The Event History Approach in Demography

Daniel Courgeau; Eva Lelievre

Population: An English Selection, Vol. 3 (1991), 63-79.

Stable URL:
http://links.jstor.org/sici?sici=1169-1018%281991%292%3A3%3C63%3ATEHAID%3E2.0.CO%3B2-J

Population: An English Selection is currently published by Institut National d'Études Démographiques.

THE EVENT HISTORY APPROACH IN DEMOGRAPHY

Period analysis, then cohort analysis, have both proved unsatisfactory for investigating certain aspects of population, where the conventional sources of demographic data fall short. These can be approached by means of event history analysis, which has been made possible by surveys collecting a wealth of biographical data. Daniel COURGEAU* *and* Eva LELIÈVRE* *have used bivariate and multivariate models to study the complex interaction between life events. Their analysis reveals various types of dependence (reciprocal, unilateral) or independence between events. Parametric or semiparametric models are then used to take population heterogeneity into account and to show the different strategies followed. The problems raised by unobserved heterogeneity are solved in part, but collaboration with other disciplines seems necessary.*

Sociologists, psychologists, demographers ... are all interested in life histories, which provide a wealth of information for the different social sciences. But each discipline approaches these histories in its own way, fashioning them to satisfy its own particular curiosity, and focusing on certain aspects while neglecting others. Yet the study of life histories has the same goal in all these different fields: a better understanding of the mechanisms of change in individual and community behaviours.

The outward structure of the life history is the same in all cases. It can be identified by the succession of transitions and states which mark the individual's life course, and which can be defined and situated in time and space. But an anthropologist, a sociologist and a demographer will not define and locate an event or state in the same way, and views and perceptions will also differ within each discipline, from one researcher to another. The number of possible states or types of transition will be reduced or multiplied accordingly.

An individual's spatial mobility, for instance, can be approached in a variety of ways: international migration, internal migration, temporary moves, commuting, etc. (Courgeau, 1988). The divisions of the space in which an individual moves may therefore by defined in more or less complex ways.

* INED.

Furthermore, these events belong to a variety of fields, in which our power of decision may vary. It may be null, for instance when the events are related to physical laws beyond our control: an apple falling, an earthquake. It may be very slight, when events are related to biological laws, such as puberty and the menopause. In contrast, the scope for decision is much broader with events which concern the social, economic and political spheres. In our family life, for example, marriage, the birth of a child, etc. are events over which we have a power of decision, even if this is limited in certain societies. In our working life, we can resign from a job we do not like, even if that means problems later in finding another one. Other events – more psychological, for instance – may be more complex, and more difficult to define and date, but this is possible: the beginning and the end of a friendship may be dated, for instance.

Out of this enormous network of events, researchers must focus on a few, which they consider particularly important and which can be adequately investigated by their own particular field of research. They thus consider as "negligible, and neglect, circumstances which they do not attempt to investigate or to verify" (Costa de Beauregard, 1988). They then attempt to analyse the events and states which they have selected and to find a logic to their pattern, as though the other events and states had no influence on them. The dangers of this procedure are evident. To challenge it, when this is justified, means shaking some of the foundations of social science.

For instance, the demographer will focus on the main events which mark the life course from the point of view of the family (marriage, birth of a child, divorce...), employment (entry into the labour force, change of occupation...), migration (international migration, local moves...) and health (illness, death...). He/she will ignore events which would be considered important in other fields: the psychologist will pay attention to how friendships build up or fade over the years, which will be judged of secondary interest in demography[1].

In spite of such differences, the events on which information is collected are often the same: birth, economic activity, spatial mobility, death, etc. are events which are studied in sociology, psychology, demography... The divergences are in the ways in which the data are collected (sample surveys, in-depth interviews, etc.) and analysed, and in the object of the study. Within the same discipline, the focus of research may also change over time.

In the present paper, we shall examine such changes in the field of demography, by following the progression from period analysis to cohort analysis. We shall discuss the hypotheses underlying each of these approaches and show how their distance from actual human behaviour could lead to mistaken or incomplete conclusions. We shall then show how event

---

[1] See, however, Courgeau (1972).

history analysis can solve these problems and we shall define the hypotheses implied by this new approach. Finally, we shall come back to the complementarity of the different social sciences.

