Renn, M. Rohner, A. Sen, and L.Tubiana. The usual disclaimers apply.

plague in Florence to the countryside and entertaining each other by telling amusing short stories, each of which is worth reading in its own right. It is noteworthy that for the rest of his life Boccaccio was insecure whether his work was really something to be proud of or rather a sign of disorientation in the wake of a great disaster. He would certainly have been relieved to know that in the long run his creative reaction to the Black Death helped crystallize the Italian language along the lines set by Dante, whom Boccaccio deeply admired.

What distinguished Europe from other world regions hit by the plague earlier on was the fact that here this terrifying experience was countered by a wave of innovations that transformed not only Europe, but actually the world as a whole. While it would be naïve to attribute the emergence of modern society to the experience of the Black Death, it is remarkable how in this case the experience of a tremendous epidemic was integrated into a history of creative progress.

There are many other instances of large-scale epidemics, but none of them triggered a comparable response so far. Take the case of the 'Spanish Flu' of 1918-19. It seems that it originated in March 1918 in Kansas, U.S., most likely through a virus transfer involving poultry, pigs, and humans. The huge movements of soldiers involved in World War I led to a fast spread of the epidemic across the Atlantic. In Spain, which was not involved in the war, the epidemic was more widely publicized in the media, hence the somewhat misleading name 'Spanish Flu'. A remarkable feature of the flu was that it by no means spared young, healthy people, as flu usually does. In summer 1918, somewhere in France, the virus underwent a mutation that made it even more deadly. It reached Africa, Russia – brought there by troops fighting the communists –, China, India, and Australia. In 1919, it suddenly disappeared, having killed around 30 million people, about one third of which in densely populated, poor India.

This disastrous experience certainly played a role in the wave of revolutionary movements that rolled over Europe in the following years. But even if this was the time when some of the most significant breakthroughs of modern art – from Joyce's writing to Picasso's painting and Eisenstein's film-making – were achieved, it is a period remembered more for the rise of totalitarian ideologies and regimes than for cultural achievements on a par with those of the Renaissance.

Nowadays, the HIV-Aids epidemic confronts large parts of humankind with a similar challenge. Epidemiological research suggests that it is quite possible that even larger epidemics will arise in the years to come. At the

time of writing, avian flu is in the headlines around the planet. Toxic pollution may well acquire similar dimensions. Chemical, nuclear, and biological pollutants have the potential to cause human suffering and large numbers of death in many parts of the world, in particular in developing countries.

How will humankind respond to the experience of large-scale diseases, toxic pollution, and the clear risk of even larger future human health disasters? One possibility is to stick to existing patterns of development, another one to engage in spirals of increasingly destructive conflict. The present paper explores a third possibility: a creative process leading to new forms of solidarity in an age of globalization.

2. Illness and Two Forms of Social Solidarity

Since the origins of humankind, the reality of illness is tied to the possibility of solidarity. For obvious reasons, the rules defining rights and obligations in case of illness have always been closely related to analogous rules with regard to the elderly.

Three basic mechanisms are involved in such rules. First, there is the need to deal with uncertainty: we do not know for sure when we will be hit by illness and death. Second, there is the opportunity to improve everybody's lot by exchanging some resources for health care: a little time – or other resources – by the healthy and the young makes a huge difference for the ill and the old. As everybody is likely to play both roles in the course of time, everybody is better off by accepting some obligation to help. Third, there is the problem of free-riding: some sense of solidarity must be consolidated to make sure that people actually fulfill those obligations.

All three mechanisms can be mobilized quite effectively in face-to-face networks between people sharing long stretches of their lives. This may be a key reason why, until the 19th century, solidarity with the ill – as with the elderly – was rooted mainly in kinship and in religious practices. The latter are geared to kinship via rituals of marriage, baptism, burial, and more. And it is obvious that the phenomenon of social solidarity is closely linked to the Christian concept of Caritas (Benedict XVI, 2006).

Since the origins of humankind, being ill meant being dispensed from various obligations towards one's community, while being entitled to care from various members of that community. Sometimes this entitlement included help from a person with a healing role, like a Shaman. In the course of history, the happy few endowed with exceptional power or

wealth could summon healing services from others, but for most people kinship and religion remained the basis for the solidarity they needed in case of illness.

It is useful to relate this situation to a fundamental institutional dichotomy rooted in Roman law: the polarity of private and public law. Private law deals with interactions between physical persons, public law deals with interactions between the legal person constituted by a state and either other states or physical persons. This dichotomy is particularly relevant when it comes to property rights: private property rights are rights of physical persons to take action with regard to things (in Ancient Rome, a slave was a thing, too); public property rights are rights of states to take action with regard to things.

Along these lines, one can distinguish two forms of social solidarity, private and public. Private solidarity is rooted in kinship and religion (and to some extent plain neighborhood). As far as we can tell, it characterizes human culture since its origins at least several tens of thousand years ago. Public solidarity is rooted in people being members of the same state – a phenomenon that seems to have emerged for the first time in Mesopotamia less than ten thousand years ago. It seems that the differentiation of a public and a private domain was one of the most significant steps in the evolution of human culture.

The social solidarity required to deal with illness originated in the earliest forms of human culture and remained a matter of private solidarity for millennia. (An important exception is the solidarity provided by states to soldiers injured in wars). With the expansion of market institutions, things changed in a fundamental way, leading to current systems of health care and social security (Achenbaum, 2003; Kingson and Berkowitz, 1993; Tomasson, 2002).

This development took place in several steps. First, the old patterns of solidarity broke down as reliable means to deal with illness. They did so for a variety of reasons. The development of the market economy weakened older linkages of social solidarity that could provide non-monetized health care – in particular by severing them through migration. Moreover, the pattern of industrialization associated with the spread of the market economy was extremely unhealthy for large numbers of people, both in their workplaces and in their homes. As illness often brings old people closer to death, these problems made the situation of the elderly particularly difficult. A further reason is linked to the fact that various forms of health care became services traded on markets, along with medical drugs.

Episodes of illness often are random events, and once people started to live in a market economy, such episodes often required expenses that exceeded the financial means available in the given situation.

New forms of solidarity were called for, but they did not develop smoothly. In a second step, trade unions as well as a wide variety of friendly societies – including bodies like the Ancient Order of Foresters as well as the Masonic Order – provided sickness insurance to their members, i.e., insurance for health-related loss of income. Again, face-to-face networks with considerable continuity were involved in dealing with the three mechanisms highlighted above. In the setting of formal insurance schemes, the problem of uncertainty leads to the one of adverse selection, the problem of free-riding to the one of moral hazard. Trade unions and friendly societies were able to handle these problems by the solidarity organized via their well-articulated networks. The workers' movement turned this solidarity into a dramatic challenge of emerging capitalism – much as later on a similar movement, with important linkages to the Polish Pope John Paul II (Luxmoore and Babiuch, 2006), challenged communism in Poland.

Out of the resulting turmoil, in the third step far-reaching institutional innovations evolved. In Europe – and Canada – nation states organized systems of social security and public health (Freeman, 2000). They included provision of health care, but also systems of sewage and drinking water as well as various forms of support for the elderly. This meant that private solidarity in the face of health risks was substituted for by public solidarity. By helping its citizens to cope with the risks of illness and ageing, the nation state could strengthen a sense of national identity in its citizens – an effect consciously mobilized by Bismarck in establishing the German pension system.

In the U.S., the role of the nation state in health insurance was much weaker. In 1929, a group of Dallas teachers invented a scheme under which they would pay a hospital a regular fee in order to be able to go to the hospital when the need would arise. The American Hospital Association encouraged such schemes, state-level legislation freed them from various insurance-related regulations and from taxation. In 1934, the American Medical Association adopted a set of principles for similar schemes relating to physician services. Seeing the success of these schemes, private insurance companies entered the market, too. They handled the adverse selection and moral hazard problems by focusing on groups of employed workers. The elderly and the poor were left out by

this development, and in the 1960s national health insurance schemes – Medicare and Medicaid – were introduced to address their needs. But of course now there was a huge adverse selection problem: these programs had to serve a fraction of the population with above average health risks and below average incomes.

As for pensions, private companies took the initiative in the U.S. After a series of smaller initiatives, in 1940 General Motors established the first large-scale private pension plan. Such plans became widespread, and so the risk of their failure through bankruptcy became a serious issue. In the 1970s, therefore, legislation established a national body offering re-insurance for private pension plans.

Meanwhile, it is increasingly doubtful whether nation states will be able to successfully provide the social solidarity required to deal with illness and ageing in the globalized world we have entered (Sheehan, 2002). Several factors need to be considered here. Science-based medicine has led to highly sophisticated, but also very expensive, treatments for a variety of illnesses. With existing schemes of health insurance, these treatments can be sold in large numbers even if few people can afford them out of their more immediate financial resources. Moreover, public health systems combined with increasing economic welfare have led to spectacular increases in life expectancy. This implies rapidly growing costs for retirement schemes and additional costs for health insurance systems. For health care alone, in the U.S. the costs rose from about 7% of GDP to 15% over the past three decades (Holtz-Eakin, 2004, Table 3). Other industrialized countries show similar patterns, and the fraction of GDP earned by the elderly reached even higher levels.

By itself, there is nothing wrong with large and increasing amounts of income being spent to finance retirement and health care. In a globalized market economy, however, there is something worrying when financial streams of the order of magnitude of half the GDP are administered by nation states. Experiences as well as theoretical analysis suggest that a market economy tends to get inefficient if it is exposed to too much government control. Of course, one can argue that health care and social security involve public goods that must necessarily be provided by the state. However, globalized health risks as well as global demographic developments involve public goods not at a national, but at a global scale, and nation states cannot easily provide those (Kaul *et al.*, 2003; Feachem, R.A., Sachs, J., 2002).

Moreover, the demographic and economic situation of most developing countries makes it extremely difficult to establish nation-based systems of social security and health care. As a result, in a globalized economy engendering new health risks for everybody, and especially for the elderly, billions of people are exposed to these risks without a sound structure of governance to deal with them.

3. Health Care in Atomistic Markets

Faced with these challenges, the idea of cutting back the role of the state in favor of market solutions to the problem of human illness is gaining influence. This idea is based on the historical success of two remarkable innovations in handling money: credit and insurance.

Credit links the act of an agent A entitling an agent B to spend some amount of money X with a promise by agent B that at a later moment in time he will give some amount of money Y to agent A. (Y-X)/X then is the nominal rate of interest implied by that particular credit (to the extent to which a suitable rate of inflation can be indicated, the real rate of interest is the nominal rate divided by the rate of inflation).

Insurance is based on contingent contracts, i.e., contracts where an agent A pays some price P to an agent B while B promises that under a series of pre-specified conditions C1, C2,..,Cn, agent A will get payments X=f(Ci), i=1,2,..,n. An insurer sells a number of such contracts in such a way that he can expect the sum of his revenues to be larger than the payments he will have to make. This requires two things. On one hand, the conditions specified in the contracts must be such that it is highly unlikely for conditions triggering large payments to be realized simultaneously in most contracts. On the other hand, the buyers of the contracts must have a strong interest to avoid running into those conditions without compensating payments.

The latter phenomenon is known as risk aversion: often, people prefer a limited payment to be performed with certainty to a large loss that may happen as a random event. If people were immortal, this would mean that in the long run their average stream of payments would actually be larger than if they would accumulate wealth individually and pay their random losses out of this wealth.

In reality, however, most people die long before they had a chance to accumulate enough wealth to pay for health care in case of serious illness.

Under these circumstances, health insurance redistributes income from the healthy to the ill in ways that make both of them better off: the healthy live without the anxiety generated by the possibility of falling ill without the means to cope with illness, and the ill receive the means needed for that purpose (Nyman, 2002).

The basic mechanism is a Pareto improving set of transactions, i.e., a set of transactions whose result is preferred by all parties to the original situation. To the extent that this mechanism requires nothing than the invisible hand of the market to turn individual egoism into a collectively desirable outcome, one may describe it as the operation of an atomistic market.

With this background, we can now consider an ideal type of health care and retirement pensions as being provided exclusively by private businesses. This takes three kinds of firms: producers, insurers, and banks (of course one company may perform several of these activities). Producers deliver health services and medical drugs along with other goods and services. Insurers offer contracts for health insurance. With health insurance, the insurer gets a payment for each time period (say, each month), while the insured gets a payment in case of illness, where the amount of the payment depends on the kind of illness and on the treatments available. When somebody gets ill, he can then buy various health services and medical drugs, all of which are strict private goods in the sense introduced above.

In the ideal-typical situation to be considered here, retirement then is based on private saving organized by banks. During their economically active life, people accumulate savings on bank accounts, and during retirement they live by running down these accounts. As the time of death is not known in advance, however, this is not sufficient. In addition, there is a need for retirement insurance. Here, the fee must be paid in advance. If the insured then lives longer than average, he gets a pension from the insurer. If the life of the insured is shorter than average, his remaining wealth goes to the insurer (this helps reducing the fee for retirement insurance).

If inflation does not run out of hand and all incomes are high enough, this is certainly a feasible scheme for retirement. Although the lives of different people have different duration, the basic mechanism of such a scheme still is the use of part of current production to sustain the livelihood of the elderly. If the proportion of retired people in the population increases, everybody's consumption can still increase as long as production per capita increases, too. The growth of production per capita, in turn, depends to a considerable extent on the fraction of past output ded-

icated to gross investment. Besides a relatively stable rate of inflation and a sufficient level of the lower incomes, then, such a retirement scheme also requires that a sufficient fraction of gross income be dedicated to gross investment. If these conditions are met, even a population with a rapidly increasing fraction of retired people can avoid a pensions crisis.

Schemes for health care and for retirement financing, then, require different mechanisms. Nevertheless, they should not be analyzed in isolation, if only because health risks tend to increase with age (Alemayehu and Warner, 2004; Hogan *et al.*, 2001). A combination of credit, insurance, and saving yields an ideal concept for how to finance health care and retirement in a market setting. Given the remarkable capability of a market economy to satisfy people's wishes and needs, there is little doubt that a creative response to global health risks will heavily rely on these instruments. However, additional instruments are needed to cope with the rather severe limitations of atomistic markets in the face of those risks (Jaeger *et al.*, 2001).

4. The Importance of Club Goods

A useful model of an atomistic market can be framed with a set of two kinds of agents: households and firms. Households may be characterized by utility functions and initial endowments, firms by production functions. Firms produce goods and services and exchange them with each other and with households; households supply labor services and own the firms. The goods and services are private in the following technical sense: if a good or service enters the production function of a firm, it cannot enter any other production or utility function; and if a good or service enters the utility function of some household, it cannot simultaneously enter the utility function of any other household.

Under a series of non-trivial additional conditions, one can then show that no institutional arrangement can yield better results than a competitive market. Better, that is, in terms of the preferences of the households, and given a set of technologies and a distribution of initial endowments. This result is important because it helps to understand what happens if goods are not private in the sense defined above. One then gets so-called external effects that make the atomistic market inefficient.

Private goods are often contrasted with public goods, i.e., goods that enter all production and/or utility functions at once. This is a theoretical

limit case never to be found in practice. What really matters are club goods, i.e., goods that are necessarily used jointly by some set of firms and/or households. In the presence of club goods, the operation of atomistic markets can be improved by establishing suitable clubs, i.e., collective agents jointly using some good or service in such a way as to operate internally according to non-market rules and externally as a participant in the market.

Of course, firms, households, states all are clubs in this sense of the word. However, while in atomistic markets these operate as disjoint entities, in more complex markets collective agents overlap in a rich variety of ways. To overcome inefficiencies due to externalities generated by some club good, then, one needs to either find some existing club that can take care of it or to establish a new club that fits the case.

In practice, the expression 'public good' is used with regard to some restricted community, typically inhabitants of a nation state. In this sense, public goods are a special case of club goods, with the inhabitants – or the citizens – of a nation state forming the relevant club. The concept of global public goods can then be used to advocate the emergence of a global government, along lines first drawn by Kant in his essay on the need and possibility of 'perennial peace'.

But already Kant saw that a society of nations that have learned to interact peacefully may be a more realistic goal than a single global government. It may also be less dangerous, because the existence of multiple nation states provides a system of checks and balances that might be hard to match in the setting of a global government. This problem would become especially serious if the power of such a government would be geared to a monopoly of force on its territory – which then would be the whole planet.

A creative answer to global health risks may require more imaginative solutions than an extrapolation of the nation state to the global level (Ostrom, 1998). Such an answer may be facilitated by thinking about the importance of global club goods. A particularly important example of such a good is the professional knowledge developed and maintained in health-related institutions – hospitals, medical schools, pharmaceutical companies, etc.

Currently, scientific knowledge is geared to the power of a few nation states (and thereby more closely to military purposes than is often acknowledged). In the field of health as in other fields, this has led to a situation of opaque and sometimes missing accountability, and thereby to an incentive structure that leaves room for improvement (Fried, 1998). It

has also led to claims of certainty in areas where a more pragmatic approach seems warranted (Feyerabend, 1978; Jaeger *et al.*, 2001; Servan-Schreiber, 2004). Even global environmental change may be such a field.

Accountability in the development of health related knowledge will be particularly relevant in view of global health risks. Managing these risks will require sophisticated systems of health data that raise serious issues of privacy. And the expanding use of genetic engineering will confront us both with new opportunities to reduce health risks and with risks unknown before. Global governance for human health needs to take this into account.

5. The Potential of Vocational Solidarity

We live in a world increasingly shaped by truly staggering flows of money directed at short-term gains without sound incentives to pursue goals like poverty reduction, fostering peace, and avoiding environmental risks. These are not simply noble goals for idealists. In the age of nuclear weapons, they are essential goals for responsible behavior. There is a need for global mechanisms that combine a real capability to shape long-term developments with a robust system of checks and balances (Shiller, 2003).

One possibility to develop such mechanisms is by creating vocational groups that can pool the savings of people linked by shared vocational knowledge. In view of global health risks, it is worth considering the possibility that people working in the health sector might form one or several international associations pooling and managing the savings of their members. Similar groups may emerge in other sectors. In a way, they would renew the tradition of friendly societies and trade unions that lays at the origins of the American system of health insurance. And they would enable associations like the American Medical Association, which played such an essential role in the development of that system, to help address its current shortcomings.

By now, the largest investors on global financial markets are pension funds (Drucker, 1996/1976). These funds, however, are currently managed without in any way involving the people on whose savings they thrive. Vocational groups actively managing the savings of their members can modify patterns of social solidarity – in particular, the solidarity between generations (Shubik, 1981) – at a global scale. And as there would be a plurality of such groups competing with each other for members, they would provide a structure of checks and balances.

Such vocational groups would simultaneously possess global financial clout and have access to the professional knowledge of the broad economic sectors they represent. With this background, they would be in a position to provide an efficient mechanism for managing many large-scale health risks. The mechanism would work as follows. Any firm wishing to engage in some activity deemed risky by the law would need to be certified by a vocational group owning a critical amount of assets. The group certifying the operation would then accept unlimited liability for damages caused by the firm. The firm would be granted great freedom from regulations in its activity, while being held accountable for its action by the vocational group.

Vocational groups could quite naturally act as trustees of that global club good, vocational knowledge (Abbott, 1988). They can confer educational degrees on the basis of suitable tests, they can run schools leading towards these degrees, and they can run research institutes detached or combined with those schools. To deal with the health risks associated with global change does not require a static kind of knowledge settling the relevant questions once and for all. It requires a co-evolution of professional practice and the environment in which this practice takes place. Treating vocational knowledge as a global club good means to establish such a co-evolution. In the long run, this may enable scientific traditions to make even more effective contributions to the solution of global problems than they have already begun to do.

The understanding of vocational knowledge as a club good is highly relevant for the debate about how best to translate the health costs of globalization into incentives for medical R&D (Archibugi, Bizzarri, 2005; Kremer, 2000; Nosek, 2004). It may well be that in the longer run neither private companies nor nation-states or organizations formed by these agents will be the most appropriate agents to promote vocational knowledge about the health risks associated with global change. Vocational groups may turn out to be a superior alternative.

The possibility of vocational groups to operate as trustees of vocational knowledge resonates with Catholic social thinking (Dougherty, 2003; Nell-Breuning, 1936), but also with very different traditions of social theory (Durkheim, 1997/1893). Unfortunately, it is difficult to discuss this possibility in a dispassionate manner because it is also related to the radical experience of Spanish Anarcho-Syndicalism. Things are even more difficult because somewhat similar ideas were entertained in Fascist circles. However, it is neither a sign of strong democratic values

nor a step towards intellectual progress if the analysis of global club goods like vocational knowledge is blocked by an inability to distinguish between the use of a misleading idea and the abuse of a potentially fruitful idea. In the latter case, the task is to free the idea from its abuse in order to gain new insights from it.

Such insights are needed if we are to understand the health risks involved in global change (Epstein, 1999) and the factors influencing the differential vulnerability of people to those risks (Ribot *et al.*, 1996). They are needed if we are to take ‘global dimensions seriously, in regard to the formation of international solidarity and the constructions of identity patterns that go beyond national borders’ (Sen, 1999, p. 116). Exploring the potential of vocational groups to create global solidarity is an essential way of taking seriously ‘the range of multiple identities accessible to individuals and makes ‘justice’ applicable to a corresponding diversity of socio-political realities, independent of the idea of national frameworks’ (*loc. cit.*).

6. Conclusion: Three Practical Steps

If a truly creative response to global health risks is to emerge, this is likely to take several decades. Nevertheless, there are steps that can be made with a time horizon of a few years. Three such steps shall be mentioned here.

First, the risks of toxic chemical pollution can be reduced by developing suitable systems of certification and liability. In the short term, there is no way of doing this via a vocational group of people active in chemical industries. But it is possible to require any company dealing with risky chemical substances to be member of an industrial association representing a critical volume of assets. Such associations would be granted considerable freedom from government regulation in return for unlimited liability for damages caused by its members. This would provide an incentive structure where the professional knowledge of the firms involved would be used to keep risks within acceptable limits as set by liability law. And it would provide credibility to firms claiming that some technology is actually safe according to state-of-the-art knowledge.

Second, prevention and treatment of large-scale episodes of illness can be fostered by encouraging and supporting global networks of health professionals to monitor both risk indicators for such events and successful medical practice when dealing with them. This would have three

advantages. First, it would help collecting valuable information in a cost effective manner. Second, it would provide an opportunity to limit the tendency of governments and companies to weaken standards of privacy in dealing with this kind of data. And third, it would prepare the ground for the kind of vocational solidarity discussed in the previous section as a possible component of a long-term response to global health risks.

Third, the linkage between epidemics, poverty, and malnutrition can be addressed by a focused effort to realize the potential for sustainable development currently untapped in the world region most dramatically affected by the HIV epidemic: Sub-Saharan Africa. This implies a Marshall-plan style initiative of perfectly feasible proportions. The historical Marshall Plan did not simply organize a flow of money into Western Europe. It was a remarkably efficient scheme to strengthen critical institutions in Germany and other Western European countries (De Long, Eichengreen, 1993). In particular, credits were given in foreign currency and had to be paid back in local currency. This fostered imports into Western Europe and helped to reconstruct the German financial sector. Moreover, the whole initiative was consciously developed so as to build up the expectation that Western Europe would experience significant economic growth. This expectation greatly amplified the effect of the financial flows involved in the scheme.

