Paths of Discovery
Pontifical Academy of Sciences, Acta 18, Vatican City 2006
www.pas.va/content/dam/accademia/pdf/acta18/acta18-colombo.pdf

PATHS OF DISCOVERY: PERSONAL EXPERIENCES IN A SOCIAL SCIENCE

BERNARDO M. COLOMBO

The foundation of Demography may be traced to a booklet of 'Natural and Political Observations' made on the Bills of Mortality, then in use, that John Graunt presented to the Royal Society in 1662.[1] The death of a person is a common topic in a social conversation. Graunt, observing that, beyond that, little use was made of the information recorded in the Bills, considered the possibility of extending it in other directions and for other purposes. The 'death' event became for him simply an abstract object with certain characteristics: location, time, cause of the event, sex and – roughly – age of the deceased person... He went on, making classifications of those objects according to the specific expression of those characteristics in each observed unit collected over a number of years in the city of London and in the country town of Romsey, New Hampshire. In each science, the purposeful classification of observations is a fundamental step. As to the collection of data, in the area of population it goes back to ancient times. As we know, the birth of Jesus is linked to an operation of that kind. The classification of individuals is also old. A Psalm reminds us that 'The days of our life are seventy years – or eighty if we have the strength' (NRSV, *Ps* 90:10). A text that implies classification by age at death.

Graunt took a further step, looking for configurations and regularity in time of observed phenomena, but also putting in relation different classifications for the purpose of enriching knowledge. May I quote one example.[2] He saw that, year after year, the numbers of abortions and still-

[1] Graunt J., *Natural and Political Observations Mentioned in a following Index, and made upon the Bills of Mortality,*1662, London.

[2] Glass D., 'John Graunt and his Natural and Political Observations', *A Discussion on Demography*, Proc. of the Royal Society, B., vol. 159, 1963, 2-32.

births were roughly constant. The number of christenings – taken as the number of births – after a normal period, started and continued declining over several years. Taking the figures of 1631 – a normal year without problems – he calculated the ratio of abortions to christenings. Assuming that this ratio remained constant in time, he made an estimate of the 'true' number of christenings in 1659, a recent year in which he thought that 'the neglect in the *Accompt* of christenings' continued. To check this estimate, he followed a similar procedure using the number of women who died in childbirth. The techniques of analysis of the data he used were plain exercises of arithmetic, but as a path to discovery they were 'outstanding innovations'.

To advance one more step, I take a modern example. It concerns the measurement of the level of fecundability, taken as the probability that a married woman has a conception during a month, making abstraction from any practice meant to limit procreation. This concept, suggested in 1923 by Corrado Gini, opened a new chapter in Demography.[3]

According to Gini:

p = is the probability that a woman conceives in a month, probability supposed constant in time and the same for all women

n_x = is the number of women who did not yet have the first birth up to the x^{th} month since marriage.

Then $p \cdot n_x$ = number of first born by these women during the month x. The number of first births during the month $x + 1$ – disregarding mortality – will be $(1 - p)\, p \cdot n_x$ and during the month $x + 2$ will be $(1 - p)^2\, p \cdot n_x$. That is, the number of first births in successive months forms a geometrical progression of ratio $(1 - p)$. Month by month you easily get a rough estimate of p. A very simple operation, concerning which I highlight two aspects.

First, it is possible because each single unit is cross-classified by parity and distance from marriage. Cross-classification of covariates is an essential tool for understanding phenomena influenced by the concomitant impact of several factors. In the field of population studies, it is a demanding task. Take, for instance, a Yearbook of vital statistics based on documentation taken usually from administrative records. There are many variates of interest and a huge number of possible cross-classifica-

[3] Gini C., "Prime ricerche sulla 'fecondabilità' della donna", *Atti del R.*, Istituto Veneto di Scienze, Lettere ed Arti, Anno Accademico 1923-24, vol. 58, Parte 2, Venezia, Ferrari, 1924, 315-344.

tions with a variety of details. For the book, you must make a choice among them. This requires ideas about what might be relevant for scientific investigation. That demands a great amount of culture and imagination about the course of science. In the Yearbook you are providing building blocks for future research. According to the choices you make, you open or close doors. From personal experience, I know how challenging these choices are.

Secondly, besides the availability of adequate data, you need intuition on how to use them to shed light on unexplained or unknown realities. But how do such ideas come?