I. – From period analysis to cohort analysis

For the most part, demographers study population in terms of quantities. They therefore need to compare the behaviour, with respect to various demographic phenomena, of populations living in different countries, or in the same country at different times.

The definition given to the above terms (population, behaviour, phenomenon, period) may transform the approach, as we shall see.

We can say that 'population' is the population observed in a given place (a country, for instance) at a given time. The 'period' considered will be this time, generally a year, during which we observe the events which are the 'phenomena' under study. 'Behaviour' will then be the timing and structure of each phenomenon, which we suppose acts independently of all others. This is the exact context of period analysis, which was introduced in the late seventeenth to eighteenth centuries (Halley's Table in 1693, Wargentin's Table in 1776) and was used almost exclusively up to the end of the Second World War.

Period analysis was first employed to study mortality during a short lapse of time, generally a year. It was gradually extended to the other demographic phenomena: nuptiality, fertility, etc.

We consider the application of this method to mortality. We suppose that we know the number of deaths occurring at each age in a given year, in a population whose exact age structure we also know at the beginning of that year. Supposing that all individuals of a given age stand the same chance of dying, then the probability of dying within that calendar year – the ratio of deaths for a given age to the number of persons of that age on January 1st – provides an estimate of this risk. As Landry says (1949), the demographer "works on homogeneous material, that is, considered homogeneous".

Here, demography shares the 'subjectivistic' or Bayesian approach to probabilities. In this approach, as stated by Costa de Beauregard (1988), "the probability expresses nothing more than an estimate of chance based on what we know (or think we know). According to this view, the estimation of probability is a *hypothesis* which would be tested *if possible* as a 'statistical frequency', by means of *repeats* of the experience. But there are never strict repeats of any event or situation, since there is only ever *one history*. Two experiences which are said to be identical, yet are distant from one another in time or space, are only considered identical because certain circumstances are taken to be negligible, and are neglected, with

no attempt to investigate or verify them." In the mortality example, each experience is the follow-up of an individual over a year. Two experiences are said to be identical when we suppose that none of the individual's characteristics other than age has much influence on the risk of dying.

Yet in period analysis, the influence of other factors can easily be tested. The population can be disaggregated by sex, marital status, occupation, etc., to compare various sub-populations. Being clearly defined at the beginning of the period of observation, these states or conditions set no particular methodological problems, in contrast to what we shall see with the cohort approach. It is then possible to estimate the probability of dying in each of these sub-groups, which are supposed homogeneous.

The problems arise when, to satisfy some natural curiosity, we attempt to construct synthetic measures.

For instance, we may wish to estimate the probability of survival at a given age. We can do so by combining the complements to one of the probabilities of dying from birth to the age considered. But this measure is not related to an actual cohort: it represents the cumulated effect of period mortality conditions on a *hypothetical or synthetic cohort.* To compare populations or sub-populations within a country at different dates, or populations of different countries, is therefore not so simple. In particular, it is necessary that "the different cohorts do not reach the beginning of the year studied with specific histories which to a large exent condition their mortality during that year" (Pressat, 1966). What we have is, in fact, an attempt to use period analysis to obtain cohort results, whereas the cohort is in this case hypothetical, with no real identity.

Greater difficulties arise when we study demographic phenomena such as fertility and nuptiality, which involve periods when events are postponed, followed by others in which they are made up, for instance after an economic crisis or war. As stated by Henry (1966): "during a period of recovery, behaviour is influenced by the preceding delay; to attribute to a hypothetical cohort a set of rates observed at this time means postulating the existence of a cohort which, throughout its life, would try to make up a delay which it never had".

This explains why the sum of the age-specific first marriage rates which, in a real cohort, measures the frequency of first marriages and should always be less than unity, may exceed unity in a synthetic cohort: it reached 1.5 in France in 1949, for instance. We can see the need to redefine the terms we are employing.

We suppose that a 'population' is a cohort of people born the same year in a given country or a cohort whose members all experienced a baseline event at a given time. The 'period' under study will then be the whole life of the members of this cohort, from birth on, or from the baseline event. The 'phenomena' studied will be those occurring in their lives. The 'behaviour' considered here corresponds to the occurrence of each life event, again supposed to be independent of all others. But since we are

following an individual's life course, it will be necessary to eliminate the extraneous influences of events which are not under investigation, to observe the phenomenon in the 'pure state' (Henry, 1972, p. 21).