A remarkable feature of the historical Marshall Plan was the extent to which it was based on learning by doing. A critical mass of highly competent people in different institutions were able to implement a shared vision, learning from preliminary successes as well as from partial failures. This capability would be even more important with regard to Sub-Saharan Africa. Launching a similar process is possible with a financial stream considerably smaller than the more than 50 billion Euros spent annually on German re-unification or on the aftermath of the Iraq war, as long as there is a clear long-term commitment – for about three decades – to maintain the process. Already after a first year of operation, it would be essential to review encouraging and discouraging developments, and to repeat such a review on an annual basis. Clearly, the whole operation might fail for many reasons, in particular lack of trust between investor and recipient countries. But it seems worth to take such a risk to avoid the clear and present danger of a destructive spiral of socio-economic development.

Along such lines it is possible to generate employment and strengthen vocational groups in Sub-Saharan Africa. This would help to build on the amazing degree of reconciliation – another form of solidarity – that has

been achieved in several countries of that region. The rich nations on the planet are in a position to launch such an initiative, and by so doing they could regain a sense of historical purpose that might be an even greater benefit than the economic return that can be expected from growing trade with a successful Sub-Saharan economic region.

REFERENCES

Abbott, A. (1988), *The System of Professions. An Essay on the Division of Expert Labor, Chicago*, The University of Chicago Press

Achenbaum, W.A. (1986), *Social Security: Visions and Revisions*, Cambridge, Cambridge University Press.

Alemayehu, B., Warner, K.E. (2004), The Lifetime Distribution of Health Care Costs, *Health Services Research*, 39 (3), 627-642.

Archibugi, D., Bizzarri, K. (2005), The Global Governance of Communicable Diseases: The Case for Vaccine R&D, *Law & Policy*, 27, 33-55

Benedict XVI (2006), *Encyclical Letter 'Deus caritas est'*, Vatican City, Libreria Vaticana.

De Long, J.B., Eichengreen, B. (1993), The Marshall Plan as a Structural Adjustment Programme, in Dornbusch, R., Nölling, W., Layard, R. (eds.), *Postwar Economic Reconstruction: Lessons for Eastern Europe*, London, Anglo-German Foundation for the Study of Industrial Society.

Dougherty, R.J. (2003), Catholicism and the Economy: Augustine and Aquinas on Property Ownership, *Markets & Morals*, 6, 479-495.

Drucker, P. (1996/1976), *The Pension Fund Revolution*, New Brunswick, Transaction.

Durkheim, E. (1997/1893), *The Division of Labor in Society*, New York, Free Press.

Epstein, P.R. (1999), Climate and Health, *Science*, 285, 347-348.

Feachem, R.A., Sachs, J. (2002), *Global Public Goods for Health. The Report of Working Group 2 of the Commission on Macroeconomics and Health*, Geneva, WHO.

Feyerabend, P. (1978), *Science in a Free Society*, London, New Left Books.

Freeman, R. (2000), *The Politics of Health in Europe*, Manchester, Manchester University Press.

Fried, S. (1998), *Bitter Pills: Inside the Hazardous World of Legal Drugs*, New York, Bantam.

Hogan, C., Lunney, J., Gabel, J. Lynn, D. (2001), Medicare beneficiaries' costs of care in the last year of life, *Health Affairs*, 20, 188-195.

Holtz-Eakin, D. (2004), *Health Care Spending and the Uninsured. Testimony before the Committee on Health, Education, Labor, and Pensions of the United States Senate*, www.cbo.gov/Search.htm.

Jaeger, C.C., Renn, O., Rosa, E.A., Webler, T. (2001), *Risk, Uncertainty, and Rational Action*, London, Earthscan.

Kaul, I., Conceicao, P., Le Goulven, K., Mendoza, R.U. (2003), *Providing Global Public Goods. Managing Globalization*, Oxford University Press.

Kingson, E.R., Berkowitz, E.D. (1993), *Social Security and Medicare: A Policy Primer*, Westport, CT, Auburn House.

Kremer, M. (2000), *Creating Markets for New Vaccines*, Harvard University, Mimeographed.

Luxmoore, J., Babiuch, J. (2006), John Paul's debt to Marxism, *The Tablet*, 14.1.2006.

Nell-Breuning, O.v. (1936), *Reorganization of Social Economy: The Social Encyclical Developed and Explained*, New York, The Bruce Publishing Company.

Nosek L.J. (2004) Globalization's costs to healthcare. How can we pay the bill?, *Nurs. Adm. Q.*, 28, 116-21.

Nyman, J.A. (2002), *The Theory of Demand for Health Insurance*, Stanford, Stanford University Press.

Ostrom, E. (1998), Scales, Polycentricity, and Incentives: Designing Complexity to Govern Complexity. In: Guruswamy, L.D., McNeely, J.A. (eds.), *Protection of Global Biodiversity: Converging Strategies*, Durham, NC, Duke University Press, 149-167.

Ribot, J., Magalhaes, A.R., Panagides, S. (eds.) (1996), *Climate and Social Vulnerability in the Semi-Arid Tropics*, Cambridge, Cambridge University Press.

Sen, A. (1999), Global Justice. Beyond International Equity. In: Kaul, I. *et al*. (eds.), *Global Public Goods. International Cooperation in the 21st Century*, Oxford, Oxford University Press, 116-125.

Servan-Schreiber, D. (2004), *Healing without Freud or Prozac: Natural approaches to curing stress, anxiety and depression without drugs and without psychoanalysis*, London, Rodale.

Sheehan, P. (2002), *Health Costs, Innovation and Ageing, Working Paper No. 9*, Centre for Strategic Economic Studies, Victoria University of Technology, Melbourne.

Sherman, P.B., Oakley, E.F. III (2004), Pandemics and Panaceas: The World Trade Organization's Efforts to Balance Pharmaceutical Patents and Access to AIDS Drugs, *American Business Law Journal*, 41, 353-59.

Shiller, R.J. (2003), *The New Financial Order: Risk in the 21st Century*, Princeton, Princeton University Press.

Shubik, M. (1981), Society, Land, Love, or Money: A Strategic Model of How to Glue the Generations Together, *Journal of Economic Behavior and Organization*, 2, 359-385.

Thomasson, M. A. (2002), From Sickness to Health: The Twentieth-Century Development of U.S. Health Insurance, *Explorations in Economic History*, 39, 233-53.

CLIMATE CHANGE AND HUMAN RIGHTS

WOLFGANG SACHS

Introduction

On February 16, 2005, when the Kyoto Protocol finally came into force, a long drawn-out process of consensus and institution building reached a temporary conclusion. International climate policy makers had achieved what they had been struggling over for the last century and a half. Now for the first time, despite setbacks along the way, industrial countries have a legal commitment to reduce greenhouse gas emissions. Thirteen years earlier, in May of 1992, governments had signed the United Nations Framework Convention on Climate Change. At that time it had dawned upon the world that the thin layer of atmosphere enveloping the Earth was being turned into a dumping ground for combustion-generated gases, and that this dumping ground was about to overflow. Twenty years after the bestseller, *Limits to Growth*, brought the finiteness of natural resources lying deep in the bowels of the Earth in to the limelight, the international community was forced to acknowledge that the finiteness of natural sinks in the air might be more urgent. The limits of the sky, rather than the Earth, turned out to be the more pressing issue. By 1997 an international treaty – the Kyoto Protocol – had been drawn up to define policy obligations for climate protection, and in 2005, after protracted international convulsions, the treaty acquired legal validity. As the 'entry into force' of the Kyoto Protocol is the beginning as much as the end of a historical process, it is advisable to look at some of the deeper implications of international climate policy, in particular at its implications for global fairness and equity.

At the time of the Rio Conference, it had already become clear that climate change is far from being just an ecological issue; it is also an issue of equity. In particular, climate change was identified as an issue of inter-

generational equity. It became ominously clear to observers that global warming, since it modifies important parameters of the ecology of the planet, such as sea levels or weather patterns, will affect the relations between present and future generations. Today's generation, by filling up the absorptive capacity of the atmosphere, lives at the expense of tomorrow's generation. At the same time, it came to the fore that the use of fossil fuels not only affects inter-generational equity, but also intra-generational equity, i.e., the relations between nations and social groups within a generation. Who will be allowed to reap the benefits from fossil fuel combustion? Who will have to carry the burden of emission abatement? Equity within a generation has at least two dimensions (Wuppertal Institut, 2005). First, it implies the fair distribution of burdens and benefits of fossil fuel use among nations. Secondly, however, it also implies the universal protection of human dignity by securing the fundamental rights of every human person to water, food, housing, and health. The present article will focus on the latter dimension; it will explore the links between human rights and climate change, without, however, losing sight of the broader framework of equity in climate politics.

Two Dimensions of Equity

Social science inquiries into systems of human resource use normally start where natural science inquiries have left off. The latter, by focusing on the relations between humans and the conditions of the natural world, usually speak about humans only in general terms; indeed, they mostly treat humanity as a black box. The former, in turn, usually hold rather general notions about the natural world, but attempt to shed light on the relations that are formed between people and between nations as they use the environment. From this vantage point, how the balance of power and opportunities within humanity is affected by particular patterns of resource use is an essential question.

Who has the advantages and who the disadvantages when nature is used? This may be considered the key question of environmental justice. It takes off from the widespread observation that the benefits and the burdens of resource use often do not fall to the same social actor, but are unevenly spread across different regions and social groups. While some might be able to enjoy the benefits, others might be forced to shoulder the burdens. Economists are well aware of this divergence in effects; they speak about the 'internalization of positive effects' when an actor is able

to seize the benefits of resource use, while they speak about the 'externalization of negative effects' when an actor is able to shift the burdens coming with a particular resource use to other social groups. However, this process has not only a biophysical, but also a social profile (Sachs, 2003). As organizations internalize benefits and externalize costs, societies are structured into winners and losers. Power relations ensure that positive consequences crystallize at the top end and negative consequences at the bottom end. This shifting of costs may take place in a temporal, spatial or social dimension: that is, costs may be shifted temporally from present to future, spatially from center to periphery, and socially from upper classes to lower classes.

Two critical dimensions can be distinguished in the distribution of benefits and costs. They point to the two most important concepts of justice: human dignity and equality. Both dimensions differ in their starting-point and in their conclusions. The demand for human dignity starts from the absolute necessity of certain living standards, and insists that these must be achieved for all, whereas the demand for equality focuses on relations among people and presses for the leveling out of inequalities. In other words, the dignity concept of justice rests upon a non-comparative approach that looks at the absolute provision of certain fundamental goods and rights, while the distributive concept of justice rests upon a comparative approach that looks at the proportional distribution of various goods and rights (Krebs, 2002). Both dignity and equality go to make up the ideal of justice; therefore, any policy striving for equity will keep in mind both human rights issues and distributive issues.

The use of fossil fuels, as with any environmental resource, results in burdens as well as benefits. Issues of human dignity arise with regard to the distribution of dangers. Potential threats are not distributed equally across the globe; both developing countries and lower social classes are likely to be disproportionately affected, possibly to such a degree that fundamental rights might be violated. The IPCC (2001) has confirmed that developing countries are most at risk of climate change. Thus, under certain conditions, global warming may undermine people's right to a secure livelihood. Issues of distributive justice, in contrast, arise in particular with the unequal distribution of access to the atmosphere as a deposit for greenhouse gases. This is because emissions not only produce the burden of marginalization, they also produce the benefit of power, and the right to use the atmosphere as a dumping ground represents a source of economic power. Disparity in access leads to disparity in economic opportunities; it parti-

tions the world society into winners and losers. Such a situation is unjust if it allows certain nations to maximize their freedom to flourish at the expense of the freedom of others. Therefore, at the international level, equity calls for a rebalancing of opportunities among nations.

What rise of the global mean temperature can be tolerated? This question is technical in appearance, but highly political in reality. It hides fundamental issues of how to live together in an interconnected world. In particular, it decides about the human rights impact of anthropogenic climate change. As is well known, the United Nations Framework Convention on Climate Change calls for the stabilization of greenhouse gas concentrations at levels that, '...would prevent dangerous anthropogenic interference with the climate system' (Article 2). Such levels should be achieved, '...within a time frame sufficient to allow ecosystems to adapt naturally to climate change, to ensure that food production is not threatened and to enable economic development to proceed in a sustainable manner'. Up to this date climate negotiations have refrained from defining what may constitute dangerous anthropogenic interference with the climate system (Hare, 2003). What kind of threat qualifies as 'dangerous'? Twenty centimeters of sea level rise or two meters? One degree rise in medium global temperature or three degrees? And in what time frame, in twenty years or in eighty years? Different impacts are associated with different levels of temperature rise; who will be affected, how, and to what extent largely depends on how far global warming is allowed to go.

In the end, however, it is a matter of political and ethical judgment to determine how much climate change is tolerable. This is not only because any assessment of risk implies a value judgment, it is also because the avoidance of a risk often implies missed opportunities, in the case of climate change, these are usually opportunities for economic growth. Moreover, most evaluations of the dangers will have to implicitly or explicitly confront the question: dangerous to whom? The potential cost associated with continuing to risk high levels of danger are of a different kind, and fall on a different set of people than the cost associated with avoiding dangerous change. Any decision on what is to be considered a dangerous level of impact is clearly a political and ethical issue. It basically implies two valuations: what kind of danger is acceptable, and what kind of danger is acceptable for whom? It is the response to the latter question that determines the degree of environmental injustice involved in climate politics.

Impacts

Estimating possible impacts of global warming on human communities is a notoriously difficult endeavor, for at least three reasons. First, most effects resulting from rising global mean temperatures affect people only indirectly; they may lead to changes in the conditions of ecosystems that in turn may impact conditions of human systems. While already different ecosystems are not equally affected by climate change, human communities exhibit an even wider range of vulnerability. They are exposed differently to changes in natural variables, depending, for instance, on if they are peasants or city dwellers. And they are capable of coping with adverse circumstances in different ways, depending, for instance, on if they are poor people without means or rich people with insurance policies. This diversity of settings renders general statements about causal relationships very problematic. Secondly, human development is not going to stand still over the rest of the century; there are, however, different development paths that societies – or the world as a whole, for that matter – can take. It is impossible to anticipate what development path will eventually prevail, regardless of any climate policy. But levels of atmospheric CO_2 concentration will vary according to the development path chosen; the IPCC SRES scenarios, therefore, imply different impacts according to development paths. And thirdly, the vulnerability of human societies also varies according to development paths taken; for instance, the A-family of SRES scenarios generally leads to higher vulnerability than the B-family of scenarios. For these reasons, any sweeping affirmation about the impacts of global warming on people should be taken with caution; nevertheless, severe impacts are to be expected whose general pattern has emerged in recent research (IPCC, 2001; Exeter Conference, 2005). The most important impacts are likely to affect natural assets that underpin human existence – water, food, health.

With regard to *water*, it is important to note that currently 30 countries with a combined total population of over 500 million have less than 1000 m3 of renewable water available per capita per year; they are considered water-scarce, a condition which by the year 2025 is likely to affect some 50 countries with a combined population of about 3 billion (Shah *et al.*, 2006). Projected climate change will exacerbate water shortage in many water-scarce areas of the world, though it will alleviate them in some other areas (IPCC, 2001). The hydrological cycle is expected to intensify, which essentially means more droughts and floods, and more variable and extreme

rainfall. Generation-old patterns of rainfall may be shifting with corresponding consequences for plants, animals and people. Several hundred million to a few billion people are expected to suffer a water supply reduction of 10% or more by the year 2050 for climate change projections corresponding to a 1% per year increase in CO_2 emissions. Regions where water stress is likely to increase due to climate change include central and southern Africa, central and southern America, and the watersheds around the Mediterranean, while South and East Asia are likely to see an increase in water resources. (Arnell, 2004). Finally, too much of the wrong water can be dangerous as well. Rising sea levels obviously increase the risk of coastal flooding that could displace large numbers of people. Some of the most vulnerable regions are the Nile delta in Egypt, the Ganges-Brahmaputra delta in Bangladesh, and many small islands, such as the Maldives, the Marshall Islands, and Tuvalu.

Furthermore, climate change will leave its imprint on the conditions for *food* production across the globe. In temperate zones, small increases in temperature might boost yields for some cereals, while larger changes are likely to decrease yields. In most tropical and subtropical regions, potential yields are projected to diminish with most increases in temperature. For instance, damage to the world's major crops begins when daytime temperatures climb above 30°C during flowering. For rice, wheat, and maize, grain yields are likely to decline by 10% for every one degree C increase over 30°C (Halweil, 2005). If, in addition, there is also a large decrease in rainfall in subtropical and tropical dryland/rainfed systems, crop yields would be even more adversely affected. In tropical agricultural areas, yields of some crops are expected to decrease even with minimal increases in temperature (IPCC, 2001). In sum, 20-40 poor and food-insecure countries, with a projected population in 2080 in the range of 1-3 billion, may lose on average 10-20% of their production potential in cereals due to climate change (Fischer *et al.*, 2002). Moreover, it is expected that the income of poor farmers will decline with a warming of 1.5-2°C above preindustrial levels. (Hare, 2003). In fragile rural areas, such a change will aggravate the fate of people that derive their livelihood from direct access to forest, grasslands, and water courses. In developed countries crop production, in contrast, is likely to benefit from climate change at least initially, compensating for the declines projected for developing countries. Thus while global production appears stable, regional differences in crop production are likely to grow stronger through time, leading to a significant polarization of effects, with substantial increases in the risk of hunger amongst the poorer nations,

especially under scenarios of greater inequality (Parry *et al.*, 2004). Declines in food production will most likely hit regions where many people are already undernourished, notably Africa.

Finally, as public health depends to a large extent on safe drinking water, sufficient food and secure shelter, climate change is bound to have a range of *health* effects (McMichael *et al.*, 2003). On the first level, a shortage of freshwater caused by climate change will increase risks of water-borne disease, just as shortage of food will increase the risk of malnutrition. On the second level, climate change, via both a shift in background climate conditions and changes in regional climatic variability, will affect the spatial and seasonal patterns of the potential transmission of various vector-borne infectious diseases. With global warming, it is expected that there will be an increase in the geographic range of potential transmission of malaria and dengue – two vector borne infections, each of which currently affects 40-50% of the world population. A rise in temperatures, for example, would result in an increased prevalence of malaria in higher altitudes and latitudes. Within their present ranges, these and many other infectious diseases would tend to increase in incidence and seasonality, although decreases would occur for some infectious diseases in some areas. The human-induced warming that the world is now experiencing is already causing 150,000 deaths and 5 million incidents of disease each year from additional malaria and diarrhea, mostly in the poorest nations (Patz *et al.*, 2005). However, in all circumstances actual disease occurrence is strongly influenced by local environmental and social conditions. On the third level, climate change will be accompanied by an increase in heat waves, often exacerbated by increased humidity and urban air pollution, which would cause an increase in heat-related deaths and episodes of illness. The impact would be greatest in urban populations, particularly affecting the elderly, the sick, and those without access to air-conditioning. Furthermore, a reduction in crop yields and food production will predispose food-insecure populations to malnutrition, leading to impaired child development and diminished adult activity (IPCC, 2001).

Summing up these possible effects of global warming on sea levels, water availability, and the incidence of malaria, it has been estimated that in the case of greenhouse gas emissions that result in a global temperature rise of a rather moderate 2°C, by the year 2050 some 25 million additional people will be threatened by coastal flooding, 180 to 250 million by malaria, and 200 to 300 million by water shortages (Parry *et al.*, 2001).

Human Rights

There has been injustice in the world ever since Cain killed his brother Abel. Similarly, the expulsion of people from their land, the assault on their physical well-being, and the withdrawal of their means of subsistence have always been standard instruments in the repressive exercise of power. But only since the middle of the 20th century have such ways of holding others cheap been thought to involve contempt for human rights. In the past, according to the age and the local circumstances, they may have been seen as violations of the ruler's duties, as infringements of rights and customs, as sins against God or as evidence of oppression. In today's world, however, there exists the international consensus that instances of humiliation and impoverishment have to be measured against the norm to guarantee the fundamental rights of every human person. By birthright, people are considered bearers of rights for protecting their dignity, regardless of their nationality or cultural affiliation. These rights are equal, i.e., everyone enjoys the same rights, they are inalienable, i.e., they cannot be forfeited, and they are universal, i.e., every human being is a holder of fundamental rights (Donnelly, 2003). Especially in an age of globalization, it is increasingly the discourse of human rights that sets the terms of reference for disputes over power and its victims.

Before the Second World War it was just states that could claim rights. The rights of persons were first recognized at the international level only with the Universal Declaration of Human Rights, in 1948. This may be seen as the juridical revolution in human rights (Ignatieff, 2001), which later was complemented by a revolution in their advocacy and enforcement. The legal breakthrough came after Nazi crimes and horrors – the worst imaginable disaster for the rule of law – had shown the depths to which a totalitarian *raison d'État* can lead. That experience gave rise to a codification of the basic political rights of each and every individual in the world vis-à-vis state power. Subsequently, the juridical revolution made further progress with the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights (in force since 1976), and with the World Conference on Human Rights held in Vienna in 1993. Although symbolic politics have often seemed to be dictating the course of things, it has become clear over the years that even rhetorical affirmations can be a political weapon in the hands of the powerless. For this reason, what Ignatieff calls an 'advocacy revolution' has occurred in the 1980s and 1990s. Groups belonging to international civil society – the

best-known being Amnesty International and Human Rights Watch – have put various states in the dock for their violations of basic rights. In the name of human rights, numerous campaigns have begun to interfere in hitherto internal affairs of states. With the appointment in 1993 of a High Commissioner on Human Rights, the United Nations gave itself an instrument of its own to investigate excessive internal sovereignty claims on the part of various states. However, an 'enforcement revolution' can hardly be considered to have happened on a large scale. Nevertheless, such institutions as the International Courts of Justice in Arusha and The Hague, or the International Criminal Court set up in March 2003, demonstrate that a trend is under way which might make human rights violations justiciable, over and beyond the principle of state sovereignty. In any case, the three 'revolutions' mentioned have combined to give human rights greater prominence throughout the world.

For a long time, however, people's economic, social and cultural rights played a subordinate role in this growing legal awareness. This was largely due to the Cold War; for the Western bloc inscribed civil and political rights on its banner, while the Eastern bloc did the same with economic and social rights. The two sets of human rights were ritually played off against each other, with the result that social rights were taken no more seriously in the West than democratic rights were in the East. But meanwhile this confrontation has resolved itself, and the inseparability and interdependence of political and social human rights have been largely accepted (Steiner/Alston, 1996). Indeed, it would be hard to understand why disease or malnutrition should be less important than press censorship or religious persecution in affecting people's ability to act. Without social and economic rights, the minimum basis for equality of civil and political rights is lacking and, conversely, social and economic rights without civil and political rights are robbed of the motive power of freedom. A minimalist conception of human rights that refers only to negative political freedoms therefore discriminates against the have-nots and those whose livelihood is threatened; recognition of their dignity requires the protection of economic, social and cultural rights.