May I quote a personal experience. As a young research assistant in the University of Venice, I was confronted with the problem of understanding the reasons for the recovery of the birth rate in several countries during the Second World War: countries directly involved in the operations, or neutral but near the regions where battles were going on. An English colleague had seen in that a making up of births delayed during the great economic crisis, without a real change in the final family size.[4] In order to test this hypothesis, in each instance a series of specific rates standardized with respect to both parity and distance from marriage had to be computed. Detailed tables – with appropriate cross-classifications – were usually published each year, providing the figures for the numerators of such rates. The data for the denominators were missing everywhere, even in census years. I was at this point when one evening I went to the movie theatre to relax. Walking back home, with that problem in mind, I suddenly had an idea about how the data available for the numerators could, through appropriate elaborations, be used to create the figures for the denominators of those rates. It worked. I thus had the possibility of showing that fertility in those countries really increased, without any making up.[5]

I also tried to give an explanation of that phenomenon in a few pages of considerations which I supposed plausible. When the teacher who was my guidance finished reading the manuscript, he asked me: 'Can you prove that?'. I had to answer: 'No'. Those pages were omitted in the published paper. That lesson made me very cautious in evaluating conclusions

[4] Hajnal J., *The Analysis of Birth Statistics in the light of the recent International Recovery of the Birth-Rate,* Population Studies, 1, 2, 1947, 137-164.

[5] Colombo B., *La recente inversione nella tendenza della natalità,* CEDAM, Padova, 1951, VIII+183.

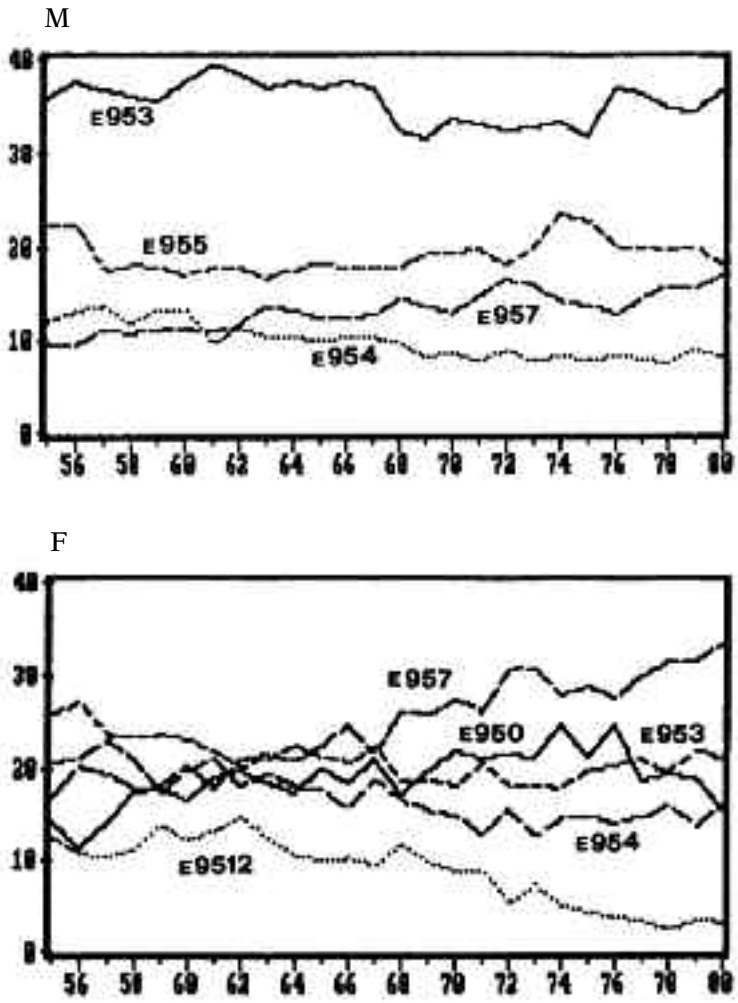

Fig. 1. Percentage of suicides of various categories on the total number of suicides for the given sex: Italy, 1955-1980.
Note: Suicides and self inflicted injury by
E950= poisoning by solid or liquid substances
E951= poisoning by any gas in domestic use
E952= poisoning by other gases
E953= hanging, strangulation and suffocation
E954= submersion (drowning)
E955= firearms and explosives
E957= jumping from high places
E9512= E951+E952

reached in similar areas. The risk of spurious correlations, ecological fallacies, explanations of the 'post hoc ergo propter hoc' type, and so on, now immediately raise my critical attention.