This is the context of longitudinal or cohort analysis, which started to develop after the Second World War. The rare cohort studies conducted prior to that time did not stem from a systematic criticism of period analysis, and did not contain the seeds of the change which was to occur after the war.

In the case of first marriages, what are the hypotheses which the demographer must make to observe nuptiality in the 'pure state'? According to Henry (1972, p. 77): "Premature death prevents single persons from marrying; in this way, mortality influences nuptiality. To eliminate this effect, we would need to know how many of the single persons who died before their 50th birthday would have married had they survived. Since this is impossible, we resort to the hypothesis that the individuals who died single at a given age *a* would, had they survived beyond age *a*, have had the same marriage behaviour as those who did not die."

Given this assumption, we can estimate a first marriage rate from the surviving population, which will be the same as that calculated knowing all the individuals who, in the absence of mortality, would have been unmarried at the age considered.

This type of analysis, which observes the behaviour of real cohorts, does not raise the same objections as period analysis. It is now dominant in nuptiality and fertility studies, where it is used to compare countries or cohorts within a country. But a certain number of problems remain which cannot be resolved while the conventional sources of demographic data (vital registration, censuses...) are used.

In most manuals on demography, each phenomenon is dealt with separately. In a section devoted to nuptiality, for instance, mortality and migration will only appear as extraneous factors, the influence of which must be eliminated to obtain the 'pure state' of nuptiality. Yet, as Henry indicates (1972, p. 77): "For emigrants, we might be tempted to substitute their nuptiality abroad for the nuptiality they would have had if they had not emigrated, whereas this nuptiality abroad depends on conditions which may be very different." We see that nuptiality in the 'pure state' is not the only form worthy of investigation, since changes in the marriage behaviour of emigrants are as important as the behaviour of the sedentary population. More broadly speaking, we can say that it is necessary to analyse the *interaction* between different demographic phenomena[2]. But vital registration or census data are inadequate for this task, and other sources of information are required.

---

[2] This wish was expressed by Pressat (1966, p. 35): "We do not have the elements necessary to specify the link between the two events... Yet the study of correlations between demographic phenomena, although as yet an unexplored field, should greatly expand our knowledge". The term 'interaction' will be used throughout our paper in its statistical sense.

There is another fundamental drawback of cohort analysis, described by Henry (1959), which cannot be solved with conventional demographic data. This problem, which does not apply when the cohorts are homogeneous (an underlying hypothesis which is very frequently made in demography), stems from the statistical *heterogeneity* of the cohorts observed. Henry (1959, p. 31) has shown that: "The use of probabilities supposes that the cohorts are homogeneous, either for the phenomenon under investigation, or for the extraneous factor, or that there is no correlation between the two risks, the one corresponding to the phenomenon under investigation and the one corresponding to the extraneous factor ... Given the differences of all kinds which exist between men, we can be certain that no group is perfectly homogeneous; furthermore, empirical evidence and reflection suggest that, in most cases, there is not independence between risks".

To assume that an individual who has died would have followed the same behaviour, had he survived, as another individual actually observed, boils down to assuming this homogeneity or this lack of correlation in the population observed.

To get round this problem, which once again cannot be solved by conventional data, Henry has advocated (1959, p. 32) that: "To know with accuracy the practical level of heterogeneity in human groups, research in differential demography would have to extend to individual physical and psychological characteristics, and study both the dispersion and the correlation of demographic factors within the somewhat rudimentary groups considered so far".

In addition to such difficulties, there is the fact that these characteristics cannot be defined once and for all, but vary in the course of the individual's life. This problem did not exist with period analysis using a hypothetical cohort. The event history approach offers a means of actually tackling these problems, which were hitherto posed in purely theoretical terms.

II. – The event history approach

An extension of cohort analysis, this approach provides the means of investigating the interaction between demographic events, and population heterogeneity. We shall first examine how event history surveys have made this type of analysis possible.

Event history surveys Such surveys enable the demographer to break through the hard administrative shell which surrounds the other sources (vital registration, population registers...). They allow him/her to set a problem clearly and to elaborate a questionnaire which will yield all the information required to shed light on this problem.