When human beings do not have the basic capability to support themselves with dignity, their human rights are under threat. Most societies regard as basic requirements: the capability to obtain adequate nourishment, to avoid unnecessary illness and premature death, to have adequate housing, to earn one's own livelihood, to be assured of physical safety, to have equal access to justice, to appear in public without feeling ashamed,

and to take part in the life of a community [OHCHR, 2002]. The International Covenant on Economic, Social and Cultural Rights declares that 'the State Parties to the present covenant recognize the right of everyone to an adequate standard of living for himself and his family, including adequate food, clothing and housing ...' (Article 11) and 'the right to the highest standard of mental and physical health' (Article 12). Under the influence of this formulation – which echoes Article 25 of the Universal Declaration of Human Rights – the debate on development has changed its color in the subsequent decades; overcoming hunger, illness, and misery is not seen any longer as a matter of charity or solidarity, but as a matter of human rights. The need-centered approach in development has thus been largely replaced by a rights-centered approach; moreover, it has also been adopted by leading international organizations, such as UNICEF, WHO, and UNDP.

Rights generate duties, but needs – in the best of cases – just compassion. Anyone who speaks of rights asserts that certain institutions and authorities have an obligation to give an account of themselves; the language of rights strengthens the power of the marginalized. First of all, of course, governments constitute the prime duty bearers in human rights law. Indeed, securing human rights should be the first priority of governments, as Article 1 of the Vienna UN Conference on Human Rights has affirmed. However, in a transnational world where the influence of states has become more and more circumscribed while the influence of corporations and multilateral institutions is on the rise, a case can be made that non-state actors will have to observe corresponding duties as well. It is, after all, difficult to imagine how there can be universal human rights without universal human duties. While the Universal Declaration limits itself to the allusion that 'everyone has duties to the community in which alone the free and full development of his personality is possible' (Article 29), the 'Declaration on Human Duties and Responsibilities', as it was proposed in 1997 under the auspices of UNESCO, is much more explicit: 'Members of the global community have collective, as well as individual duties and responsibilities, to promote universal respect for and observance of human rights and fundamental freedoms' (Declaration 1997).

It is remarkable that the text, speaking about members of the global community, is not simply referring to states, but also to transnational corporations, international organizations, associations, even to all communities of people, including the individual person. Indeed, things would look bad for human rights if only states continued to be considered duty-

bearers in a globalized world. Instead, as philosophers increasingly argue (O'Neill, 2000; Scheffler, 2001), all actors that exert power in a more and more border-less world carry responsibility for the protection of fundamental rights. The basic dignity of people is to be safeguarded against any form of denigrating power, regardless from whom and where it originates. For rights cannot be maintained universally, unless the duty of observing them is shared universally. In the end, it is nothing but the golden rule of ethics that underpins this conclusion, demanding that 'what you do not wish to be done to yourself, do not do to others'. As interactions across borders intensify, rendering state borders ever more porous, this rule provides a minimum moral ground for the recognition of universal basic rights in the emerging world society.

Rights-Based Climate Policy

The bitter consequences resulting from climate change – in particular several decades from now – will spread across the globe, albeit in varying degrees. Even rich countries in temperate zones are not able to shield themselves against adverse impacts, as the 25,000 deaths caused by heat waves in Europe in the summer of 2003 have dramatically shown. Yet researchers converge on the general assumption that developing countries are most at risk of climate change, with damage at even low levels of warming and increasing rapidly with rising temperature (IPCC, 2001). Countries – and regions within countries – are disproportionately affected for basically two reasons: higher impacts and higher vulnerability. As indicated above, adverse impacts of climate change are likely to be more concentrated in areas of Africa, South America, and Asia. Impact profiles differ according to kind of impact and geography, but water stress and flooding, declining agricultural productivity and weakening ecosystem services, crop pests and human diseases are more likely to occur in subtropical and tropical countries, in coastal areas, and in arid and semi-arid agricultural areas. Higher vulnerability, however, derives from the fact that in many places at risk a great number of people already live under fragile conditions, economically and healthwise. The ability to prepare for and to cope with threats varies widely according to income and living conditions. The impact of a hurricane in Orissa, for example, may be much more severe than the impact of a similar hurricane in Florida. Likewise, in 1999 there were two to three times more disaster events reported in the US than in India or Bangladesh, but there were 14 times

and 34 times more deaths in India and in Bangladesh than in the US (UNEP, 2002). Wealth, technology, and infrastructure facilitate adaptation and the ability to cope. The poor generally tend to have much lower coping capacities; they are more exposed to disasters, drought, desertification, and slow economic decline.

Climate perturbations are likely to be superimposed on economic insecurity. As a consequence, climate impacts are at times likely to aggravate the living conditions of people up to a point where their basic rights are in jeopardy. It is for this reason that climate impacts may turn into a matter of human rights. As people already living at the edge see themselves pushed over into disaster, climate effects may trigger an infringement upon economic and social human rights. This is not to say that climate-related threats (hurricanes or heat waves, for instance) to human physical integrity under conditions of greater affluence may not constitute a human rights violation as well, but they are going to be more occasional and less structural in terms of their occurrence, just as they are going to be more accidental and less predictable in terms of their location. Impacts in poorer regions, in contrast, often add to an already structurally precarious livelihood situation; it is the compounded effect of economic insecurity and climate stress for large numbers of people that centers around the question of how much climate change should be allowed into a human rights issue.

However, climate-related human rights are matched only by imperfect, not by perfect duties. Like with most economic, social and cultural rights, the link between the right and the corresponding duty is blurred. Just as a violation of the right to food, health, or shelter, can often not be traced back to the action of a clearly identifiable duty-bearer, also climate effects cannot be attributed to a culprit with name and address. Who exactly should be held responsible for hunger and widespread illness? While it might be possible to identify the victims, it is often impossible to identify the responsible agent or the causal relationship between a specific action and a specific damage. In fact, an objection often raised against the concept of economic, social and cultural rights holds that rights make no sense unless they are combined with exact duties imposed on specified actors who would make sure that these rights are fulfilled. But the objection is flawed, for it militates against the basic idea that people have some claims on others and on the design of social arrangements regardless of what laws happen to be enforced (HDR, 2000). The absence of culprits or judges does not nullify rights. A strictly legal conception,

which maintains that there are no rights unless they are justicable, misses out on the universalist nature of human rights entitlements.

Furthermore, climate rights call for extra-territorial responsibility, even more so than do economic, social and cultural rights. Climate perturbations most clearly surpass the jurisdiction of single states, they are in fact a striking example for the transnational character of threats in a highly interdependent world. Under such circumstances, the human rights obligations of states and non-state actors cannot simply stop at territorial borders; rather, they reach geographically to other countries as well. As the Special Rapporteur to the Human Rights Commission on the Right to Food has recently stated: 'Governments must recognize their extraterritorial obligations towards the right to food. They should refrain from implementing any policies or programs that might have negative effects on the right to food of people living outside their territories' (UNCHR, 2005). When the right to food is threatened by climate change, the principle of extra-territorial obligations becomes even more relevant, given that rich countries are largely responsible for climate perturbations in poorer countries. Just as climate effects reach to the ends of the earth, the geographical scope of responsibility has become global as well.

However, this responsibility is in the first place a negative one; it implies avoiding harmful action rather than intervening to provide conditions for an unmutilated life. In other words, climate responsibility is first of all a matter of self-limitation on part of high-emitting nations and social groups, not a matter of benevolent imperialism bent on improving the world. It is, incidentally, the liberal core of human rights law to emphasize negative obligations, i.e., to call on power-holders to refrain from actions that infringe upon people's integrity. Since institutions are nothing but consolidated systems of action, the human rights imperative can be reformulated by saying that social institutions – including, one might add, energy systems – should be shaped in such a way that they do not structurally and permanently undermine fundamental rights (Pogge, 2002). Under human rights law, governments are supposed to carry out a triple task with regard to the rights to food, health, and housing (Steiner/Alston, 1996). They are first and foremost obliged to respect these rights by avoiding violating them through state measures; they are further required to protect them against powerful third parties, such as industries or landlords; and they are, in the end, expected to fulfill them only through positive action by facilitating access to food, health or housing. It would follow to apply the same hierarchy of obligations to climate rights; the right to live in freedom from

human-induced climate perturbations has first to be respected by avoiding harmful emissions nationally, it has, secondly, to be protected against third-party emissions of countries or corporations through international cooperation, and it has, thirdly, to be fulfilled by upgrading people's capability to cope with climate change through adaptation measures, such as dam-building, resettlement, or land redistribution.

Mitigation and Adaptation

In 2005, the Inuit Circumpolar Conference filed a legal petition to the Inter-American Commission of Human Rights demanding that the US limit its emissions. This move on part of the people living in the Arctic represents the first legal case brought against a high-emitting nation in defense of economic, social and cultural human rights. The petition is based on both long-term projections and current experience (Watt-Cloutier, 2004). Scenarios project massive thinning and depletion of sea-ice, with the result that ice-inhabiting marine species – seals, walrus, and polar bear – may be pushed to extinction by 2070-2090. And current experience shows that Inuit hunters run into growing difficulties in predicting weather conditions, the solidity of sea-ice, and the location of species to be harvested. Many indicators suggest that global warming is threatening the ability of Inuit to survive as a hunting-based culture.

From a human rights point of view, the classical policy responses to dangerous climate change, mitigation and adaptation, acquire an additional urgency. As to mitigation, human rights considerations need to enter into the definition of what constitutes dangerous climate change. They direct attention to the most vulnerable sections of the world population, suggesting a frame of evaluation that is consistent with the basic law that governs world society. However, negotiations at present fail to define a target of tolerable climate change that would sufficiently protect the fundamental rights of the most vulnerable people. The IPCC low concentration scenario results in a CO_2 concentration of 450 ppm CO_2 and a total greenhouse gas concentration equivalent to about double pre-industrial levels. This would produce a long-term temperature increase of about 2.5°C at the present best estimate of climate sensitivity (Hare, 2003). However, a survey of possible impacts (Exeter Conference, 2005) suggests that a target that avoids systematic threats to human rights would need to keep the global mean temperature increase below 2°C above pre-industrial levels. It is obvious that such a target calls for mitigation commitments far beyond the Kyoto Protocol.

One reason, however, for the neglect of a human rights approach so far is the prevalence of a utility-based framework of evaluation in climate research and politics (Rayner/Malone, 1998). In this framework, benefits of climate mitigation are weighed against its cost in order to optimize both the amount and the time of protection measures. Achieving optimal welfare on the national or global level is the overriding goal. Yet the focus on aggregate welfare is largely incompatible with a focus on rights. For a rights-based framework centers on individual, local or ethnic rights that are not to be violated even at the expense of the aggregate good. It concentrates on the distribution of advantages/disadvantages across single groups, not on the maximization of welfare at the collective level (Höffe, 1989). It is therefore immune against considerations like the one, for example, that the flooding of the Maldives might be a cost to be justified by the aggregate benefit of unhindered growth. The utility approach is all too often inclined to trade away rights for higher aggregate welfare, while human rights are clearly absolute rights; they cannot be traded for higher incomes or disregarded because of a majority opinion.

Finally, human rights considerations also call for vigorous measures to facilitate adaptation to unavoidable climate change. Inasmuch as mitigation is insufficient, the polluter-pays principle requires that high-emitting nations offer compensation for damages caused. In particular, in a human rights perspective they are obliged to prevent violations of economic, social and cultural rights by adequate protective measures. These may range from upgrading health care, to investments in construction, to the building of dams. Governments, however, have so far not been very forthcoming; only a levy on projects in the framework of the Clean Development Mechanism is earmarked for this purpose up to this date. In any case, there can be no doubt that the adherence of the more affluent countries to human rights principles will be put to a hard test as long as emissions remain at current levels.

Whose Atmosphere?

During climate negotiations, both developed and developing countries – apart from the Island States – have shown little interest in defining low-danger emission caps. All parties disregard the fact that, when it comes to capping emissions, the choice is between livelihood rights and the desire for affluence. The task of keeping the temperature rise below 2°C appears too large and too threatening to the economic interests of consumers and

corporations. In particular, it still seems to have escaped the attention of developing countries that climate protection is of the utmost importance for the dignity and survival of their own people. It is time they became protagonists of climate protection, because climate protection is not simply about crops and coral reefs, but fundamentally about human rights.

The Kyoto Protocol fails to live up to this challenge. It does not demand serious reductions from the North, nor does it include newly industrializing countries from the South. Nevertheless, for the second commitment period of the Kyoto process, an ecological breakthrough cannot be expected unless the South assumes commitments as well. In fact, current emissions from developing countries alone would already overstress the absorptive capacity of the atmosphere, even if all of the industrial countries were to vanish from the Earth by a stroke of magic (Ott *et al.*, 2004). Without the participation of the newly industrializing countries, global climate protection is bound to fail.

At this point, however, fairness might become the only realistic option (Athanasiou/Baer, 2002). For it turns out that the human rights issue can probably not be solved without addressing the issue of an equitable distribution of benefits from fossil fuel use – the second dimension of climate equity. This is why developing countries will refuse to cooperate as long as they have reason to fear that reduction commitments on their part will consolidate the inequality among nations for eternity. They will perceive any request to cooperate as an attempt of the well-to-do to pull up the ladder by which the rich themselves have climbed to success and power. Saving the climate at the price of long-term inferiority is not an option for them. Indeed, why should countries such as India, Brazil or China enter an agreement that would constrain them for an undetermined period of time to emit fewer greenhouse gases than industrial countries? It is difficult to avoid the conclusion that without greater fairness between North and South all appeals for effective climate protection will be in vain.

Fortunately, the terrain for greater fairness among nations is not entirely unprepared. Already the Climate Convention of 1992 has underlined the significance of international distributive equity. As Article 3.1 states, 'The Parties should protect the climate system for the benefit of present and future generations of humankind, on the basis of equity and in accordance with their common but differentiated responsibilities and respective capabilities. Accordingly, the Parties of the developed countries should take the lead in combating climate change and the adverse effects thereof'. Only developed countries are expected to assume reduction com-

mitments and financial burdens, while developing countries have just reporting duties. This unequal distribution of commitments arises from the unequal responsibility of countries for climate change. As it happens, industrialized countries are responsible for the bulk of carbon dioxide emissions in the past and in the present. Since 1800, approximately 80% of the rise in cumulative emissions is attributable to the industrialized countries; at present they are responsible for nearly 50% of global carbon dioxide emissions, yet the industrialized countries represent only 25% of the world population. In the light of this situation, only industrial countries are subject to reduction commitments under the Kyoto Protocol.

However, the way in which reduction commitments have been distributed among the industrial countries in Kyoto was a matter of accident and political shrewdness rather than systematic consideration. There is no explanation why, for instance, Australia, was granted a further rise in emissions of 8% while Japan was forced to reduce emissions by 6%. In any case, rules for the distribution of reduction commitments will be at the center of attention the moment developing countries are expected to come onboard a governance system for climate protection. The atmosphere, however, belongs to nobody in particular and to everybody equally; in other words, the atmosphere is a global common good. In the future, who should be allowed to use it, and by how much? What principles should determine the fair distribution of the 'cake' that is available?

Among observers of the negotiations, this issue has been hotly debated for some time (Brouns, 2004). For instance, some put forth the grandfathering principle, according to which each nation has to accept equal reduction commitments, disregarding the present unequal distribution of emissions. However, as such a principle would maintain the global welfare gap; it can hardly be considered fair. Brighter prospects are offered by the capabilities principle that demands commitments according to the capability of countries to reduce emissions. Economically strong countries are expected to carry the bulk of the reduction load, regardless of how efficiently they use energy. This proposal may be fair, but it is ecologically counterproductive, as wasteful countries would enjoy an advantage. A third principle calls for a distribution of commitments according to the historical responsibility of countries for loading the atmosphere with greenhouse gases. Each country's obligation would be measured by its relative contribution to global warming. Indeed, in 1997 Brazil introduced a proposal along these lines to the climate negotiations; the issue of equity has been squarely on the agenda of environmental diplomacy

ever since. Countries are expected to assume obligations according to their share of cumulative emissions, given that the ominous concentration of greenhouse gases in the atmosphere has been built up over 150 years. Such a scheme would place the biggest burden by far on industrial countries. However, it is doubtful to what extent responsibility can be assumed for actions that have been adopted in ignorance of their consequences. After all, the possibility of a greenhouse effect was known to just a handful of specialists before the 1980s.

The situation is different when it comes to the equal entitlement approach (Meyer, 2000). This calls for a framework that respects the principle of an equal per capita right to the Earth's atmosphere. Most other allocation schemes would repeat a colonial style approach, granting disproportionate shares to the North. If the use of a global common good has to be restrained through collective rules, it would violate the principle of equity to design these rules to the advantage of some and the disadvantage of many. The equal right of all world citizens to the shared atmosphere is therefore the cornerstone of any viable climate regime. Therefore, for the second commitment period of the Kyoto Protocol, a process should be initiated whereby each country is allocated emission allowances based on equal rights per capita. This is hard on the North, but not unfair, as in exchange for accepting the rule of egalitarianism in the present, industrial countries would not be held liable for emissions accumulated in the past.

Assuming an equal right to the Earth's atmosphere, broadly speaking it is possible to envisage different development paths for North and South. All countries are expected in the long run, to converge upon a similar level of fossil energy-use per capita. The North will contract, while the South will expand towards a convergence with the North. Over-users will have to come down from their present level, while under-users are permitted to raise their present level, albeit at a gradient that is much less steep than the one industrial countries went through historically, leveling off at the point of convergence. However, the convergence of North and South on equal emission levels cannot be achieved at the expense of contraction, i.e., the transition to globally sustainable levels of emissions. Once again, sustainability gives rise to equity. Indeed, the vision of 'contraction and convergence' combines ecology and equity most elegantly; it starts with the insight that the global environmental space is finite, and attempts to fairly share its permissible use among all world citizens, taking into account the future generations as well.

REFERENCES

Arnell, Nigel W. (2004), Climate change and global water resources: SRES emissions and socio-economic scenarios, *Global Environmental Change*, Vol. 14, No. 1, April 2004, 31-52.

Athanasiou, Tom, Baer, Paul (2002), *Dead Heat. Global Justice and Global Warming*, New York: Seven Stories Press.

Brouns, Bernd (2004), Was ist gerecht? Nutzungsrechte an natürlichen Ressourcen in der Klima- und Biodiversitätspolitik, *Wuppertal Paper No. 146*, Wuppertal: Wuppertal Institute.

Declaration on Human Duties and Responsibilities (1997). To be found at: www.cdp-hrc.uottawa.ca/publicat/valencia.

Donnelly, Jack (2003), *Universal Human Rights in Theory and Practice*, Ithaca: Cornell University Press.

Exeter Conference (2005), *Avoiding Dangerous Climate Change. A Scientific Symposium on Stabilisation of Greenhouse Gases*, Exeter, 1-3 February, 2005 (www.stabilisation2005.com).

Fischer, Günther, Mahendra Shah and Harrij van Velthuizen (2002), *Climate Change and Agricultural Vulnerability*, Laxenburg: IIASA.

Halweil, Brian (2005), The Irony of Climate, *World Watch Magazine*, March/April, 18-23.

Hare, William (2003), Assessment of Knowledge on Impacts of Climate Change – Contribution to the Specification of Art. 2 of the UNFCCC, Expert study for the German Advisory Council on Global Environmental Change for the Special Report *Climate Protection Strategies for the 21th Century: Kyoto and Beyond*, Berlin: WBGU.

Höffe, Otfried (1989), *Politische Gerechtigkeit*, Frankfurt: Suhrkamp.

Ignatieff, Michael (2001), *Human Rights as Politics and as Idolatry*, Princeton: Princeton University Press.

IPCC (2001), *Climate Change 2001. Synthesis Report*, Cambridge: Cambridge University Press.

Krebs, Angelika (2002), *Arbeit und Liebe. Die philosophischen Grundlagen sozialer Gerechtigkeit*, Frankfurt: Suhrkamp.

McMichael, A.J. *et al*. (2003), *Climate change and human health – risks and responses*, Geneva: WHO/UNEP/WMO.

Meyer, Aubrey (2000), *Contraction and Convergence. A Global Solution to Climate Change*,Totnes: Green Books.

Office of the UN High Commissioner on Human Rights (OHCHR) (2002), *Draft Guidelines: A Human Rights Approach to Poverty Reduction Strategies*, Geneva.

O'Neill, Onora (2000), *Bounds of Justice*, Cambridge: Cambridge University Press.

Ott, Hermann *et al*. (2004), *South-North Dialogue on Equity in the Greenhouse. A Proposal for an Adequate and Equitable Global Climate Regime*, Eschborn: GTZ, www.wupperinst.org/download/1085_proposal.pdf

Parry, M.L., C. Rosenzweig, A. Iglesias, M. Livermore, G. Fischer (2004), Effects of climate change on global food production under SRES emissions and socio-economic scenarios, *Global Environmental Change*, Vol. 14, No. 1, April 2004, 53-67.

Parry, Martin *et al*. (2001), Millions at Risk. Defining Critical Climate Change Threats and Targets, *Global Environmental Change* 11 (2001), 3, pp. 181-183.

Patz, J.A., D. Campbell-Lendrum, T. Holloway, and J.A. Foley, Impact of Regional Climate Change on Human Health, *Nature*, 438, 310-317.

Pogge, Thomas (2002), *World Poverty and Human Rights*, Cambridge: Polity Press.

Rayner, Steve, and Elizabeth Malone (1998), *Human Choice and Climate Change*, Vol. 4, Columbus: Batelle Press.

Sachs, Wolfgang (2003), Environment and Human Rights, *Wuppertal Paper No. 137*, Wuppertal: Wuppertal Institute.

Scheffler, Samuel (2001), *Boundaries and Allegiances. Problems of Justice and Responsibility in Liberal Thought*, Oxford.

Shue, Henry (1980), *Basic Rights. Subsistence, Affluence and U.S. Foreign Policy*, Princeton: Princeton University Press.

Steiner, Henry J., Alston, Philip (eds.) (1996), *International Human Rights in Context: Law, Politics, Morals*, Oxford: Oxford University Press.

UNCHR (United Nations Commission on Human Rights) (2005), *Rapport of the Special Rapporteur on the Right to Food*, Jean Ziegler, 24 January 2005, E/CN.4/2005/47.

UNDP (United Nations Development Programme) (2000), *Human Development Report 2000*, Oxford: Oxford University Press.

UNEP (United Nations Environment Programme) (2002), *Global Environment Outlook 3*, London: Earthscan.

Watt-Cloutier, Sheila (2004), *Climate Change and Human Rights*. www.carnegiecouncil.org/viewMedia.php/prmID/4445.

Wuppertal Institut (2005), *Fair Future. Begrenzte Ressourcen und globale Gerechtigkeit*, München: C.H. Beck.