Demographic analysis can offer clear illustrations of ongoing phenomena. Take, for instance, the baby boom of the postwar period, a development common to Western Countries. Its peak was reached in 1960 in North America and in 1964 at the same time in several countries of the European region. But when you ask: 'Why is that?', demographic analysis alone is unable to give answers.

Or take the deaths due to suicides recorded in a recent quarter of a century in Italy – from 1955 through 1980.[6] Charting the figures of the two sexes according to the different forms of self-murder, one can see that for the males the lines corresponding to the different manners run parallel over the years, while for the females they change levels and even cross (see Fig. 1). Maybe the collaboration of different disciplines can shed some light on the relevant factors having an impact on these social trends, but within a limiting condition. In this field, experiments are denied. In the interpretation of social events you can hardly give unequivocal proof, as you are unable to falsify it.

Population dynamics is at the crossroads of physiology and behaviour and raises problems of interpretation on both sides. Take, for instance, the sex of newborns, with the slightly higher frequency of males, the most well-known observation made by John Graunt. I studied in depth the question of the systematic differences occurring in the comparison of chosen selected groups of births classified by sex.[7] I found confirmed, for example, in the Second World War, what had been observed in the first one: an increased masculinity ratio. I will not elaborate on the hypotheses made by various authors about factors thought to be behind the results. Surely, to confirm or disprove any of them, a third world war would be too costly an experiment. Similar intriguing evidence defies interpretation in several other instances. Many hypotheses have been advanced to provide explanations, hypotheses that often died with their authors, so that we can repeat what Sir Francis Bacon happened to state: 'What was a question once, is a question still'.

[6] Colombo B., Fortuna I., 'La regolarità degli eventi rari in Demografia', *Atti e Memorie dell'Accademia Patavina di Scienze, Lettere ed Arti*, Parte III, 1992, 65-83.

[7] Colombo B., 'On the sex ratio in man', *Population Studies: Animal Ecology and Demography*, Cold Spring Harbor Symposia on Quantitative Biology, 22, 1957, 193-202.

Along the same line, I take another example from personal experience. Making use of the very good and sizeable Italian vital statistics, I found that the probability of multiple births was increasing regularly and independently of both age of the mother and parity. At higher ages and parities it reached a level of more than four times the initial lower one.[8] A colleague showed that this trend depended on dyzigotic twins.[9] I contacted a geneticist, asking for an explanation: he had none.

Certainly demography must not be taken as a technique specialized in raising problems. It offers in fact, powerful tools to understand what is happening today and to foresee the scenarios of the future. Through Lotka's stable population theory and the model life tables of Coale and Demeny, for instance, it gives the possibility of deriving from small and uncertain pieces of information a complete and coherent picture of the general structure and dynamics of a population. This path is useful for understanding the situation of countries with poor or missing data.

The advent of the computer has drastically enriched the instruments of analysis of the discipline allowing to go to the roots of what we see happening at the population level. Births, deaths, marriages, migrations depend on individual decisions or conditions, in the context of a specific milieu. Now, using all the information provided by single units, we can try to derive from them, through appropriate models and procedures, how and why things follow the trend we see in the general population.

Intuition and imagination are here involved at various steps: in the clear expression of the question to be answered; in the identification of the variables – and their possible interactions – to be taken into consideration; in the definition of a model which – through a regression equation – links them to the empirical observations that need to be explained. A model may carry out several functions: it gives a synthetic description of ascertained evidence; it clarifies the links between variables and the causal relations with the observed data; it allows us to make predictions. Within the limit of stability of the characteristics of single units and of the involved links with specified covariates, it makes it possibile to foresee future developments. But it also allows to make, through extrapolation,

[8] Colombo B., *Appunti di Statistica per ostetrici e ginecologi,* Clinica Ostetrico-Ginecologica, Università di Padova, Anno Accademico 1960-61.

[9] Rigatti Luchini S., 'La frequenza dei parti monozigotici e dizigotici, in relazione all'età della madre al parto e all'ordine di nascita, in Italia', *Atti della XXIX Riunione scientifica Società Italiana di Statistica*, vol. 2, 2, Bologna, 1978, 73-83.

extensions to peculiar sets of population: for instance to those provided with extreme features of characteristics considered relevant. This last exercise needs to be done very carefully. What you are doing, in fact, are elaborations on virtual populations. If you want to be comfortable with the results, you need the support of some evidence.