To investigate the various interactions between demographic events, all these events must be recorded and dated with accuracy. The timing of their occurrence and durations between events are useful for revealing their dependence or independence.

At the same time, it is necessary to collect as much information as possible on the characteristics of the individuals and their life, in order to measure accurately the population's heterogeneity and its impact on these interactions. When the characteristics are liable to change during the life course, these changes must also be recorded and dated.

To collect information on all these events, there are two possible types of survey, both of which have advantages and disadvantages.

The first is the *continuous* or *follow-up survey*, in which individuals are followed throughout their lives. Depending on the more specific aims of the survey, the beginning of observation may be their birth or any other stage in the life course (their marriage, for instance, to study divorce, widowhood, remarriage...). In some cases, the respondents may be interviewed each year, so that a small number of events or changes will be recorded each time. In the case of a cohort sample, the first interview will, on the other hand, be much more complete, since information on all the individual's characteristics at the beginning of observation will be recorded (number of sibs, father's and mother's occupation...). The advantage of this method is that information collected year by year will be more reliable and more accurately dated. But the time between the start of observation and data analysis, which is very long since information must be collected over many years, is liable to discourage researchers. The results may also be biased by the loss of individuals who move and cannot be traced, or who at some point refuse to respond.

For these reasons, the *retrospective survey* is often preferred, since information can be collected in a single round on all of an individual's characteristics and history. The results can then be analysed directly and there is no loss of sample. There are, however, several drawbacks. First, in a survey of this kind, the schedule may be very long and cumbrous: in the Triple Biography survey (3B survey) of family, occupational and migration histories which we conducted in 1981, the interview lasted on average 70 minutes and in 10 % of cases exceeded 100 minutes (Riandey, 1985). Second, there is a risk of selectivity bias in the sample by virtue of survival. There is also a problem of inaccurate date reporting, which must be reduced as much as possible. Such errors, due to memory failure, have been tested in Belgium, where population registers made this possible (Duchêne, 1985; Courgeau, 1985). Although the sample was small (50 couples born between 1911 and 1920), we observed that, while family events were correctly dated, migrations were much less clearly situated in time (the year of migration tallied with the population register data in 47 % of cases for men and 58 % for women, rising to 73 % and 87 % respectively with an interval of more or less one year). However, an analysis of the type presented below showed that the results were not greatly modified

by inaccuracies of this kind, given that the *order* of events was correctly memorized. To confirm these findings with a larger sample, a survey of 500 couples is now under way in Belgium.

Finally, the problem of right-censored intervals[3] is encountered in both continuous surveys (unless the cohort is followed up until extinction) and retrospective surveys. This is caused by the fact that the survey stops at a given time (continuous survey) or that a person is interviewed on his/her past and we do not know what happens after the survey (retrospective survey). This information, that the person has not experienced the event under study before the retrospective survey or before the end of the continuous survey, is in itself very important. We shall now examine how to use this information.

The interaction between events We go on to consider the actual analysis, and discuss the first question that we raised: how are we to analyse the interaction between demographic events?

We present the simple case of interaction between two events, which can easily – given a large enough sample – be extended to any number of phenomena. Suppose that, in a population of men and women who started their working lives in agriculture, we wish to study the interaction between leaving farming and marrying[4]. First, we assume that the populations of farming men and women are initially homogeneous, but that heterogeneity is then introduced by change of condition (departure from the farming world or marriage). Second, we show how to introduce heterogeneity into the original population.

This original population of single farmers is subject to three types of decrement:

— some individuals will continue to farm but will marry;

— others will continue to be single but will leave farming;

— others will continue to be single farmers but will no longer be observed by the survey (right-censored intervals).

We suppose that the two events, leaving farming and marrying, cannot occur simultaneously. The population exposed to risk is in all three cases the same: the population of single farmers.

Once they are married, some will leave farming. In this case, the population at risk is the population of married farmers. In the same way, once they have left farming, some will marry; the population at risk is then the single population having left farming. There will, of course, be simultaneous withdrawal from observation of married farmers, and of single persons and married persons who have left farming.

---

[3] In demography, an interval is said to be right-censored when the beginning of the interval is known, but the observation occurs before the end.