PROMOTING HUMAN HEALTH AND FACING GLOBAL CHANGES: THE CUBAN EXPERIENCE[1]

ISMAEL CLARK ARXER

The 20th century has witnessed the globalization and internationalization of economy and society, the increasing application of scientific knowledge with the discovery of new technologies, and the evolution of economic, financial and commercial processes at planetary scale. All of those processes, in their common interaction, have experienced an exponential acceleration in its growth. In spite of the fact that those processes are presented as promoting human development, the reality we face at the beginning of the 21st century is that they have brought new and increasingly harsh crisis situations that periodically affect all of humanity (Meadows D.H. *et al.*, 1972).

Mainly in the last decade, this has been happening under economic principles sustained by the general application of the so called 'market laws'. Instigated by the principles of the Washington Consensus, the organizations of the Bretton Woods system established standards for economic development that were applied all around the world in very different local situations, often with well known negative results (Williamson J., 1999).

Now, with over five hundred years of growing globalization and the rule of capitalistic principles over the world economy, human society, although capable of increasing awareness of nature, of the universe that surrounds us, and of its own internal organization, faces a situation in which very few solutions seem possible to cope with the challenges that loom on the horizon, including challenges to the mere survival of our own species on this planet (Wallerstein, 1997).

[1] The paper was sent originally in Spanish and translated and read at the Pontifical Academy of Sciences by Sergio Jorge Pastrana.

We all know the results of colonialism and neo colonialism, which we continue to suffer even today. Until now, this process of indiscriminate appropriation of natural resources and control and manipulation of underdeveloped economies, far from disappearing, continues to go ahead by a variety of means, including war.

We live in a world that declares one thing and does another. Goals are established to balance the situation between the developed centre of the world and its underdeveloped surroundings, but in reality the policies and economic measures that are applied increase the concentration of wealth, the polarization of social groups, social exclusion, extreme poverty, and the squandering of the natural resources of the planet.

In this context, international institutions, governments, and the majority of the social forces in the world continue to be hypnotized by the mirage of the growth of the gross domestic product as the right measure and level to identify achievements in development. The concept that wealthier societies are those that achieve a higher human development is commonly and practically accepted with scarce dissent.

The presence of cyclic crises and the menace they represent for the survival of the human species call for alternative options to reverse those negative processes that make increasingly uncertain the long-term viability of human societies.

A closer look at the Cuban experience in the field of public health and human welfare in the last forty five years could help understand how to increase human development without necessarily looking first for some economic growth in order to increase afterwards that social development, but doing practically the opposite. In a significant manner, our conception is practically the opposite of those sustained by the supporters of total economic deregulation (Spiegel & Yassi, 2003).

In Cuba, since 1959, profound changes of political and social nature have promoted important human development goals, with special emphasis and well acknowledged results in the public health field (Gordon *et al.*, 2004), (WHC, 2002), sustained by systematic measures addressed to focus the principal efforts of all of Cuban society towards definite goals:

1) A high level of social equity (Galbraith, 2002). This has included legal, economic, and also educational measures for the elimination of any kind of discrimination due to race, gender, social origin, religion, etc., the overwhelming participation of women in social work, and the achievement of total employment. Moreover, it has also included the agrarian

and housing reforms, the equal distribution of economic results, and the organization of a vast system of social security.

2) The pre-eminent and sustained promotion of education, starting with the wiping out of illiteracy in one year, which at that moment reached practically 60% of the population, and also making all education public and totally free, up to the university level. Today the educational level of all of our population is at least of nine grades and there are almost nine hundred thousand university graduates, including more than seventy thousand doctors, in a population of little over eleven million persons. The ongoing process of bringing the university to the municipal level will soon, without doubt, multiply this achievement even more.

3) The articulation of a national health system, totally free of charge and universal, based on the systematic formation of human resources, that provides total coverage at the primary health care level, and which also has an effective secondary level, capable of reducing diseases and mortality, besides creating a third health care level of a high scientific and professional quality, provided with the best technological means in this field. This effort has also implied the establishing of an advanced infectious disease surveillance system, and the progressive transformation of the national health system into one of the cornerstones of all of the economic and social structure of our society.

4) The creation of a proficient national scientific capacity (Wagner *et al.*, 2001) based on a continuously sustained political will addressing the promotion and support of local scientific development, and articulated with the educational promotion of talents, and the creation of high level research institutions, functionally integrated with several other entities of the national economy, and all devoted to support both scientific research and the practical implementation of scientific results.

5) The promotion and practice of solidarity, with social relations depending on cooperation instead of competition, creating the conditions for promoting social ethics with a high sense of solidarity. This means, even today, identifying the human groups within society that either need more attention or are more vulnerable, in order to act adequately vis-à-vis each one of them, using all the social means at our reach (Alvarez & Máttar, 2004). At the same time, as a natural extension of this social ethics, it is of customary practice, and a state policy, to share what has been accomplished in Cuba with other less advanced societies.

A great deal of evidence can be supplied to show the interrelation, in the Cuban experience, of the abovementioned important factors. We shall mention only a couple of them:

– Over twenty thousand Cuban doctors, nurses and health personnel work as volunteers, providing direct health assistance to people in dozens of countries all over the world. Their work in health education, as well as in the task of prevention and in the provision of direct medical assistance, both at primary and secondary level, show in every case a positive direct impact in sensitive indicators, such as infant mortality rate.

– A new vaccine for Haemophilus influenzae, obtained from a synthetic antigen produced by a Cuban group of research has been developed recently in collaboration with a Canadian group, and has been added to the numerous vaccines and biotechnological products obtained in our country and presently used by other countries as well, among them: those aimed to prevent meningitis produced by meningococcus B, hepatitis type B, etc. Recently, the government of the United States of America gave authorization – very seldom given when Cuba is involved, as you probably know – to establish an agreement with an institution of that country devoted to testing and developing a therapeutic vaccine obtained by Cuban scientists for the treatment of a number of malignant cancer tumour processes.

– The national production of medical instruments and equipment, conceived basically to cover national needs at lower costs, has reached increasingly important export levels, and already includes electrocardiographs, defibrillators, vital functions monitors, equipments and means for immuno-enzymatic microanalysis, and modern neurophysiology and brain imaging devices.

These and other examples that could be mentioned are not intended to showcase an idealistic or apologetic vision of the Cuban experience. Though, all that has been achieved in Cuba in this field can be better assessed if we remember that our country continues to face an economic blockade that has lasted for over 45 years. This inhumane harassment has caused, until now, a calculated loss of not less than seventy nine thousand million dollars, not to mention the human suffering and calamities that our people have faced for so long due to its ever increasing, never decreasing, enforcement.

The purpose of these notes is to state our conviction, based on practice, that with relatively modest material resources it is possible to achieve substantial results whenever those resources are not limited by the application of senseless and inhuman economic models, but instead, the promotion of education and knowledge sharing, and the practice of social ethics based on solidarity are the guiding principles in the organization of society.

That seems to be, in our opinion, the only possible way for underdeveloped countries to face the challenges of the present processes of global change. At the same time, to work towards the elimination of the increasing divide in development is also in the best interest of rich countries, so it is imperative to request their contribution to those efforts in terms of ethics and moral standards.

REFERENCES

Álvarez, E. y J. Máttar (coords.), *Política social y reformas estructurales: Cuba a principios del siglo XXI*, Naciones Unidas, Comisión Económica para América Latina y el Caribe (CEPAL), Naciones Unidas, Programa de las Naciones Unidas par el desarrollo (PNUD) e Instituto Nacional de Investigaciones Económicas (INIE), 2004.

Galbraith, James K., A perfect crime: inequality in the age of globalization, *Daedalus*, Winter 2002. Issued as Volume 131, Number 1, pp. 11-25.

Gordon, B., R. Mackay & E. Rehfuess (eds.), *Inheriting the world: The atlas of children's health and the environment*, WHO, 2004.

Meadows, D.H. *et al.*, *Los limites del Crecimiento. Informe al Club de Roma sobre el Predicamento de la Humanidad*, Fondo de Cultura Economica, Mexico, 1972.

Spiegel, J.M. and A. Yassi, Lessons from the margins of globalization: appreciating the Cuban health paradox, *Journal of Public Health Policy*, Vol. 25, No. 1, pp. 85-110, 2003.

Wagner C., I. Brahmakulam, B. Jackson, A. Wong & T. Yoda, *Science and Technology Collaboration: Building Capacity in Developing Countries*, – MR 1357.0, 2001

Wallerstein, I. (1997), Ecology and the Capitalist Cost of Production: No Exit, Keynote address at PEWS XXI, *The Global Environment and the World-System*, Univ. of California, Santa Cruz, Apr. 3-5, 1997.

Williamson, J. (1999), What should the bank think about the Washington consensus?, Paper prepared as a background to the World Bank's *World Development Report* 2000, July 1999.

World Health Chart 2002. Downloaded from www.whc.ki.se.

GLOBAL ENVIRONMENTAL CHANGE AND HUMAN HEALTH

M.O. ANDREAE, U.E.C. CONFALONIERI, A.J. McMICHAEL, D.E. BLOOM, D.L. HEYMANN, P. MARTENS, W. STEFFEN & M.E. WILSON

The Earth's environment is showing signs of strain from a burgeoning world population and its impact on the planet's life support systems, creating headlines almost every day: increasingly crowded subsistence farming of pigs and poultry in southern China and Southeast Asia makes the emergence of new strains of avian influenza more likely, escalating the risk of new epidemics in humans. Climate change over the preceding three decades doubles the probability of severe heat waves in Western Europe, and in August 2003, a heat wave in Europe kills in excess of 30,000 people. Coastal deforestation and the clearing of mangrove forests exacerbate the impact of the 2004 tsunami in Asia. Poverty-driven patterns of human settlement in Central America multiply the damage to life, limb, and infrastructure caused by Hurricane Mitch and Hurricane Stan. The list goes on…

These amplifications of environmental impacts on human populations are facets of the ongoing, now dramatic, extension in both scale and type of the human impact on the natural environment, and the resultant risk to health. The sharp rise in human population in the 20th century, accompanied by energy and resource intensive economic development, has rendered this human impact so large and pervasive that it now entails perturbation of the Earth System itself (Andreae *et al.*, 2004; Steffen and Lambin, 2006). These disruptive changes to natural biogeophysical systems weaken Earth's life-support systems, and are referred to as Global Environmental Change (GEC). In this paper, we will use the term GEC to represent the biological, chemical, and geophysical aspects of Global Change, i.e., the pervasive change in the Earth System that also includes the social, economic and political aspects (Steffen *et al.*, 2004b).

1. Global Environmental Changes

The best known component of GEC is global climate change, which, like stratospheric ozone depletion, affects the entire planet. Diverse sources of greenhouse gases (GHG) around the world – power-generating plants, factories, vehicles, wet farmlands, deforested areas, and others – change the composition and, hence, the heat-trapping properties of the lower atmosphere. The resultant change in climate occurs at a global scale, affecting populations everywhere, notwithstanding the many original local sources of the gaseous emissions. One of the great problems that stands in the way of finding a course of action to mitigate climate change is just this disconnect between local cause and effect. Because it is only the integrated effect of a vast number of local emissions that results in climate change, it becomes difficult for individuals and policymakers to associate their actions (emissions) with their consequences (climate change). In fact, it appears that the populations most vulnerable to human-induced climate change are not those that cause most of the GHG emissions, but people in the developing world who contribute minimally to the atmospheric burden of these gases (Shah *et al.*, 2006).

Along with these globally integrated atmospheric changes, human actions are also causing marked changes in the local, regional, and global cycles of nitrogen and phosphorus, and have initiated the global dissemination (especially to higher, non-industrial, latitudes) of semi-volatile persistent organic chemical pollutants (POPs). Various other worldwide environmental changes also reflect the escalating extent and intensity of human pressure on the global environment. These include disruptions to ecosystems, land degradation, biodiversity losses, depletion of freshwater resources, and critical pressures on ocean fisheries (Schellnhuber *et al.*, 2004; Dobson, 2006; Meybeck, 2006).

The complexity and scale of these various GECs, which often entail changes to ecological and geophysical processes, make the identification and quantification of the resultant health risks difficult. This paper explores the relationships between various types of global change, their social, demographic and economic drivers, and their impacts on population health. It suggests some areas for future research, and proposes policy guidelines for the achievement of sustainable development and, hence, the reduction in risks (present and future) to human health that result from existing practices.

While awareness of the processes of globalization and global environmental change has increased in recent years, there has been insufficient

attention paid to the interconnectedness and likely consequences of these processes. These large-scale changes represent a new dimension to the state of the world in which we live – an era now referred to by some as the Anthropocene, due to the overwhelming dominance and influence of the human species (Steffen and Lambin, 2006). The changes reflect the continued expansion of the human population, the magnitude and growing intensity of economic activity (including escalating levels of consumption and waste generation), and the associated impacts on social structures, wealth distribution, geopolitical relations, and environmental systems and resources.

Much of the recent discourse on 'sustainability' has focused on whether humans can maintain current levels of social and economic activity without depleting the natural environmental resource base. There has been little recognition of the risks that GEC pose to the health, perhaps even survival, of human populations, both present and future. Yet (viewed anthropocentrically) the reason for seeking to optimize social structures, environmental integrity and economic productivity is essentially to improve human well-being, health and survival. We ought, then, to recalibrate our discussion of 'sustainability' to take explicitly into account the implications of global environmental changes for human health.

2. Drivers of Social and Environmental Changes, Interactions, and Health Impacts

The continuing global population growth (Lutz, 2006), increasing levels of material wealth and consumption, prevailing technologies, and aspects of globalization (increases in human inter-connectivity – economic, physical, cultural, microbial, electronic, etc.) are the fundamental drivers of these environmental changes. Over recent decades neo-liberal market-driven economic policies, with reduced governmental controls on industrial pollution, have amplified these environmental pressures further. Meanwhile, these underlying large-scale drivers are also having increasingly pervasive effects on social, economic and political conditions around the world.

Much attention has been paid to how urbanization, social change, trade liberalization, and environmental change affect outcomes such as social relations and community cohesion, economic development, levels of poverty, air quality, and food production and distribution. Less attention has

been paid to how these changes affect human population health. Yet, the trend in human health, observed over decadal time, is clearly a key indicator of whether society at large has achieved a sustainable way of managing the natural and social environments (McMichael, 2002; Huynen and Martens, 2006). The huge disparity in economic wealth across nations, and the resultant inequitable distribution of public health status can be visualized in the form of 'cartograms' (Gastner and Newman, 2004), where countries are drawn in a size proportional to population, gross domestic product (GDP), and child mortality (Figure 1, see page 420).

Over the short term, material conditions have improved in most populations, and average life expectancy has continued to rise. However, over the past decade, life expectancy has decreased in a dozen or more countries (McMichael *et al.*, 2004); infectious diseases seem to have become more labile in distribution, resurgence, and emergence (Morens *et al.*, 2004; Weiss and McMichael, 2004; McMichael, 2006); and the total number of malnourished persons in the world has risen slightly over the past half-decade, after declining during the 1990s (FAO, 2005). Meanwhile, there are rising concerns about recent trends in many material environmental indicators – freshwater availability, urban air quality, supplies of energy (the 'peak oil' debate), and the prospects for continuing to feed the world population on the basis of sustainable production methods.

Figure 2 provides a simple representation of how the various major components of 'global change' – encompassing changes in demographic patterns, social-cultural relations, the economy, technology, and the environment – affect population health. The interactions between these components are also shown.

A fuller model would incorporate the ways in which the health subsystem may feed back into the socioeconomic subsystem and ultimately into GEC. The interconnected components of global change, with the added dimension of health feedback, are shown conceptually in Figure 3.

The implications of population health deficits for economic productivity and, in turn, socio-economic development have recently been examined in detail (Bloom and Canning, 2000; Commission on Macroeconomics and Health, 2001; Sachs and Malaney, 2002). The unabated burden of malaria in Africa has been calculated to have halved the growth rate of income per capita during 1965 through 1990 (Gallup and Sachs, 2001). In China, air pollution takes a high toll on human health and thereby also causes large economic costs (Brajer and Mead, 2004). Finally, the extreme example of the ballooning impact of HIV/AIDS in Sub-Saharan Africa on economic

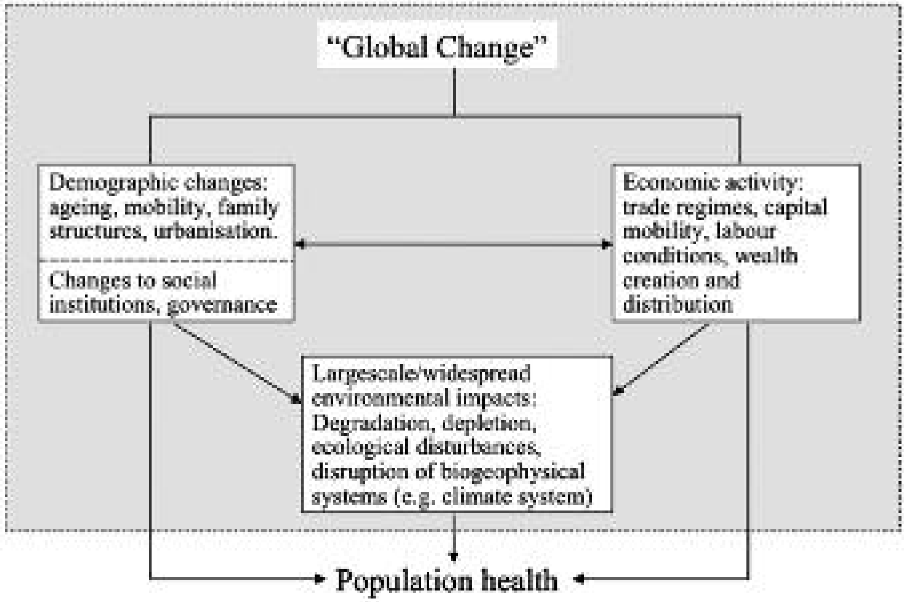


Figure 2. Schematic representation of the main components of 'global change' (comprising the area shown against shaded background), and the paths by which they affect human health and disease.

Figure 3. Relationships between socioeconomic conditions, environmental changes, and human health, showing the complexity of this system.

productivity, as the ranks of teachers, farmers, health workers, and others are depleted, is now well known (Piot *et al.*, 2001). The initial impact may be of an arithmetic kind, but the future impact may increase exponentially due to the erosion of social, educational and cultural institutions.

While poor population health can impair society's overall economic advance, the typically uneven distribution of poor health within the population often exacerbates poverty. Poverty, especially in combination with

population, increases the pressure on the natural environment. These interactive relationships have recently been examined in detail by the Millennium Ecosystem Assessment (Hassan *et al.*, 2005).

Ultimately, the outcome of the interactive processes between human society, ecosystems and climate will depend on the resilience of social and economic systems, ecosystems, and climate (Butler *et al.*, 2005; Carmichael, 2006; Huynen and Martens, 2006; Jäger, 2006). Limited health effects and ecosystem disruptions, coupled with appropriate social responses, can be accommodated in scenarios where population health improves in both poor and wealthy populations. On the other hand, a combination of degraded ecosystem services and poor governance could lead to 'system failure' with dramatic harm to human health and well-being, precipitating further collapse of social and economic systems (Butler *et al.*, 2005). This, in turn, can be expected to have major consequences for climate scenarios. The highly coupled character of the system depicted in Figures 2 and 3 implies the probability of strongly non-linear responses, including abrupt 'state transitions', as we enter previously unexplored regimes of the Earth System as a consequence of human perturbations (Steffen *et al.*, 2004a; Steffen and Lambin, 2006).

3. Effects of Global Environmental Changes on Human Health

The escalating levels of environmental change are having observable effects on many plant and animal species (Houghton *et al.*, 2001; Hassan *et al.*, 2005). For example, global climate change over the past quarter-century has affected the seasonal cycles, geographic distribution, numbers, and, in some cases, survival of plant and animal species (Parmesan and Yohe, 2003; Root *et al.*, 2003). In some cases, e.g., extinctions of amphibians from infectious disease, a clear connection between climate change and animal disease has been identified (Pounds *et al.*, 2006). In human populations, however, patterns of health and disease are affected by many social, environmental, behavioral, health-care and other factors. Hence, the attributable impacts of GEC on human health are less clear-cut than in non-human systems. Still, a recent assessment by the World Health Organization (WHO) has estimated that the climate change which has occurred by the year 2000 (about 0.7°C over the last 100 years) is already responsible for over 150,000 deaths and ca. 5 million 'disability-adjusted life years' per year (Figure 4, see page 421) (World Health Organization, 2002).

Changes in the environment affect human health via diverse pathways. This is well illustrated by the diverse ways in which regional climate change

and land-use change affect infectious disease patterns (Foley *et al.*, 2005; Patz *et al.*, 2005; Confalonieri *et al.*, 2006; Heymann, 2006; McMichael, 2006). Global patterns of parasitic and infectious disease species diversity show close connections to climate and total species diversity, with maxima in tropical regions (Guernier *et al.*, 2004). Changes in biological and ecological relationships often modify the distribution and behavior of vector species and intermediate host species, thereby altering the transmission of infectious disease (Dobson, 2006). Examples include changes in the patterns of tick-borne encephalitis in Sweden and Lyme disease in the northeastern United States, and the emergence of Nipah virus disease in humans as a result of the combination of intensified pig-farming, increasing deforestation, and altered ecology in Malaysia (Lindgren *et al.*, 2000; Ostfeld and Keesing, 2000; Harvell *et al.*, 2002; Weiss and McMichael, 2004; Olival and Daszak, 2005; Eaton *et al.*, 2006).

The rate of entry of microbes, both novel and previously encountered, into human populations is increased both where high human-population densities are in close contact with animal reservoirs of infectious disease, and where food production methods have been intensified (Weiss and McMichael, 2004). Pathogens can be exchanged between domestic and free-living wild animals in both directions, and can be passed to humans from both domestic and wild animals. This transfer of pathogenic microbes between animals and humans, and the resulting potential for emerging infectious diseases, poses a serious threat to human health (Daszak *et al.*, 2000; Wolfe *et al.*, 2005).

Meanwhile, malnutrition – which remains widespread in many developing countries and is sometimes exacerbated by environmental changes including land degradation, fisheries depletion, and climatic extremes – creates large immune-compromised populations into which infectious diseases spread more easily.

Global environmental changes also affect the occurrence of non-infectious diseases and the risk of physical injury and death. A very important example is that of malnutrition and associated disorders as a result of the combined stresses on agricultural yields from various forms of environmental change – including land degradation, freshwater shortages, climate change, and altered patterns of plant pests and diseases (Gregory *et al.*, 2005; Meybeck, 2006). Models that forecast the effect of climate change on agricultural yields over coming decades indicate a pattern of gains in developed countries and declines in developing countries, where food production is often already insufficient. According to the WHO,

about 800 million people are presently undernourished, with almost half of them living in Africa (World Health Organization, 2002). The anticipated nutrition declines in the developing world would further exacerbate malnutrition and the risk of famine (Shah *et al.*, 2006).