[4] For a more detailed description, see Courgeau and Lelièvre (1986).

Although more complex, it is possible, using a method similar to that of cohort analysis, to estimate the probabilities of marrying for farmers and those who have left farming, and the probabilities of leaving farming for single and married persons. We can, of course, suppose that the probabilities of marrying for individuals who have left farming depend not only on age, but also on time spent outside farming, and reciprocally, that married persons' probabilities of leaving farming depend not only on age, but also on duration of marriage.

This method of estimation, which is based on maximum likelihood, uses both data on the persons having experienced the different events and data on censored intervals at the time of the survey[5]. It also permits estimation of the variance of the different probabilities, by which we evaluate changes in behaviour from one state to another.

In the example given here, the main findings were[6]:

— for men, the probabilities of marrying were much higher once they had left farming, while the probabilities of leaving farming did not vary with marital status;

— for women, the probabilities of marrying were the same whether they stayed or left, while the probabilities of leaving farming were very high when they were single and much lower when they were married.

This type of analysis shows the complex dependencies which can be observed between two events, when their occurrence is followed over time. We can speak of *total independence* if, at any time, the probability relative to the first event, knowing that the second has not occurred, is equal to the probability relative to the first event, knowing that the second has occurred, *and* if the probability relative to the second event, knowing that the first has not occurred, is equal to the probability relative to the second, knowing that the first has occurred. None of the analyses we have made so far has indicated a total independence of this kind between events; it is therefore necessary to analyse interaction.

If only one of these identities is always verified, while the other is not, then there is *unilateral dependence*. In our example, we see that, for men, there is unilateral dependence of leaving farming on marriage, since there is a difference between the marriage probabilities in and out of farming, but no difference in the probabilities single and married men have of leaving farming. For women, there is unilateral dependence of marriage on leaving farming. Finally, if neither of the above identities is verified, there is *reciprocal dependence* between the two events.

---

[5] For an event observed at time t, the likelihood is equal to the probability of not having experienced the event before t, multiplied by the probability at t. For an individual who has not yet experienced the event at time of survey, this likelihood is equal only to the probability of not having experienced the event before t. For further information on the estimation of this likelihood and on the method of maximum likelihood, see Courgeau and Lelièvre (1989).

[6] See Courgeau and Lelièvre (1986).

Intermediate situations may, of course, arise. For instance, total independence may be observed up to a certain age, followed by unilateral dependence and, finally, reciprocal dependence. The methods we propose make it possible to test all these possibilities.

It is also useful to take into account the date of occurrence of the previous event and to study the effect of duration between the two events. In this case, we may show that dependence exists for short durations, and disappears after a long period.

This type of analysis also reveals more subtle forms of dependence. Thus, when we analysed the fertility of women who had migrated to metropolitan areas, we saw that it was lower than that of women living in non-metropolitan areas[7]. We then wondered whether the fertility of women living in non-metropolitan areas, but who would in the future migrate to metropolitan areas, was not already different from that of women who were going to stay all their lives in non-metropolitan areas. This hypothesis was tested on women whose childbearing years were over. We were able to show that the fertility timing of future migrants was significantly different from that of women who were going to remain in a non-metropolitan area. Furthermore, this timing was not significantly different from that of women who had already migrated to a metropolitan area. This reveals an *a priori dependence* of fertility on future migration. In this case, we can no longer postulate homogeneity of the original population: we can suppose a selection of migrants who are already characterized, in their region of origin, by low fertility. It is noteworthy that for migration in the other direction – metropolitan to non-metropolitan – there is no such a priori dependence. Before migrating, these women had the same fertility as those who were going to stay in the city all their lives, but afterwards, it increased sharply.

This analysis can, of course, be extended to more complex interaction, involving many more than two events.

For instance, starting from a couple's marriage, we observe, on the one hand, births of successive children, and on the other hand, successive moves. A sequence exists within each set of events, but not, a priori, between the sets: a second birth may occur with no prior move, or only after several moves. The methods used to analyse these complex interactions are similar to those used for only two events, but they require large numbers of observations[8]. However, by making restrictive assumptions on dependence between the different events, it is possible to obtain significant results with smaller numbers[9].

---

[7] For more details, see Courgeau (1989).