The risk of injuries and deaths increases during extreme weather events (floods, hurricanes, etc.) and very hot weather (McMichael *et al.*, 2006). Extreme events, such as the European heat wave of 2003, which caused 22,000-45,000 excess deaths during a two-week period (Patz *et al.*, 2005), are expected to increase during the 21st century as the average temperature of the Earth increases and climatic patterns become more variable (Stott *et al.*, 2004).

The warming effect of anthropogenic GHG has been offset partly by the net cooling effect of atmospheric aerosol particles, which are released primarily by burning fossil fuel and biomass (Houghton *et al.*, 2001; Andreae *et al.*, 2005). Aerosols also affect the behavior of clouds and thereby the formation of rainfall, with as yet unknown consequences for water supply in susceptible regions. Aerosol particles also cause diverse adverse health effects, contributing to both non-infectious and infectious diseases, with resultant global excess mortalities in the millions annually (Pope *et al.*, 2002; Smith and Mehta, 2003; Schwartz, 2004). It has recently been estimated that exposure to aerosol pollution reduces the average life expectancy by 8.1 months in the 25 EU countries, reaching as high as 13.2 months in Belgium (Amann *et al.*, 2005). Aerosol pollution is not limited to industrialized areas, and now covers vast regions of the world, particularly in Asia and other developing regions. The resulting regional haze from anthropogenic aerosols also reduces solar radiation and thereby reduces crop yields, promoting malnutrition and associated health risks (Chameides *et al.*, 1999). The future dilemma, discussed below, is that a reduction in aerosol pollution will also accelerate global warming.

From a socio-economic perspective, the ecology of modern urban living has changed the calculus of benefits and risks to health. One example, now being widely discussed, is the surge in prevalence of obesity in many countries, especially in urban populations. This population-level phenomenon reflects the ready access of a community to energy-dense foods and to a pattern of daily living with diminished need for physical activity. This surge in obesity is occurring especially (and without precedent) among children and teenagers; it will result in a rise in the incidence of serious adult diseases (diabetes, cardiovascular disease, and others) and, very likely, a decrement in that generation’s average life expectancy (Olshansky *et al.*, 2005).

4. Effects of Changes in Population Health on the Environment

In the previous section, we have discussed how environmental changes affect human health. Influences in the opposite direction, i.e., the effects of changes in population health on the global environment and climate, are less obvious and less well understood. Yet, there is evidence for important, though indirect, causal connections that link environmental change to population health and growth. Historical analysis of the growth of the human population, from prehistory to the present, indicates that agricultural expansion and deforestation has resulted in increases in atmospheric CO_2 and methane. Conversely, massive pandemics, such as the one that followed European colonization of the Americas and the resulting introduction of infectious diseases, have led to regrowth of forests and a concomitant drop in atmospheric carbon dioxide concentration (Ruddiman, 2003; Ruddiman and Carmichael, 2006).

Looking from the past into the future, we see that policies intended to improve human health by tightening aerosol emission regulations may have negative climate effects. As discussed in the previous section, the growing awareness that aerosols impair human health results in increasing pressure for regulations limiting aerosol emissions. These regulations, however, would also eliminate the aerosol's cooling effect on climate. As a consequence, greenhouse-induced climate change may actually accelerate as a result of policy decisions on aerosols that are primarily motivated by public-health concerns. Much preferable would be a balanced policy that incorporates the parallel reduction of both greenhouse gas and aerosol emissions (Andreae *et al.*, 2005).

At the most fundamental level, the pervasive improvement in public health in the 19th and 20th centuries has ultimately made possible the massive global change that the world is experiencing at present. Improvements in health care, urban sanitation, domestic hygiene, nutrition, and literacy have resulted in greatly reduced infant/child mortality and have facilitated rapid growth in the human population. By reducing costs associated with mortality and disease, they have also allowed increasing accumulation of wealth in the hands of individuals, enabling the development of a consumer society. This, together with an increasingly energy-intensive and carbon-intensive economy over the past century, has caused the rapid build-up of anthropogenic greenhouse gases in Earth's atmosphere.

The underlying relationship between demographic and economic changes and environmental impacts was originally expressed by Ehrlich and Holdren in the form of the 'IPAT Equation' (Ehrlich and Holdren,

1971). Although this simple equation cannot adequately express the complexities of interactions between humans and the environment (and makes no explicit reference to 'health'), it illustrates the fundamental quantitative relationship between population size and vigor, human economic activities, and environmental impact (Chertow, 2001).

IPAT is an identity stating that environmental impact (I) is the product of human population size (P), level of affluence (A), and type of technology (T):

$$I = P \cdot A \cdot T$$

In the context of CO_2-driven climate change, this equation is reformulated as:

Carbon emission = Population * (Unit GDP/capita) * [(Unit CO_2)/(Unit Energy) * (Unit Energy)/(Unit GDP)]

Although human population health is not present explicitly in the IPAT identity, its indirect influence via socioeconomic effects is quite profound (see also Figure 2). Health effects are parameterized in two terms of this equation. First, it is implicit in the population term, since public health acts directly on demographic structure and population size. Second, because of the connection between health and wealth (Bloom and Canning, 2006), the second term, per capita GDP, is also an implicit function of population health. Ultimately, the complex processes hiding behind the simple conceptual relationship reflected in the IPAT identity will have to be implemented explicitly in Earth System models.

5. Research and Surveillance: Indicators, Scenarios, Modeling

A prime research task is to elucidate further the fundamental relationships between environmental change, socio-economic processes, and population health. For this protean research task, suitable population-level indicators of human well-being and health are needed. However, few such indicators exist.

Mortality, for which data are plentiful (albeit often incomplete from poorer countries), is often an unsatisfactory measure of health and well-being. Life expectancy, while being a better metric, is often dominated by early-age deaths; furthermore, it does not shed light directly on sickness, disability, and overall well-being. More useful and better-integrated indicators of health and well-being are needed, along with indicators of vulnerability.

These would incorporate information about where people live (geographical conditions), resource availability (nutrition, water, energy), socio-economic conditions (including institutions that support and protect people), health status (life expectancy, disability, disease incidence), as well as people's perceptions about their living conditions (Heymann, 2006; Jäger, 2006).

The relationships between environment, socio-economic conditions, and health now extend across an unprecedentedly large span of time and space. The climate change process, in particular, spans decades, perhaps centuries. Hence, many of the health consequences of today's actions and their environmental impacts will only be realized well into the future. Once set in motion, however, these slow but massive environmental changes are not easily reversed. Some effects, such as biodiversity losses, ecosystem collapse and topsoil loss, are irreversible on human time scales.

This unfamiliar situation, unprecedented in human experience but now gathering momentum at an increasingly global scale, requires the development of research and evaluation tools to 'look into the future' – i.e., Earth System models that incorporate the interactions between climate and the human dimension, including health (McMichael, 2006). In the simplest case, population health can be part of economic and population submodels, the results of which can be used as input scenarios for climate system models. In contrast to the very simplistic scenarios used today as input to climate projections (Nakicenovic and Swart, 2000), they should include potential future risks (e.g., pandemics, famine, migration, and conflict) (Huynen and Martens, 2006). The climate projections so obtained can then be used as forcings for socioeconomic models.

This approach is only valid, however, as long as climate feedbacks on the socioeconomic part of the Earth System are weak compared to forcings inside the human subsystem. If, however, these feedbacks turn out to be strong, fully coupled models would be required, a challenge that goes way beyond what is possible today. These models will have to be developed, probably using EMICs (Earth System Models of Intermediate Complexity) as an initial step, and then tested with environmental, socioeconomic, and health data. They can then be applied to look for future hazards to the well-being of humanity and the environment, and to explore paths to sustainable development. While these complex models are being developed, progress can be made with simplified approaches to connect health and climate, e.g., in relation to climate change and future patterns of transmissibility of mosquito-borne infections, such as dengue fever (Hales *et al.*, 2002).

6. Policy Recommendations

Although much new research is needed to elucidate these complex contemporary environmental problems, some policy guidelines can already be identified. These include:

– Recognition of the fundamental significance of population health within the 'sustainability' policy framework

– Inter-governmental commitment to change and cooperation, with the goal of reducing the rate and magnitude of GEC

– An understanding that shifts to new technologies, which create more economic and social advantage than disadvantage

– Particular commitment to economic and related policies that reduce material inequalities and, hence, vulnerabilities

– Improvements in international systems of infectious disease surveillance and control

– New, bold, forms of interdisciplinary research – essential to the tackling of complex environmental-change phenomena and their social, cultural and political remediation.

Sustainable development is, in the long run, defined by its capacity to sustain human health and well-being. It is characterized by environmental practices that maintain the integrity and productivity of natural and managed ecosystems, a healthy economy, full employment, comfortable material living standards, efficient social security and assistance, good educational systems, adequate public infrastructure, and political freedom and stability (Carmichael, 2006; Jäger, 2006). Attaining these population health-promoting conditions will demand technologies, community behaviors and inter-governmental agreements that conserve the ability of the world's natural environment to support human life (Jäger, 2006).

This collective, global, task will require shared insight, commitment, resources, and political will among nations. The improvement of social, economic, and environmental conditions is inherent to attainment of the eight U.N. Millennium Development Goals, which range from halving extreme poverty to halting the spread of HIV/AIDS and providing universal primary education, all by the target date of 2015 (United Nations, 2005). Further, because the adverse impacts of environmental change (such as a reduction of agricultural yield) are most likely to occur in already disadvantaged regions, an emphasis on achieving greater international social and economic equality is essential (Jäger, 2006; Sachs, 2006). This includes the immediate need to fund ongoing programs to

contain or eradicate infectious diseases, such as polio, and to reduce other major scourges such as malaria, tuberculosis, HIV/AIDS, and child diarrhea. These take a huge toll on human life, especially in developing countries (Heymann, 2006).

A key requirement in operational efforts to prevent and limit health losses, especially infectious disease epidemics and pandemics, is the availability of coordinated international surveillance and effective early-warning systems (Heymann, 2006; Jäger, 2006). The recent experiences of severe acute respiratory syndrome (SARS) and avian influenza, as unexpected causes of human infection and death, underscore this need. The existing regional infectious disease surveillance systems should be upgraded and put to optimal use within a better-coordinated international framework. To predict the spread of disease once it has been detected by such surveillance systems, improved spread models are required that take into account the complex patterns of biological, social, and spatial factors involved (Ferguson *et al.*, 2003; Hufnagel *et al.*, 2004; Brockmann *et al.*, 2006). Given effective early warning systems and a strategy for prompt intervention, it may be possible to contain emergent pandemics before they can cause massive damage to human health and world economics (Ferguson *et al.*, 2005; Longini *et al.*, 2005) – or at least postpone the emergence of a pandemic (Mills *et al.*, 2006).

Meanwhile, gains in wealth must be accompanied by measures to reduce the longer-term environmental consequences of increases in population size, affluence and consumption (Jäger, 2006). Energy-inefficient, high-throughput, waste-generating economies must be reoriented towards low environmental impact and the recycling of materials. Industrial and commercial activity that depends on throw-away items generates enormous emissions and other wastes that contribute to environmental degradation and its risks to health.

Emissions of greenhouse gases and atmospheric aerosols offer one complex example of how human-health risks must be taken into account alongside consideration of environmental consequences *per se*. Health-motivated reductions in aerosol emissions without simultaneous cuts in CO_2 output, as would be achieved for instance by stack scrubbing technology, would improve health but also increase global warming, because the cooling effect of the aerosols would be reduced. Aerosol emission cuts must therefore be accompanied by accelerated reduction of greenhouse gas emissions, particularly CO_2 (Andreae *et al.*, 2005). A special case is black carbon aerosol, emitted particularly from diesel engines. Reduction

of black carbon emissions, which in contrast to other aerosol types has a warming effect on climate, would simultaneously cut back on climate change and improve health (Hansen and Nazarenko, 2004). This illustrates the type of 'win-win' solutions that are often possible. Co-benefits are also obtained from measures to reduce CO_2 emissions, because they always lead to simultaneous reductions in the release of SO_2, NO_x, and other pollutants, with associated health and socio-economic benefits (Aunan *et al.*, 2004; van Vuuren *et al.*, 2006).

Finally, our incomplete scientific understanding of the complex relationships discussed in this paper must be advanced. This requires the integration of research in the natural sciences and the health sciences with that of the humanities and social-economic sciences (McMichael *et al.*, 2003; McMichael, 2006). Conceptual gaps between these disciplines must be bridged, and the techniques of research and analysis made compatible among disciplines. A start in this direction has been made by the Earth System Science Partnership, a joint activity of the current global change programs in climate change (World Climate Research Programme), geosphere/biosphere interactions (International Geosphere/Biosphere Programme), biodiversity (Diversitas), and human dimensions (International Human Dimensions Program).

The increased engagement of researchers in this domain will add a crucial dimension to the evolving policy debate about climate change and other large-scale environmental changes. Good interdisciplinary research will clarify the extent to which the impacts of those changes include potentially serious health/survival consequences for human communities. This will extend the policy discourse beyond independent consideration of economic disturbance, loss of environmental amenity, threats to species, and risks to built infrastructure. It will thus help us understand better the real meaning and prerequisites of 'sustainability'.

REFERENCES

Amann, M., I. Bertok, R. Cabala, J. Cofala, C. Heyes, F. Gyarfas, Z. Klimont, W. Schöpp, and F. Wagner (2005), *A final set of scenarios for the Clean Air For Europe (CAFE) programme (CAFE Scenario Analysis Report Nr. 6)*, International Institute for Applied Systems Analysis (IIASA), 105 p.

Andreae, M.O., C.D. Jones, and P.M. Cox, Strong present-day aerosol cooling implies a hot future, *Nature*, 435, 1187-1190, 2005.

Andreae, M.O., L. Talaue-McManus, and P.A. Matson, Anthropogenic modification of land, coastal and atmospheric systems as threats to the functioning of the Earth System, in *Earth System Analysis for Sustainability*, edited by H.-J. Schellnhuber, P.J. Crutzen, W.C. Clark, M. Claussen, & H. Held, pp. 245-264, The MIT Press, Cambridge, Mass., 2004.

Aunan, K., J.H. Fang, H. Vennemo, K. Oye, and H.M. Seip, Co-benefits of climate policy – lessons learned from a study in Shanxi, China, *Energy Policy*, 32, 567-581, 2004.

Bloom, D.E., and D. Canning, The health and wealth of nations, *Science*, 287, 1207-1209, 2000.

Bloom, D.E., and D. Canning, Epidemics and economics, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 304-331, The Pontifical Academy of Sciences, Vatican City, 2006.

Brajer, V., and R.W. Mead, Valuing air pollution mortality in China's cities, *Urban Studies*, 41, 1567-1585, 2004.

Brockmann, D., L. Hufnagel, and T. Geisel, The scaling laws of human travel, *Nature*, 439, 462-465, 2006.

Butler, C.D., C.F. Corvalan, and H.S. Koren, Human health, well-being, and global ecological scenarios, *Ecosystems*, 8, 153-162, 2005.

Carmichael, A.G., Infectious disease and human agency: An historical overview, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 3-46, The Pontifical Academy of Sciences, Vatican City, 2006.

Chameides, W.L., H. Yu, S.C. Liu, M. Bergin, X. Zhou, L. Mearns, G. Wang, C.S. Kiang, R.D. Saylor, C. Luo, Y. Huang, A. Steiner, and F. Giorgi, Case study of the effects of atmospheric aerosols and regional haze on agriculture: An opportunity to enhance crop yields in China through emission controls?, *Proceedings of the National Academy of Sciences of the United States of America*, 96, 13,626-13,633, 1999.

Chertow, M.R., The IPAT equation and its variants, *Journal of Industrial Ecology*, 4, 13-29, 2001.

Commission on Macroeconomics and Health (2001), *Macroeconomics and health: Investing in health for economic development*, World Health Organization, 128 p.

Confalonieri, U.E.C., M. Wilson, and A. Najar, Social and environmental vulnerability of emerging infectious diseases, in *Interactions Between*

Global Change and Human Health (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 195-212, The Pontifical Academy of Sciences, Vatican City, 2006.

Daszak, P., A.A. Cunningham, and A.D. Hyatt, Wildlife ecology – Emerging infectious diseases of wildlife – Threats to biodiversity and human health, *Science*, 287, 443-449, 2000.

Dobson, A., Biodiversity and infectious disease: Why we need Nature, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 145-157, The Pontifical Academy of Sciences, Vatican City, 2006.

Eaton, B.T., C.C. Broder, D. Middleton, and L.F. Wang, Hendra and Nipah viruses: different and dangerous, *Nature Reviews Microbiology*, 4, 23-35, 2006.

Ehrlich, P., and J. Holdren, Impact of population growth, *Science*, 171, 1212-1217, 1971.

FAO (2005), *The State of Food Insecurity in the World*, FAO, 36 p.

Ferguson, N.M., D.A.T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D.S. Burke, Strategies for containing an emerging influenza pandemic in Southeast Asia, *Nature*, 437, 209-214, 2005.

Ferguson, N.M., M.J. Keeling, W.J. Edmunds, R. Gant, B.T. Grenfell, R.M. Amderson, and S. Leach, Planning for smallpox outbreaks, *Nature*, 425, 681-685, 2003.

Foley, J.A., R. DeFries, G.P. Asner, C. Barford, G. Bonan, S.R. Carpenter, F.S. Chapin, M.T. Coe, G.C. Daily, H.K. Gibbs, J.H. Helkowski, T. Holloway, E.A. Howard, C.J. Kucharik, C. Monfreda, J.A. Patz, I.C. Prentice, N. Ramankutty, and P.K. Snyder, Global consequences of land use, *Science*, 309, 570-574, 2005.

Gallup, J.L., and J.D. Sachs, The economic burden of malaria, *American Journal of Tropical Medicine and Hygiene*, 64, 85-96, 2001.

Gastner, M.T., and M.E.J. Newman, Diffusion-based method for producing density-equalizing maps, *Proceedings of the National Academy of Sciences of the United States of America*, 101, 7499-7504, 2004.

Gregory, P.J., J.S.I. Ingram, and M. Brklacich, Climate change and food security, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 360, 2139-2148, 2005.

Guernier, V., M.E. Hochberg, and J.F.O. Guegan, Ecology drives the worldwide distribution of human diseases, *PLoS Biology*, 2, 740-746, 2004.

Hales, S., N. de Wet, J. Maindonald, and A. Woodward, Potential effect of population and climate changes on global distribution of dengue fever: an empirical model, *The Lancet*, 360, 830-834, 2002.

Hansen, J., and L. Nazarenko, Soot climate forcing via snow and ice albedos, *Proceedings of the National Academy of Sciences of the United States of America*, 101, 423-428, 2004.

Harvell, C.D., C.E. Mitchell, J.R. Ward, S. Altizer, A.P. Dobson, R.S. Ostfeld, and M.D. Samuel, Climate warming and disease risks for terrestrial and marine biota, *Science*, 296, 2158-2162, 2002.

Hassan, R., R. Scholes, and N. Ash (ed.), (2005), *Ecosystems and human well-being. Volume 1: Current state and trends: Findings of the Condition and Trends Working Group* (The Millennium Ecosystem Assessment Series, vol. 1).

Heymann, D. L., Effects of social, environmental and economic factors on current and future patterns of infectious diseases, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 290-303, The Pontifical Academy of Sciences, Vatican City, 2006.

Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (ed.), (2001), *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, UK, and New York, NY, USA: Cambridge University Press, 881 p.

Hufnagel, L., D. Brockmann, and T. Geisel, Forecast and control of epidemics in a globalized world, *Proceedings of the National Academy of Sciences of the United States of America*, 101, 15,124-15,129, 2004.

Huynen, M., and P. Martens, Globalisation and human health: towards scenarios for the 21st century, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 269-289, The Pontifical Academy of Sciences, Vatican City, 2006.

Jäger, C.C., Global governance for human health in an evolving environment, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 332-348, The Pontifical Academy of Sciences, Vatican City, 2006.

Lindgren, E., L. Talleklint, and T. Polfeldt, Impact of climatic change on the northern latitude limit and population density of the disease-transmitting European tick Ixodes ricinus, *Environmental Health Perspectives*, 108, 119-123, 2000.

Longini, I.M., A. Nizam, S.F. Xu, K. Ungchusak, W. Hanshaoworakul, D.A.T. Cummings, and M.E. Halloran, Containing pandemic influenza at the source, *Science*, 309, 1083-1087, 2005.

Lutz, W., Global demographic trends, education, and health, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 252-268, The Pontifical Academy of Sciences, Vatican City, 2006.

McMichael, A.J., The biosphere, human health and sustainability, *Science*, 297, 1093, 2002.

McMichael, A.J., Our changing biogeophysical and socioeconomic environments: influence on infectious diseases, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 213-229, The Pontifical Academy of Sciences, Vatican City, 2006.

McMichael, A.J., C.D. Butler, and C. Folke, New visions for addressing sustainability, *Science*, 302, 1919-1920, 2003.

McMichael, A.J., M. McKee, V. Shkolnikov, and T. Valkonen, Mortality trends and setbacks: global convergence or divergence?, *The Lancet*, 363, 1155-1159, 2004.

McMichael, A.J., R.E. Woodruff, and S. Hales, Climate change and human health: present and future risks, *The Lancet*, 367, doi:10.1016/S0140-6736(06)68079-3, 2006.

Meybeck, M., Global changes in aquatic systems and their interrelations with human health, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 61-111, The Pontifical Academy of Sciences, Vatican City, 2006.

Mills, C.E., J.M. Robins, C.T. Bergstrom, and M. Lipsitch, Pandemic influenza: Risk of multiple introductions and the need to prepare for them, *PLoS Med*, 3, e135, doi:10.1371/journal.pmed.0030135, 2006.

Morens, D.M., G.K. Folkers, and A.S. Fauci, The challenge of emerging and re-emerging infectious diseases, *Nature*, 430, 242-249, 2004.

Nakicenovic, N., and R. Swart (ed.), (2000), *Special Report on Emissions Scenarios*, Cambridge, United Kingdom: Cambridge University Press, 612 p.

Olival, K.J., and P. Daszak, The ecology of emerging neurotropic viruses, *Journal of Neurovirology*, 11, 441-446, 2005.

Olshansky, S.J., D.J. Passaro, R.C. Hershow, J. Layden, B.A. Carnes, J. Brody, L. Hayflick, R.N. Butler, D.B. Allison, and D.S. Ludwig, A

potential decline in life expectancy in the United States in the 21st century, *New England Journal of Medicine*, 352, 1138-1145, 2005.

Ostfeld, R.S., and F. Keesing, Biodiversity and disease risk: The case of lyme disease, *Conservation Biology*, 14, 722-728, 2000.

Parmesan, C., and G. Yohe, A globally coherent fingerprint of climate change impacts across natural systems, *Nature*, 421, 37-42, 2003.

Patz, J.A., D. Campbell-Lendrum, T. Holloway, and J.A. Foley, Impact of regional climate change on human health, *Nature*, 438, 310-317, 2005.

Piot, P., M. Bartos, P.D. Ghys, N. Walker, and B. Schwartlander, The global impact of HIV/AIDS, *Nature*, 410, 968-973, 2001.

Pope, C.A., R.T. Burnett, M.J. Thun, E.E. Calle, D. Krewski, K. Ito, and G.D. Thurston, Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, *JAMA – Journal of the American Medical Association*, 287, 1132-1141, 2002.

Pounds, J.A., M.R. Bustamante, L.A. Coloma, J.A. Consuegra, M.P.L. Fogden, P.N. Foster, E. La Marca, K.L. Masters, A. Merino-Viteri, R. Puschendorf, S.R. Ron, G.A. Sanchez-Azofeifa, C.J. Still, and B.E. Young, Widespread amphibian extinctions from epidemic disease driven by global warming, *Nature*, 439, 161-167, 2006.

Root, T.L., J.T. Price, K.R. Hall, S.H. Schneider, C. Rosenzweig, and J.A. Pounds, Fingerprints of global warming on wild animals and plants, *Nature*, 421, 57-60, 2003.

Ruddiman, W.F., The anthropogenic greenhouse era began thousands of years ago, *Climatic Change*, 61, 261-293, 2003.

Ruddiman, W.F., and A.G. Carmichael, Pre-industrial depopulation, atmospheric carbon dioxide, and global climate, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 158-194, The Pontifical Academy of Sciences, Vatican City, 2006.

Sachs, J., and P. Malaney, The economic and social burden of malaria, *Nature*, 415, 680-685, 2002.

Sachs, W., Climate change and human rights, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 349-368, The Pontifical Academy of Sciences, Vatican City, 2006.

Schellnhuber, H.-J., P.J. Crutzen, W.C. Clark, and M. Claussen (ed.), (2004), *Earth System Analysis for Sustainability*, Cambridge, Mass.: The MIT Press.

Schwartz, J., Air pollution and children's health, *Pediatrics*, 113, 1037-1043, 2004.

Shah, M.M., G. Fischer, and H. v. Velthuizen, Food, water, health, and infectious diseases: Focus on global change, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 230-251, The Pontifical Academy of Sciences, Vatican City, 2006.

Smith, K.R., and S. Mehta, The burden of disease from indoor air pollution in developing countries: comparison of estimates, *International Journal of Hygiene and Environmental Health*, 206, 279-289, 2003.

Steffen, W., M.O. Andreae, P.J. Crutzen, P. Cox, U. Cubasch, H. Held, N. Nakicenovic, R. Scholes, L. Talaue-McManus, and B.L. Turner, II, Abrupt changes: The Achilles heels of the Earth System, *Environment*, 46, 8-20, 2004a.

Steffen, W., and E. Lambin, Earth system functioning in the Anthropocene: Human impacts on the global environment, in *Interactions Between Global Change and Human Health* (*Scripta Varia* 106), edited by M.O. Andreae, D.E. Bloom, U. Confalonieri, & A.J. McMichael, pp. 112-144, The Pontifical Academy of Sciences, Vatican City, 2006.

Steffen, W., A. Sanderson, P.D. Tyson, J. Jäger, P.A. Matson, B. Moore, III, F. Oldfield, K. Richardson, H.J. Schellnhuber, B.L. Turner, II, and R.J. Wasson (2004b), *Global Change and the Earth System: A Planet Under Pressure*, Berlin: Springer-Verlag, 336 p.

Stott, P.A., D.A. Stone, and M.R. Allen, Human contribution to the European heatwave of 2003, *Nature*, 432, 610-614, 2004.

United Nations (2005), The Millennium Development Goals Report 2005, *United Nations*, 48 p.

van Vuuren, D.P., J. Cofala, H.E. Eerens, R. Oostenrijk, C. Heyes, Z. Klimont, M.G.J. den Elzen, and M. Amann, Exploring the ancillary benefits of the Kyoto Protocol for air pollution in Europe, *Energy Policy*, 34, 444-460, 2006.

Weiss, R.A., and A.J. McMichael, Social and environmental risk factors in the emergence of infectious diseases, *Nature Medicine*, 10, S70-S76, 2004.

Wolfe, N.D., P. Daszak, A.M. Kilpatrick, and D.S. Burke, Bushmeat hunting deforestation, and prediction of zoonoses emergence, *Emerging Infectious Diseases*, 11, 1822-1827, 2005.

World Health Organization (2002), *The World Health Report 2002*, WHO, 248 p.

TABLES

Figure 1. Photographs of Atmospheric Brown Clouds (Brown Haze in the horizon): From Los Angeles (Top left), Sierras in N. California (Top right) and Nepal (the two photos were taken at the same time from an aircraft). All photos were taken by the author except for picture of Mt. Everest, which was taken by Mr. Hung Nguyen.

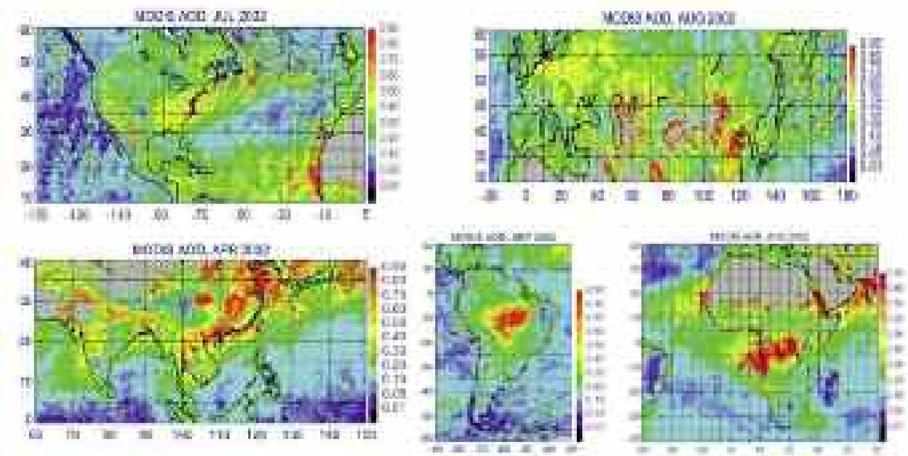


Figure 2. Aerosol Optical Depth (an index of Number of particles in the atmosphere). Obtained from MODIS instrument on TERRA Satellite. Source: Ramanathan and Ramana, 2002.

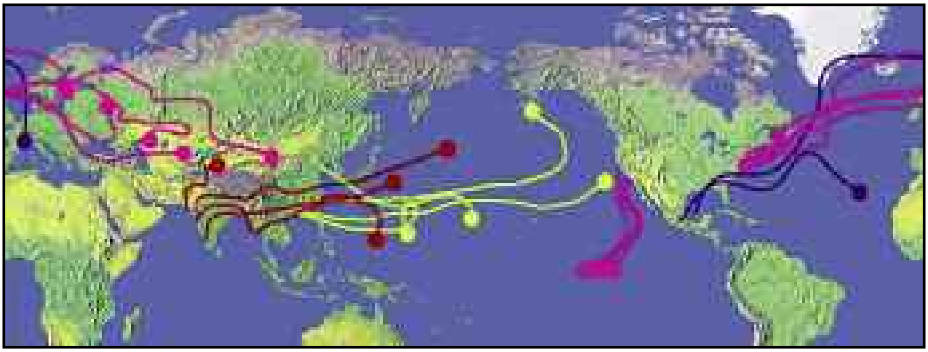


Figure 3. The path taken by an air parcel in a week at 3 km from the surface.

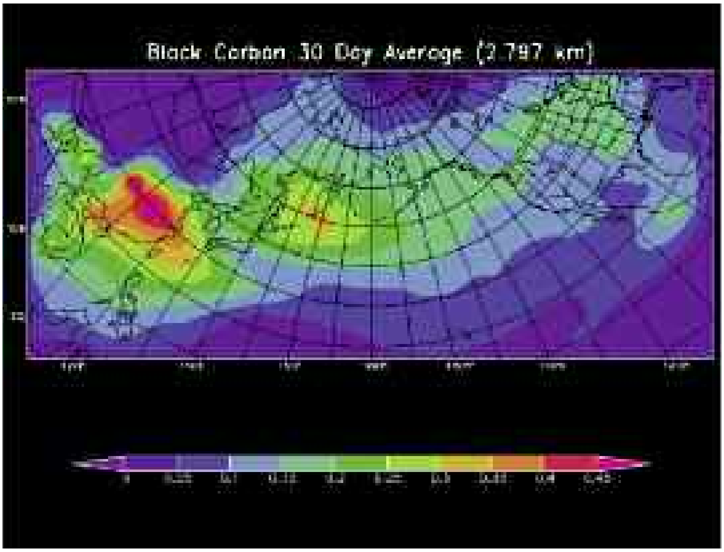


Figure 4. Simulated black carbon concentration by an aerosol-transport. Model developed by G. Carmichael (Private communication).

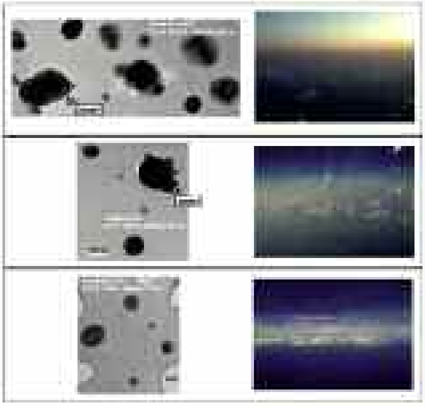


Figure 6. Fingerprint for manmade particles in the brown clouds over the Arabian Sea as captured during the INDOEX experiment. The right hand side shows the photograph of the sky, and the corresponding panel on the left right hand side shows X-Ray microscope images of the particles captured on a filter taken from C-130 aircraft in the region shown on the photo. Source: Ramanathan *et al.*, 2001a.

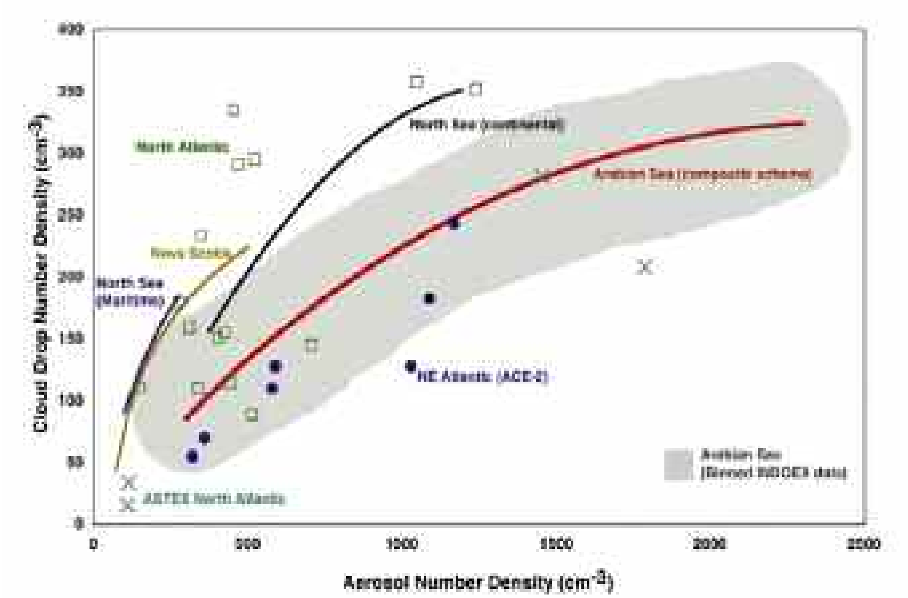


Figure 8. Aircraft data illustrating the dependence of cloud drops on aerosol number density (both natural and anthropogenic). Adapted from Ramanathan *et al.*,18.

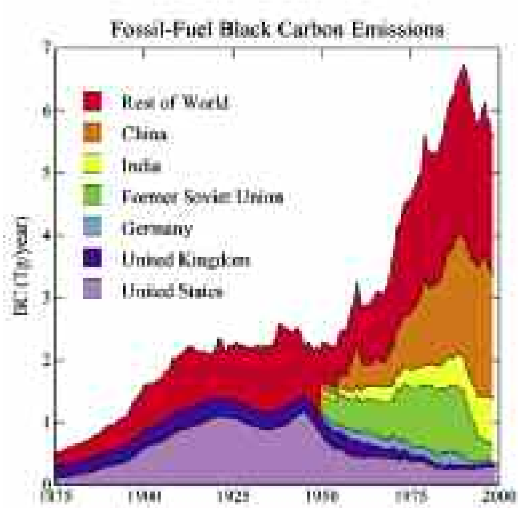


Figure 9. History of black carbon emissions due to fossil fuel combustion.

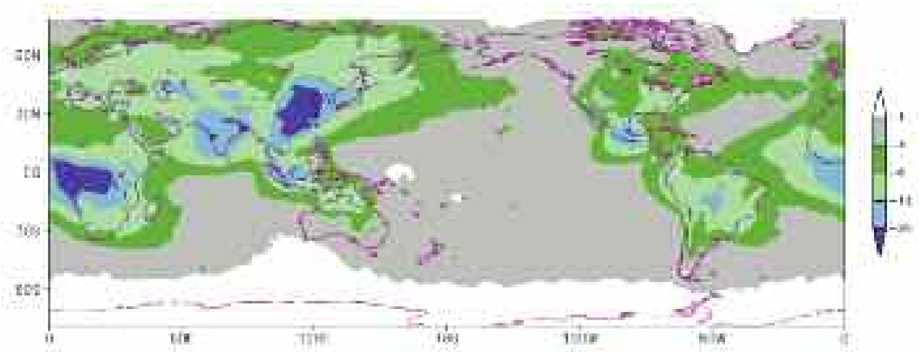


Figure 10. Global reduction in solar radiation at the surface due to brown clouds for years 2000-2004. Unit: Wm^{-2}. Source: Chung, Ramanathan, Kim and Podgorny, 2005.

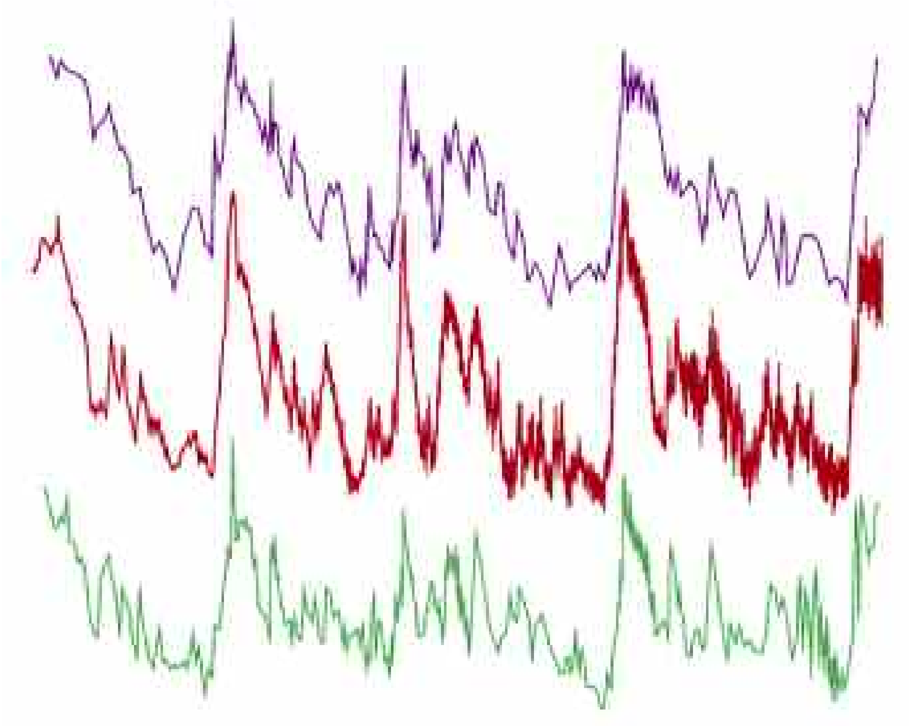


Fig. 1. The 420,000-year Vostok (Antarctica) ice core record, showing the regular pattern of atmospheric CO_2 and CH_4 concentration and inferred temperature through four glacial-interglacial cycles (adapted from Petit *et al.*, 1999).

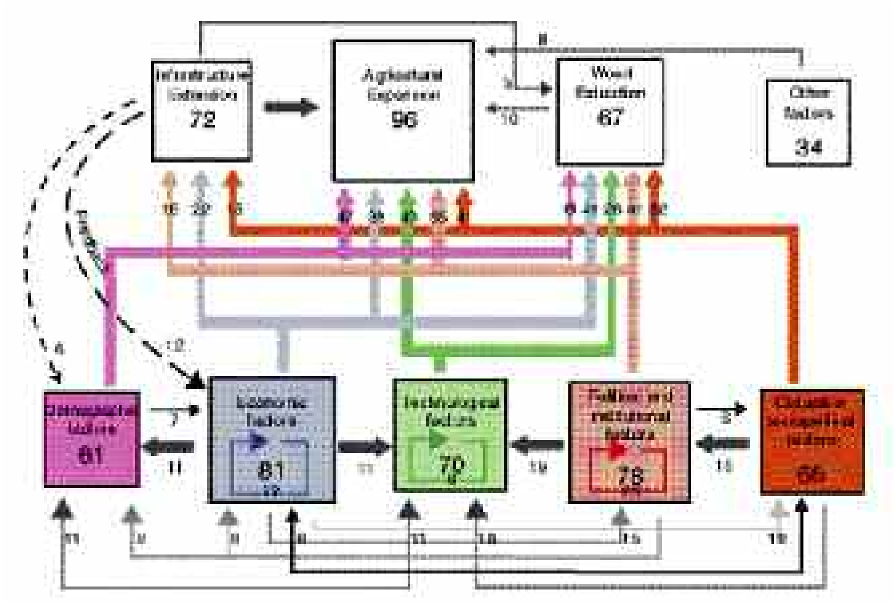


Fig. 5. The causative patterns of tropical deforestation from 1850 to 1997 based on a synthesis of 152 case studies examining both proximate and underlying causes (Geist and Lambin, 2002).

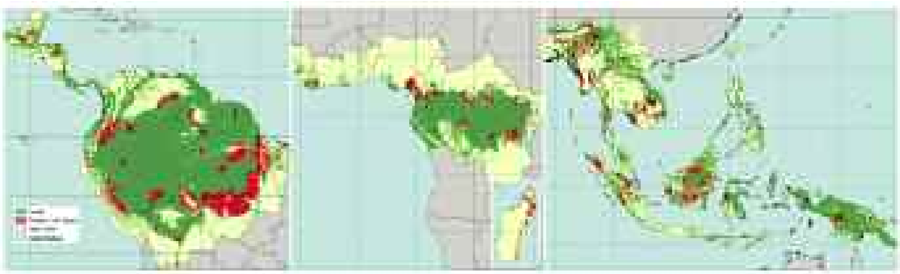


Fig. 6. Locations of rapid tropical deforestation over approximately the last 20 years (Lepers *et al.*, 2003). The map indicates the number of times each 0.1° grid was identified as undergoing rapid deforestation. Data derived from Achard *et al.* (2002), DeFries *et al.* (2002) and Landsat Pathfinder for the Amazon Basin. Pink colours indicate that only one of the data sets identified the grid as undergoing rapid deforestation, red indicates two data sets and black indicates three.

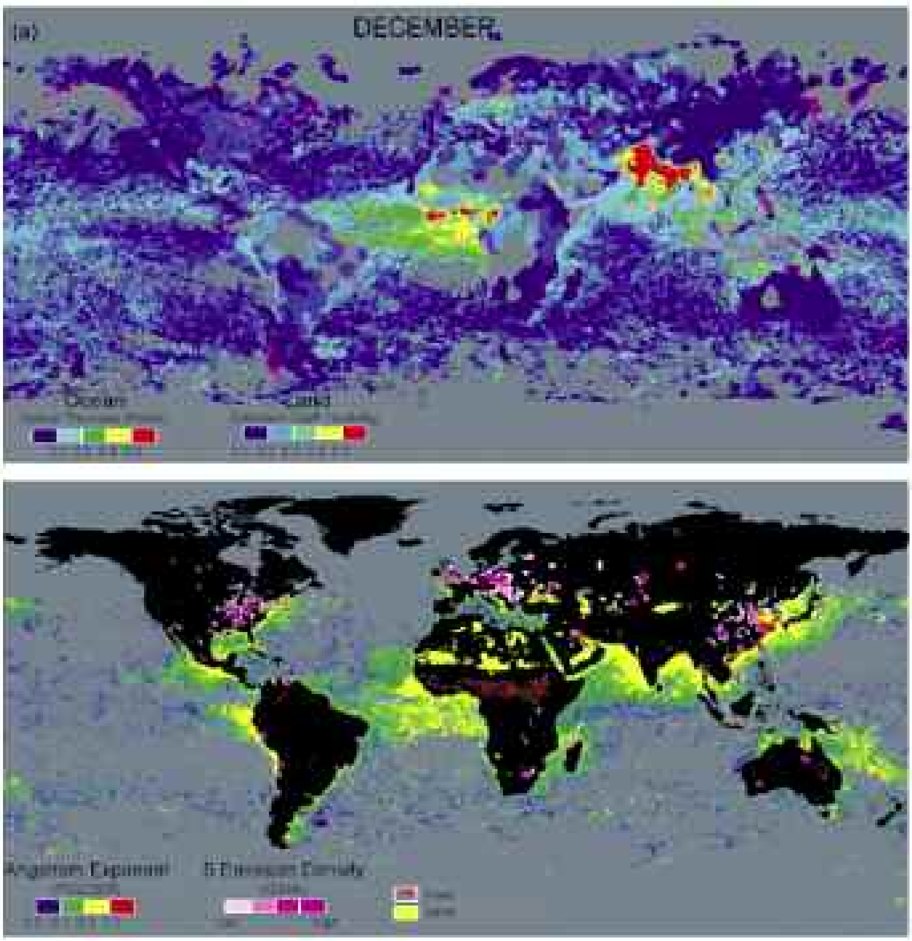


Fig. 7. Global horizontal patterns for (a) December 1996 and (b) June 1997 of aerosol optical thickness and Ångström exponents over the oceans from the POLDER instrument aboard the ADEOS satellite and extinction coefficients over land areas from visibility data (reproduced with permission of LOA, LSCE and CNES, France; and NASDA, Japan).