[8] On the methods of analysis, see Courgeau and Lelièvre (1988). With a maximum of ten children and ten moves, we have 180 different sets of estimates.

[9] For instance, the 3B survey has shown that, for women married before their 23rd birthday, family size influences mobility: the more preceding births, the higher mobility. This is not observed for women married at ages 23 and above. In contrast, preceding mobility has an influence on family size, whatever the age at marriage. For further details, see Courgeau (1985).

Finally, we have the case of three events (or more) which follow no sequence. For instance, cohabitants may marry, have a child outside marriage, or separate, and there is no order underlying these events. Once more, the method of analysis is similar to the bivariate case considered above. The number of observations must be relatively high.

Heterogeneity of the populations observed and interaction between events

The next stage is to introduce heterogeneity. The population can be disaggregated into sub-populations by sex, number of sibs, whether the individual is the oldest child, father's and mother's occupation, etc. If we attempt to introduce all these characteristics simultaneously (as we should do), it is obvious that each of the sub-populations will rapidly dwindle to nothing, unless very large initial populations are studied.

Furthermore, certain characteristics change in the course of time and do not define stable groups. The number of sibs, for instance, may increase or decrease (when we consider surviving sibs) during an individual's life, and he/she will therefore change groups.

Finally, since we can never be sure of not having omitted a characteristic which might influence the interaction under study, it is necessary to have an idea of the error introduced by such an omission. This is the thorny problem of *unobserved heterogeneity*.

An elegant solution to these various problems is provided by Cox's proportional hazards model (1972), which postulates that all the members of a population, to whichever sub-population they belong, have the same baseline hazard, but that the fact of belonging to a given sub-population multiplies this hazard by a factor which is independent of the period studied. When a person belongs to several sub-populations at the same time, this factor is the product of the factors related to each separate sub-population[10]. It is preferable, however, to ensure, for each of the characteristics considered separately, that a multiplicative model of this kind is adequately verified, before introducing them simultaneously[11].

Characteristics which change over time are no problem with this type of model[12]. It can also be extended to the bivariate or multivariate case introducing interaction between events. This reveals the strategies of different groups of individuals, which can be read through the parameters of the model.

---

[10] At date t, the probability of marrying, for instance, will be expressed by a component characterizing nuptiality in the whole of the population and by a relative risk for each individual depending on his personal characteristics (sex, occupational status, background...). This relative risk is expressed as a proportional hazard in the model used here.

[11] For more information on this verification of robustness, see Courgeau and Lelièvre (1989), in particular chap. VII.

[12] See, for instance, Murphy (1984).

To illustrate the method, we reconsider the women who started their working lives in agriculture[13], whom we previously supposed were homogeneous at the time of their entry into the labour force. The analysis shows that this assumption does not hold, and that various characteristics will strongly influence their chances of leaving farming or marrying. Before marriage, the women most deeply rooted in the farming world are eldest daughters with few sibs and a farming father. When they have all these characteristics, their probability of leaving farming is four times lower than that of women who have none of them. The marriage effect we observed above is essentially due to women who marry a farmer's son who farms. When women combine these two sets of circumstances, their probability of leaving farming will be almost ten times lower than for women who do not have these characteristics. This can be interpreted as a strategy aiming to group the two family farms together and allow the couple to stay on the land. Furthermore, in view of the size of the sample, the effect of the different characteristics on the probability of leaving farming could be calculated and we could thus verify their statistical significance. This made it possible to distinguish several groups of women, with contrasting behaviour as regards both leaving farming and marriage.

If the survey questionnaire covered all the different characteristics which might lead to withdrawal from farming or to marriage, the methods of analysis presented here would apparently solve all the problems we initially encountered[14]. Unfortunately, we can never be sure of having collected information on all the aspects which might explain a situation. Consequently, we may question the validity of our estimates. We are confronted with the problem of *unobserved heterogeneity.*

We present a certain number of results which researchers have obtained using purely parametric models, in which the time-dependence of probabilities is fully defined[15]. In this case, certain authors have drawn attention to the dangers of modelling unobserved heterogeneity without adequate information. Heckman and Singer (1984) have shown that the influence of certain characteristics may be considerably modified by the distribution chosen to represent this heterogeneity. Their effect on the probability that an individual will experience an event may even be positive with some distributions and negative with others. Since we generally have no notion of the form of the heterogeneity, this finding does not inspire confidence about the robustness of the results yielded by a parametric model of this type.