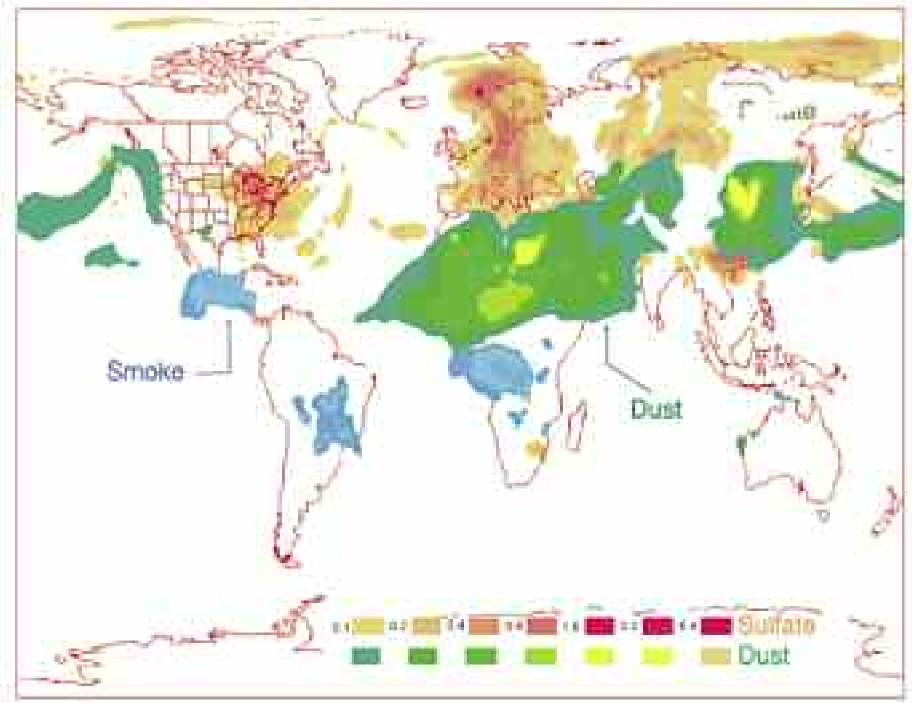


Fig. 7. (c) Global fields of mineral dust, sulphate and smoke, calculated with a chemical transport model using actual meteorology (http://www.nrlmty.navy.mil/aerosol/; Heintzenberg, 2003).

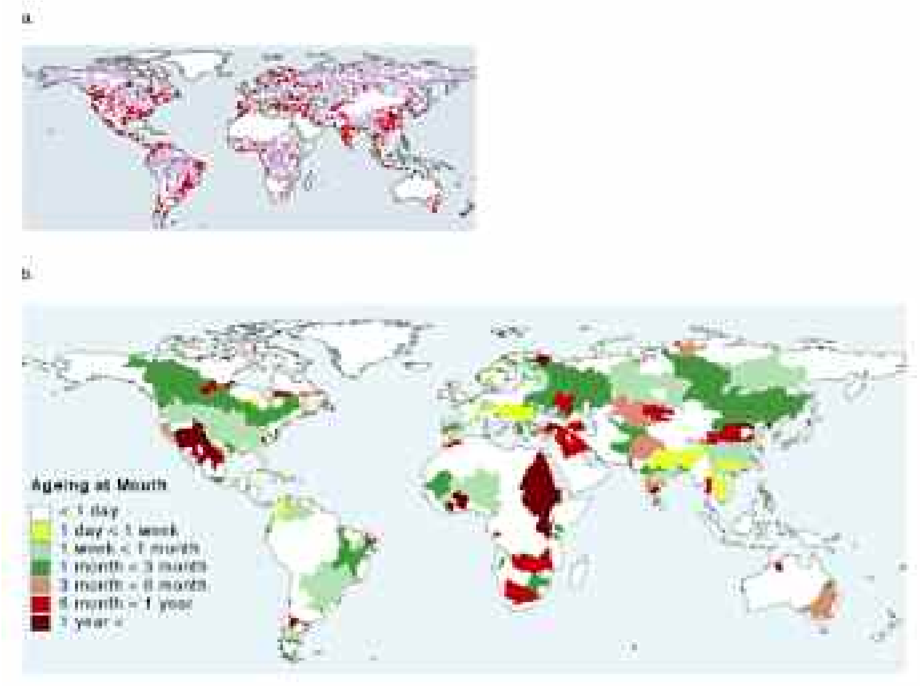


Fig. 9. (a) Global distribution of 622 large reservoirs, classified as those with a maximum capacity greater than 0.5 cubic kilometres. (b) Aging of river water, as computed at the mouth of each of the 236 regulated drainage basins. Aging varies due to river regulations on the reservoirs (Vörösmarty *et al.*, 1997).

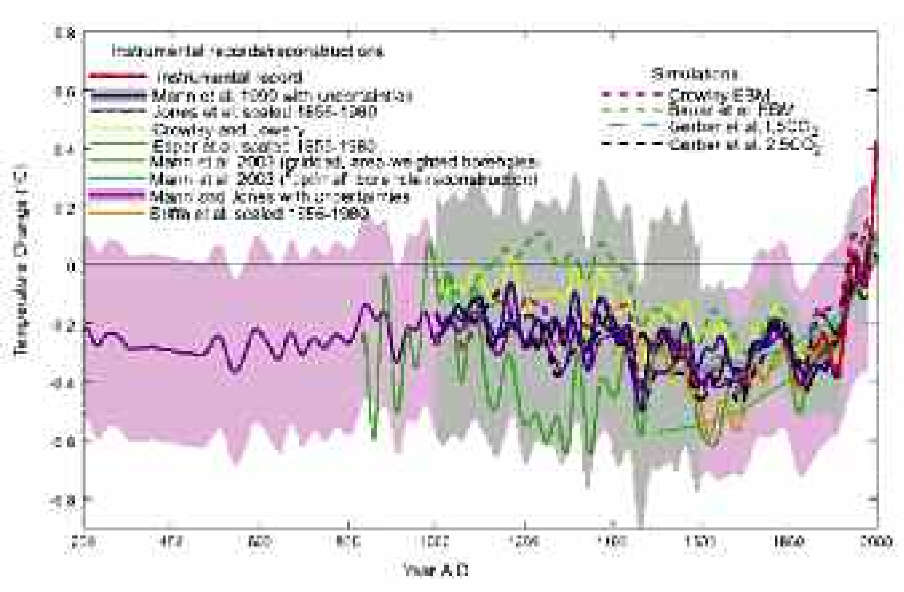


Fig. 10. Various reconstructions of mean annual surface temperature over the northern hemisphere for the last one or two thousand years (Mann and Jones, 2003).

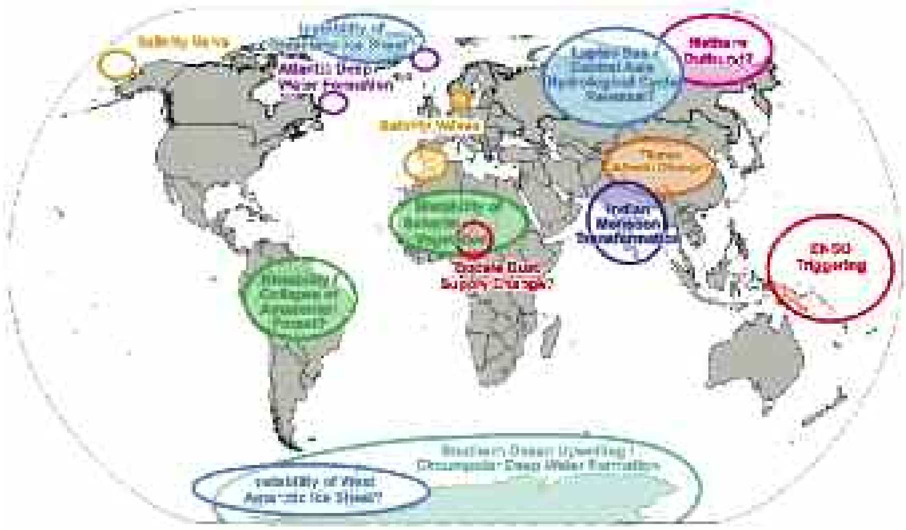


Fig. 11. Critical region analysis of hot spots or switch and choke points; an attempt to identify parts of the Earth where changes at the regional scale can cause significant changes in the functioning of the Earth System as a whole (Schellnhuber, 2002).

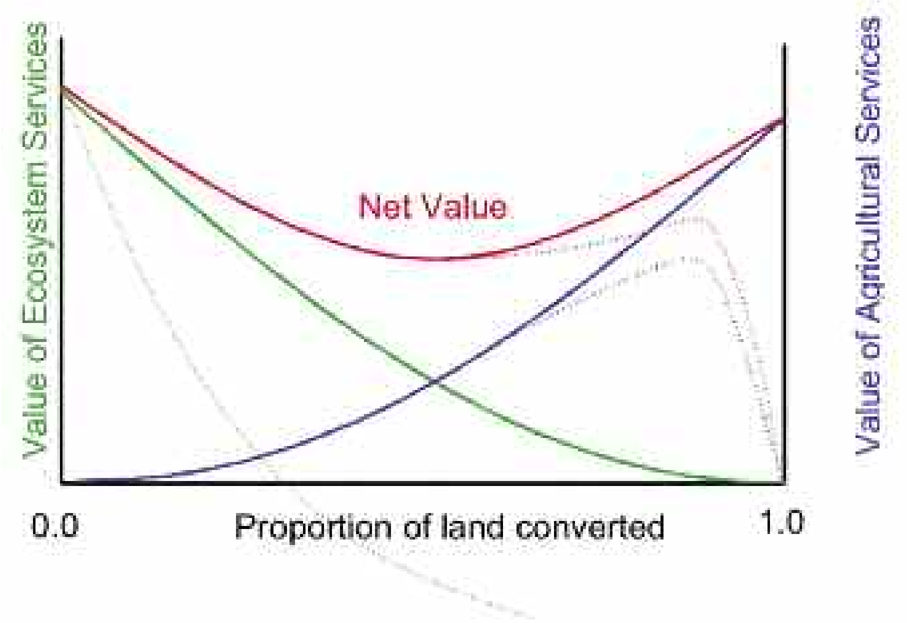


Figure 1. Change in land value as natural habitat is converted into agricultural or urban land. The net value of the land is given by the red line, which is the sum of the natural services (green) and the agricultural services (blue line). The quality of goods and services in the modified habitat may become limiting if their productivity is dependent upon services provided by the remaining natural land; this is indicated by the dotted blue and red lines. Also illustrated are the costs and future benefits of restoring the land to its original state (dashed green line).

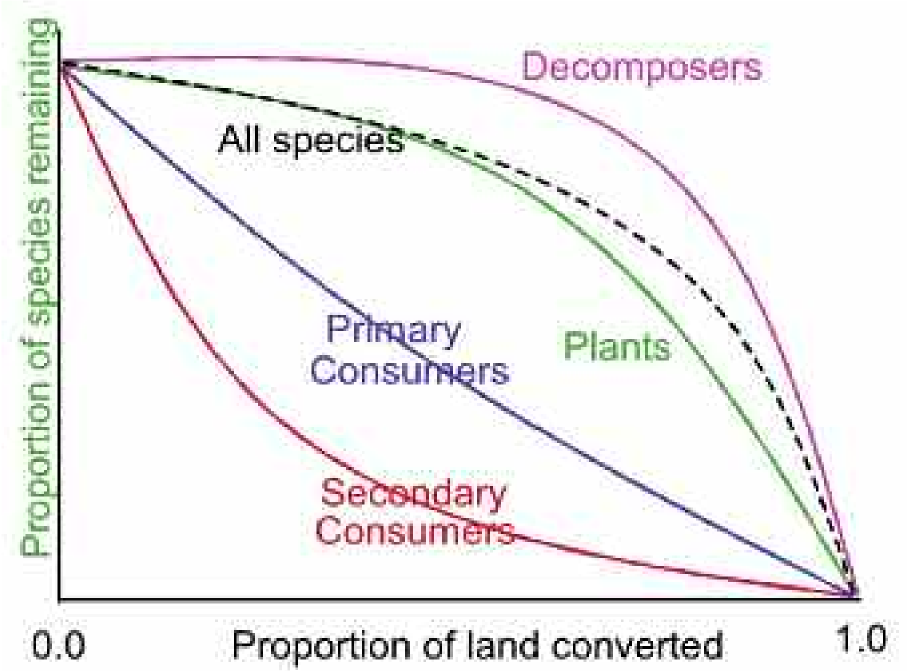


Figure 2. The loss of biological diversity and ecosystem services as natural habitat is eroded. The dashed line gives the classic species area decline in abundance such that a 90% loss of area dooms 50% of the original species to extinction. The colored lines indicate that species at different levels in the food chain will be lost at different rates. The plants at the base of the food chain will be lost at rates slightly slower than the average, the decomposers, such as worms, soil mites and bacteria will be lost at the slowest rates. In contrast, the species that feed directly on plants will be lost at a faster rate, while the charismatic tigers and eagles that feed at the top of the food chain are lost at the fastest rate.

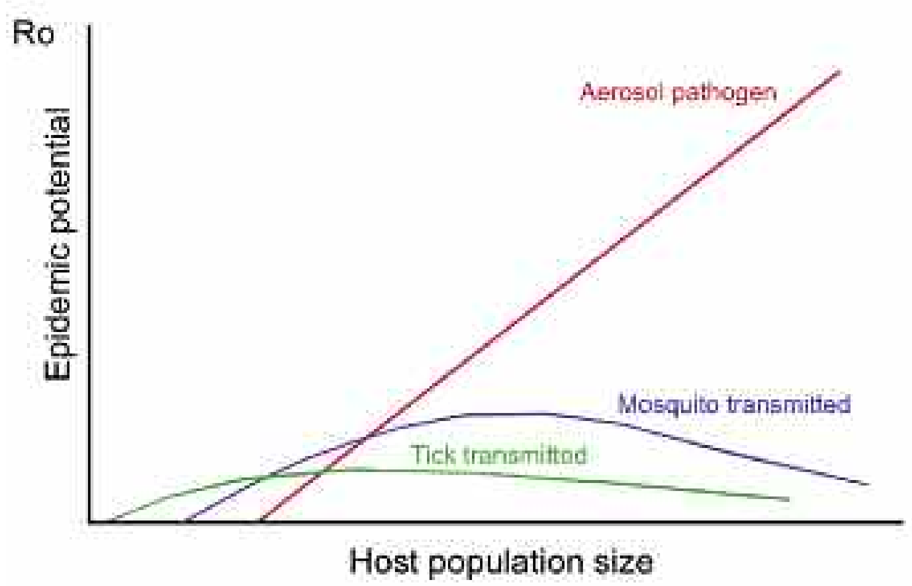


Figure 3. Relationship between host abundance and the basic reproductive number, R_0, for pathogens with direct, or vector transmission.

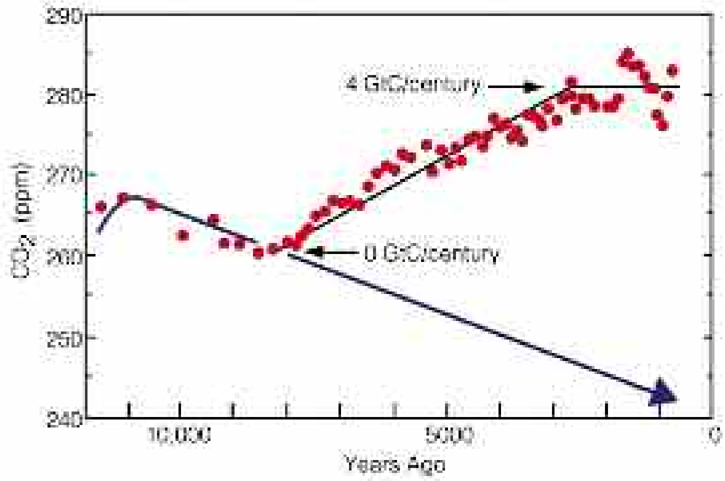


Figure 3. Observed trend in atmospheric CO_2 during the last 10,500 years from Indermuhle *et al*. (1999) compared to the trend predicted by previous interglaciations by Ruddiman (2003).

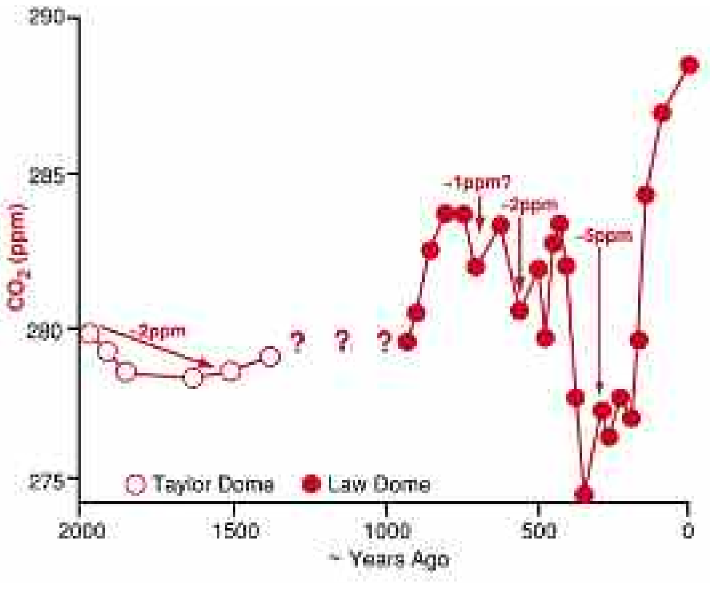


Figure 4. CO_2 signal used as 'target' for modeling exercise in this paper. Filled circles are values from Law Dome (Etheridge *et al*., 1996). Hollow circles are values from Taylor Dome (Indermuhle *et al*., 1999). Arrows mark intervals of CO_2 decreases during times of large-scale regional depopulation.

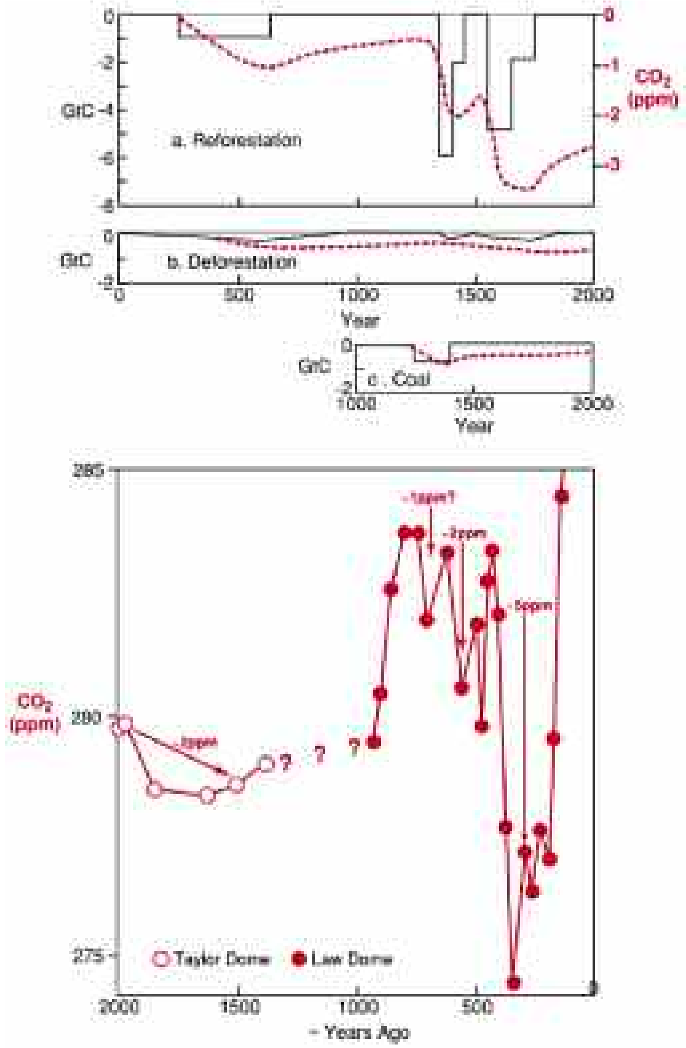


Figure 7. Model simulation of effects of major depopulation intervals on atmospheric carbon. CO_2 target signal shown in (A). Dashed lines are model simulation of effects of depopulation on CO_2, based on Figure 6. Reforestation (B) explains roughly half of the CO_2 target signal; reduced deforestation (C) and coal burning (D) contribute smaller amounts.

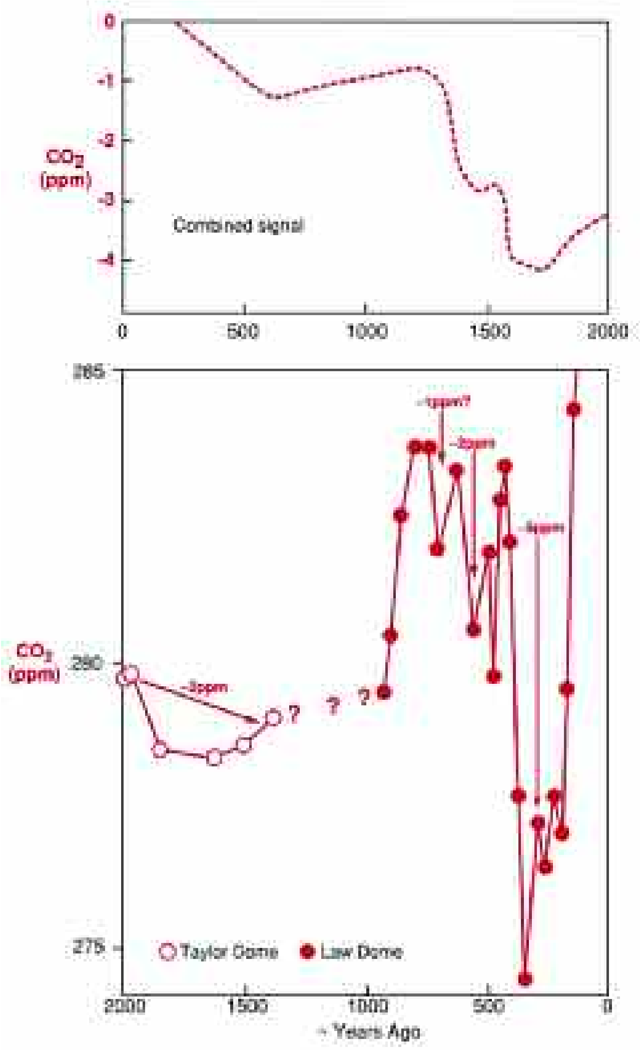


Figure 8. Comparison of combined effects from reforestation, reduced deforestation, and reduced coal burning tied to major depopulation decreases with CO_2 target signal (B).

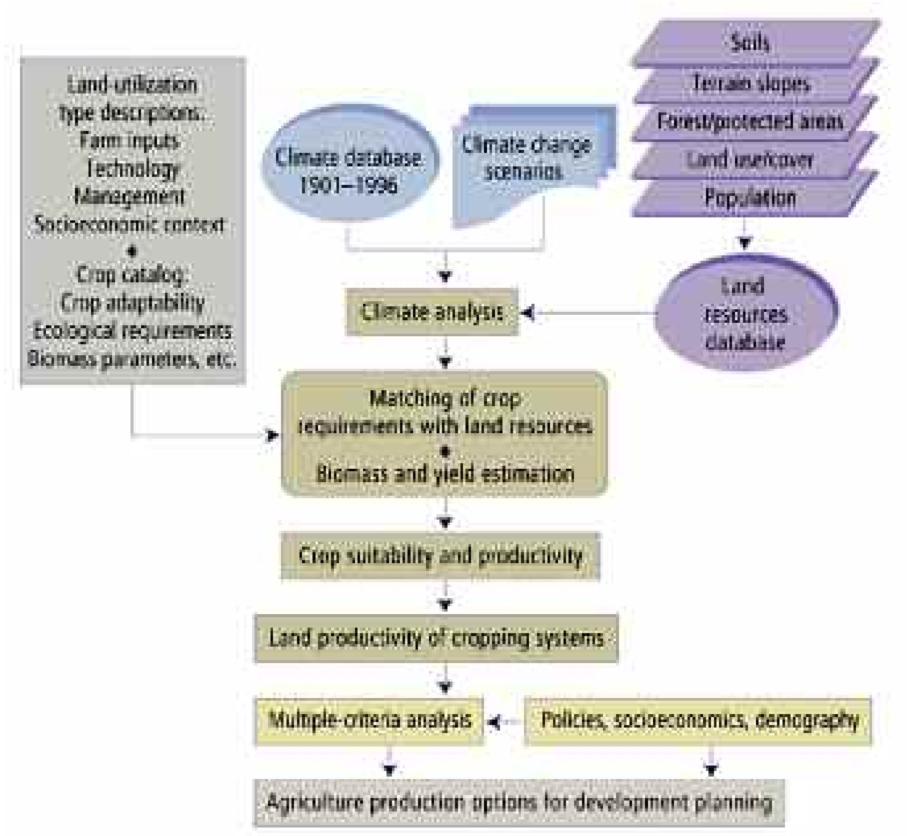


Figure 1. AEZ methodology (Fischer *et al.*, 2002b).