---

[13] For further details, see Courgeau and Lelièvre (1986). Interestingly, many of the conclusions of this analysis confirm those of the sociologist M. Chaudron (1987) in her study of 14 sib sets from two farming families.

[14] Assuming a multiplying effect of the different characteristics. Other types of model may, of course, be developed for a better fit of the situations observed.

[15] For instance, the Gompertz model used to study migration assumes an exponential dependence on duration of stay. In this model, various characteristics may be introduced by multiplication. Other distributions may express the effect of duration: for example, a log-normal distribution may be used to study nuptiality.

Further, Trussel and Richards (1985) have shown that the conclusions of the analysis depend not only on the distributions chosen to represent unobserved heterogeneity, but also on the type of time-dependence of the characteristics introduced.

Once more, Cox's semiparametric model, which we have presented here and used frequently elsewhere, proves much more robust than these parametric models, since it integrates some of the heterogeneity in the baseline hazard. More recently, Bretagnole and Huber-Carol (1988) have determined the effect of this unobserved heterogeneity. The positive/negative influence of a characteristic, estimated by omitting certain elements which affect the phenomenon under study[16], is seen to remain positive/negative when all these elements are introduced. Thus, it is not surprising that the results we obtained with incomplete biographical data on the farming world from the 3B survey confirmed, for the whole of France, the findings of a much more detailed study on the Lauragais region (Collomb, 1984) or those of a genealogical study of a few farming families (Chaudron, 1987). Bretagnole and Huber-Carol show, however, that the influence is sytematically underestimated when elements are omitted. This result shows the limits of the conclusions we can draw in the absence of unobserved heterogeneity. If a characteristic has a significant effect (positive or negative) on the occurrence of an event, this will remain significant when all the elements not observed in the first stage have been introduced: in this case, the effect will be heightened. On the other hand, if the effect is not significant, it may become significant when all the elements have been introduced. We can therefore never be completely sure of the absence of influence.

*
* *

Starting with the period approach, which analyses the events occurring at a given moment, we have discussed the limits of this method when it is used to demonstrate changes over time by postulating an underlying stability of the behaviour of different cohorts.

We saw that to follow these events through an individual's life course was a much more satisfactory basis for cohort analysis. But the conventional sources of data limited this approach, since they did not provide the means of investigating either interaction between events or population heterogeneity, which remained a major problem for analysis.

A great step forward was taken with the wealth of biographical information collected by event history surveys. This made it possible to elaborate methods of non-parametric analysis extending the techniques of cohort analysis to much more complex situations. These methods were then combined with the regression analysis prized by economists, to obtain semi-parametric models. The result is a powerful tool which simultaneously

[16] These elements should be independent of the characteristic, a condition always set when non-observed heterogeneity is introduced.

integrates interaction between events and population heterogeneity. But while certain of these models are well developed (the proportional hazards model, for instance), other more complex situations remain relatively unexplored. The use of these models is also linked to the size of the survey sample. Finally, the problems set by unobserved heterogeneity are partly resolved. We now know that a demonstrated dependence will not disappear or will not change direction when this unobserved heterogeneity is introduced.

The event history approach in demography could then find its place among the longitudinal approaches of the other social sciences. An interdisciplinary working group of anthropologists, demographers, geographers, historians and sociologists adopted the following definition of the longitudinal approach (Lelièvre and Courgeau, 1987): "Longitudinal approaches are characterized by the study of objective or subjective events or states, their sequences and interactions, in relation to a historically defined time, and as experienced by a same entity (individual, family, organization...) within a well defined group (cohort, generation...). Longitudinal approaches have precise aims, to be met by various methods of data collection and analysis, and by specific models."

We can now show how the demographic approach differs from that of other social sciences, and the dimension it may add. We shall essentially compare the demographic approach to the sociological one.

First, demography focuses more on objective events or effects (marriage, migration, etc. and conditions such as 'married', 'present in a place of residence'), leaving subjective events or states to other social sciences (psychology, for instance). Second, micro-demography examines events which have happened to the same individual. It is also beginning to study more complex entities, such as families (Keilman, Kuijsten and Vossen, 1988). The larger organizations and entities are the domain of sociology, history and anthropology.