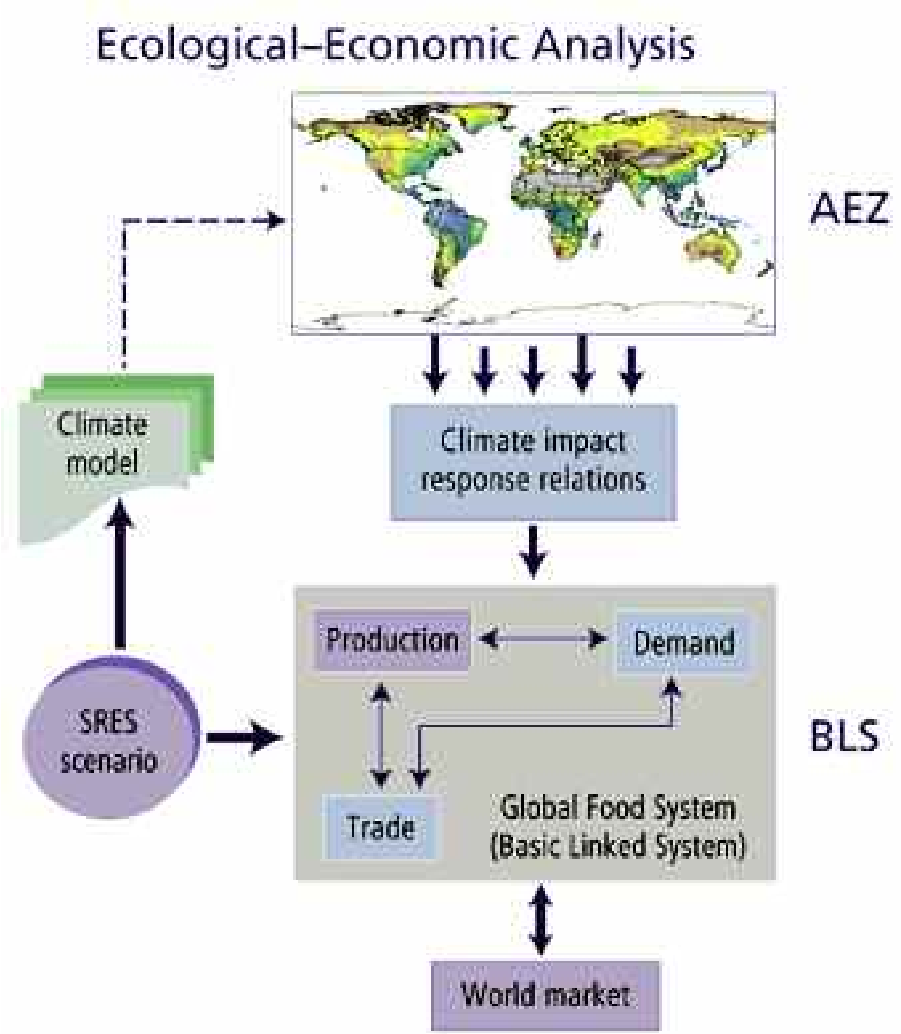


Figure 3. Integrated Ecological-Economic Analysis (Fischer *et al*., 2002b).

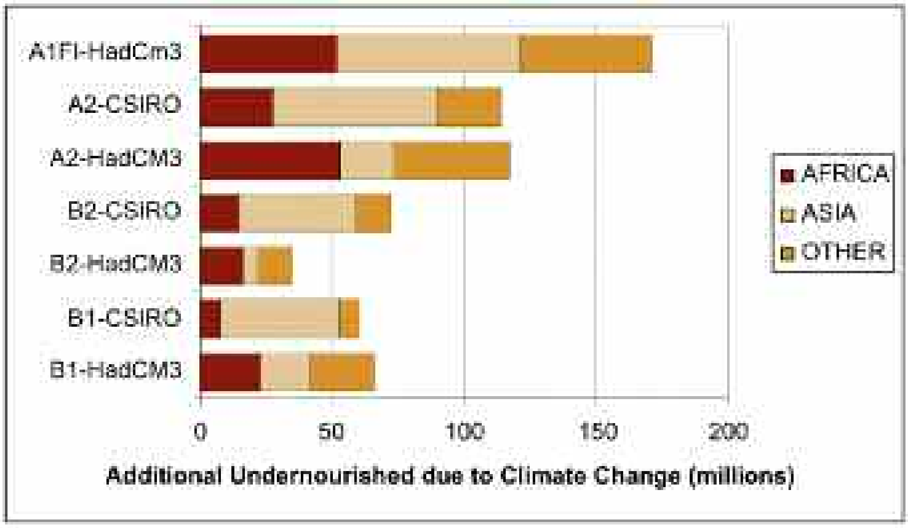


Figure 5. Additional undernourished population due to climate change (Fischer *et al.*, 2002b).

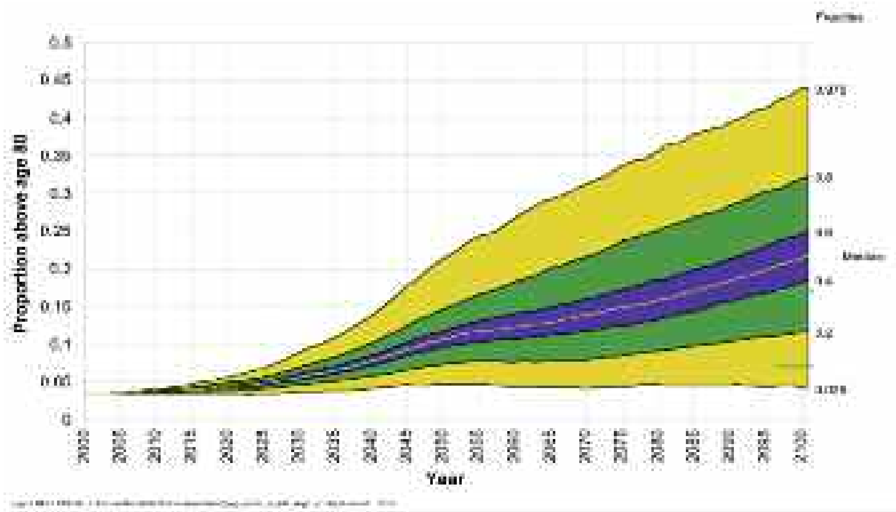


Figure 3. Proportion of population above age 80 in Western Europe (UN 'low' scenario for 2100 = 0.17, UN 'high' scenario = 0.07).

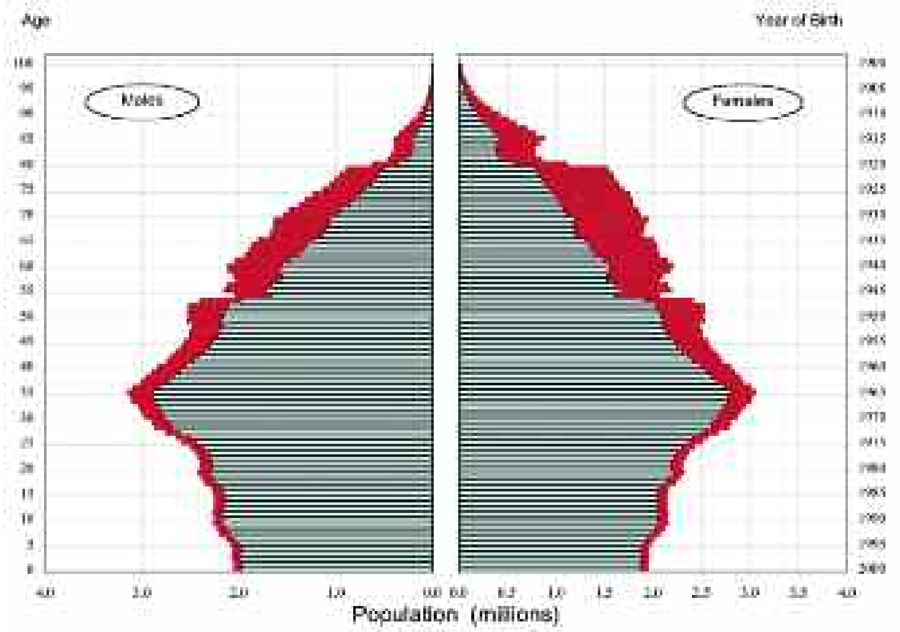


Figure 6. European Union, 2000, total and disabled population.

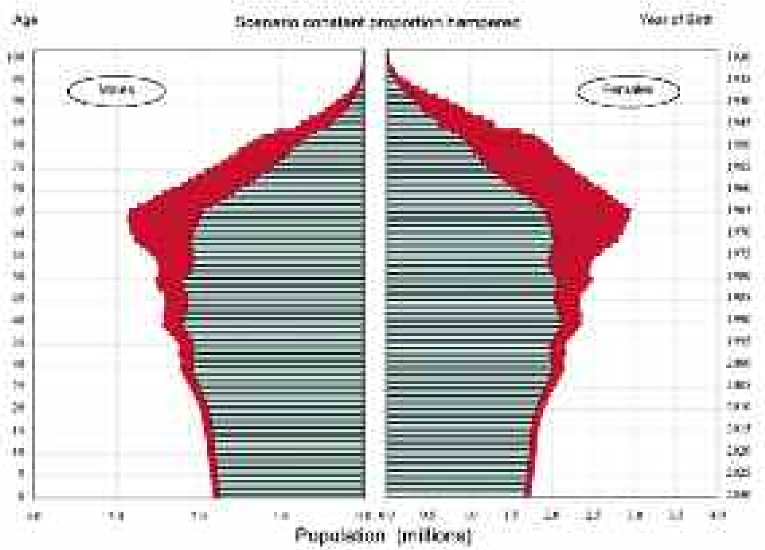


Figure 7a. European Union 2030, total and disabled population. (If desired, Figures 7a and 7b can be combined into one figure with three colors in which the difference between the two scenarios is given in yellow).

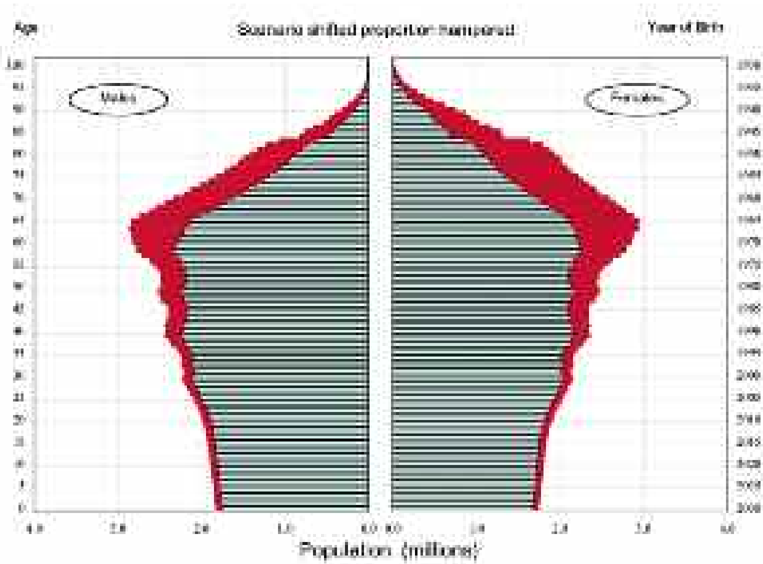


Figure 7b. European Union, 2030, total and disabled population.

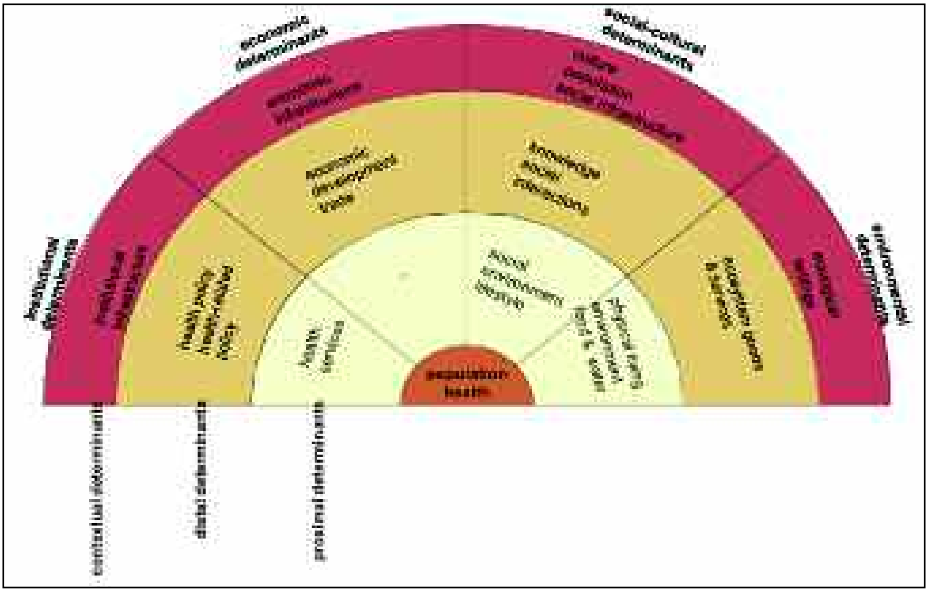


Figure 1. Multi-nature and multi-level framework for population health [3, 5].

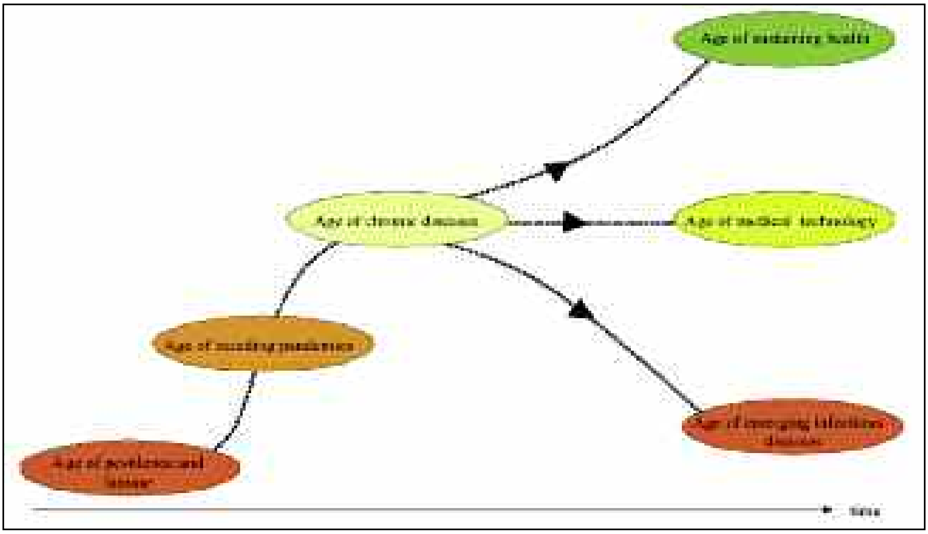


Figure 2. Future stages in the health transition. Currently, the developed world is in the 'Age of chronic diseases', while many developing countries are still in the second stage of the health transition. It is, however, possible for current developing countries to skip the 'Age of chronic diseases' and shift directly to the 'Age of sustained health', 'Age of medical technology' or 'Age of emerging infectious diseases'.

Figure 1. Source: World Bank, *World Development Indicators 2004*. Data are for 2002 or latest available year. Note: The circled outliers, from lowest to highest income, are Swaziland, Namibia, Gabon, Botswana, South Africa, Equatorial Guinea, and Luxembourg.

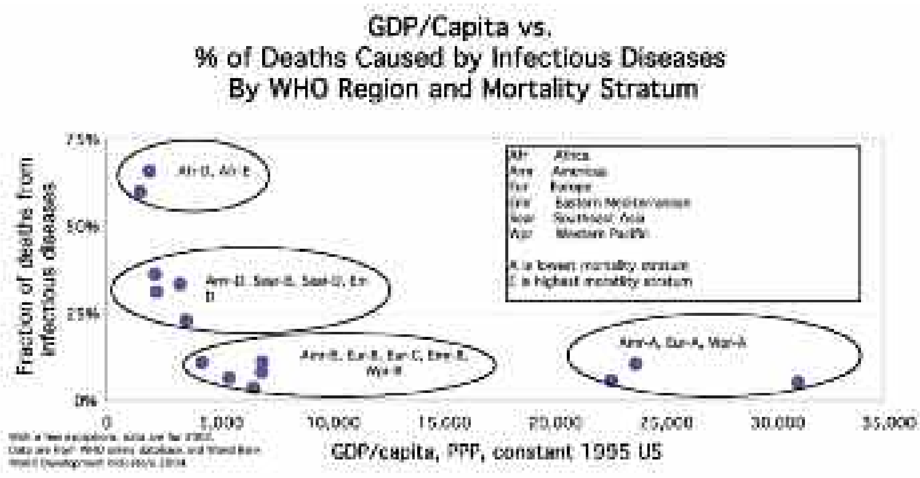


Figure 2.

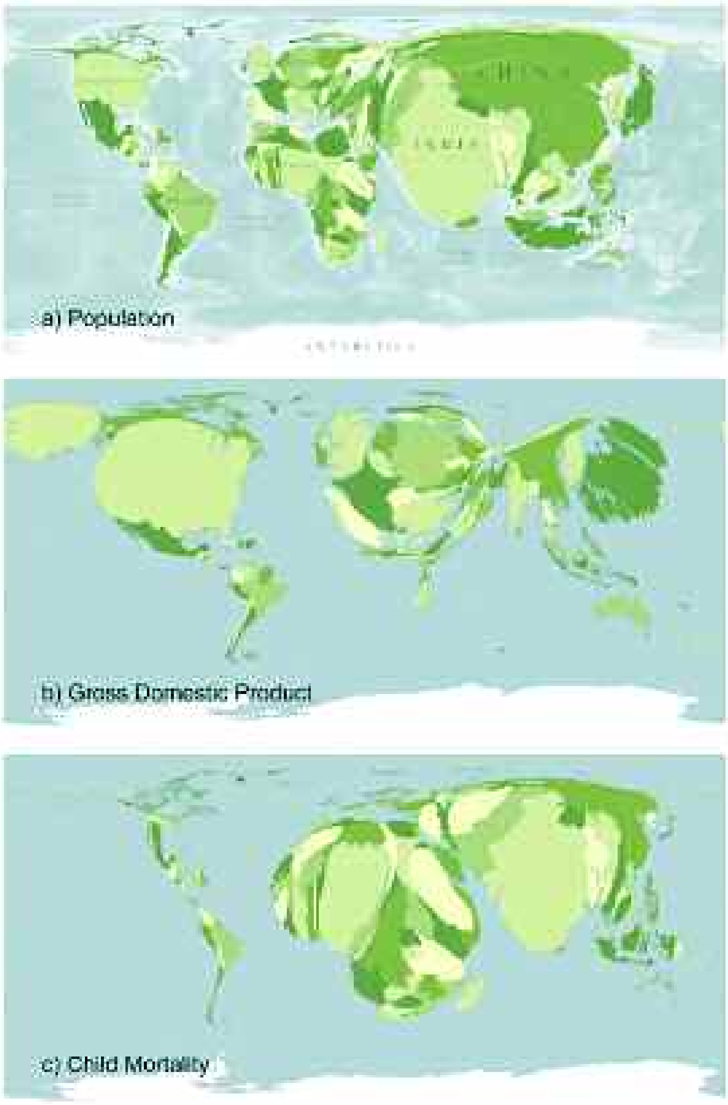


Figure 1. Cartograms of the world in which the sizes of countries are proportional to a) population, b) Gross Domestic Product (GDP), which is a measure of how much wealth a country's economy generates, and hence, to an extent, of the wealth of the country's inhabitants, and c) child mortality, as a measure of a nations public health status. Notice how America and Europe, along with Japan, dominate the GDP map and Africa dwindles almost to invisibility, while the child mortality map shows essentially the inverse distribution (Cartograms by Mark Newman; http://www.personal.umich.edu/~mejn/cartograms/).

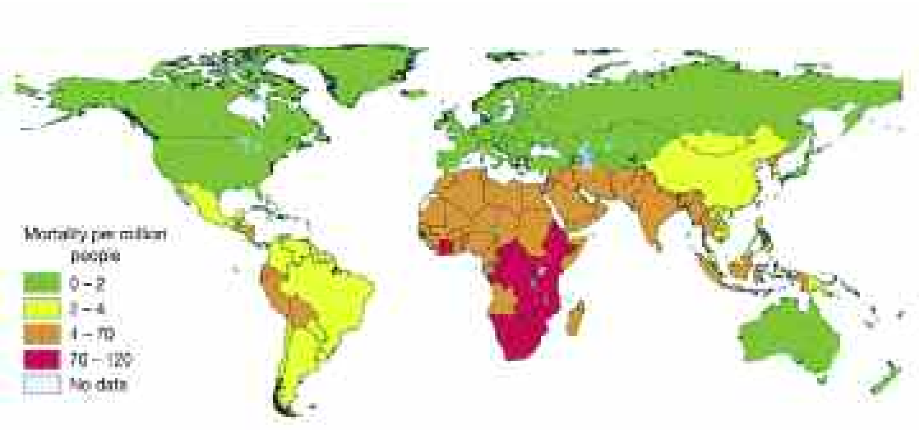


Figure 4. WHO estimated mortality (per million people) attributable to present climate change (year 2000). Existing quantitative studies of climate-health relationships were used to estimate relative changes in a range of climate-sensitive health outcomes including: cardiovascular diseases, diarrhea, malaria, inland and coastal flooding, and malnutrition. This is only a partial list of potential health outcomes, and there are significant uncertainties in all of the underlying models. These estimates should therefore be considered as a conservative, approximate, estimate of the health burden of climate change. Even so, the total mortality due to anthropogenic climate change by 2000 is estimated to be at least 150,000 people per year. (From Patz *et al.*, 2005, based on data reported in World Health Organization, 2002).