The demographic approach stands even more apart from the other social sciences by its objects and its methods of data collection and analysis.

As we have seen, in studying biographies, the demographer's aim is to analyse the course of the different events occurring in an individual's lifetime. The individual's position in society is simultaneously defined by a multitude of personal characteristics, some of which will change in the course of his/her life. It is by analyzing these that we observe the society in which (s)he lives, intervening with its constraints and obligations, which (s)he may attempt to escape. This is where the scope and limits of an individual's liberty can be perceived. That is why the demographer chooses to study probabilities, which eliminate all traces of determinism. These probabilities may moreover change during the individual's lifetime.

The sociologist takes "as starting point, not individuals or families, but 'society' itself (a given socio-historic formation). A 'society' cannot

operate unless it finds people to accept the different types of tasks which the history of production relations (and the resulting division of labour) have instituted... If all these positions are filled, even those not a priori attractive to anyone, it is because there exists a general process, a *constraining societal process*, by which the members of a society are oriented, whether they like it or not, channeled off, shared out, allocated to the different posts as defined by a (socially constructed) state of the division of labour. When one has conceived of the existence of such a process (which I have proposed elsewhere, as a process of anthroponomic distribution), one has an acceptable definition of the sociological object" (Bertaux, 1987). The study of biographies aims to identify the underlying social processes (Pitrou *et al.*, 1983).

The data collected by the sociologist may be (but are not necessarily) different from those of the demographer. The latter needs information on a large enough number of individuals to permit estimation of the probabilities that the different events will occur. He/she will therefore investigate a sample of individuals representative of the population or sub-population (s)he wishes to observe. The questions asked will be mainly closed-ended, so as to date the different events as accurately as possible and to situate them in relation to one another, while at the same time recording a maximum amount of information on the respondent's characteristics. In contrast, the sociologist may use a much smaller number of interviews to demonstrate the presence of the societal process which underlies these biographies (Léomant and Sotteau-Léomant, 1987; Menahem, 1988). These life histories will naturally be much more detailed than those recorded by the demographer. But information on the exact timing of the different events will not be necessary, as the sociologist does not need to measure the probabilities that the different events will occur. His/her concern is to detect and identify the regularities which appear beneath the diversity of individual experiences.

We can see how this approach, operating on a different level, can be complementary to the demographic approach. This is all the clearer when we observe the methods used.

In event history analysis, the demographer employs the methods we have developed in this article, to demonstrate interaction between phenomena and heterogeneity in the populations observed. When events initially considered as having no influence on the phenomenon under investigation are introduced, we can observe new significant effects. They will then modify the probabilities priorly estimated. These probabilities – in the words of Costa de Beauregard (1988) – express nothing more than an estimate of chance based on what we know (or think we know).

For the sociologist, on the other hand, the range of possible methods is much more open (see, for instance, the article by de Coninck and Godard, 1990). It is obvious that extremely detailed life histories collected for a very small number of individuals cannot be analysed using the methods presented above. But the great variety of social contexts which emerges

from such observations limits the generalization of the results. They can, on the other hand, bring to light characteristics which the demographer, working on closed questions, may have overlooked, and which may have considerable impact on the phenomenon under investigation.

We have, in fact, attempted here an epistemology of demography. It seemed useful that demographers, who are not generally accustomed to questioning their statistical practices or the implicit theories which underlie their analyses (Blum and Blanchet, 1988), should step back and consider the value, the scope and the origins of their science.

Daniel COURGEAU and Eva LELIÈVRE

REFERENCES

Bertaux, D., «Du monopole au pluralisme méthodologique dans la sociologie de la mobilité sociale», *Annales de Vaucresson*, 1987, no. 26, 305-319.

Blum, A., Blanchet, D., «L'explication en sciences sociales. La recherche des causes en démographie», *Population*, 1988, **4-5**, 914-920.

Bretagnolle, J., Huber-Carol, C., «Effects of omitting covariates in Cox's model for survival data», *Scandinavian Journal of Statistics*, 1988, **15**, 125-138.

Chaudron, M., «La constitution des intinéraires sociaux sexués dans la fratrie d'origine», *Annales de Vaucresson*, 1987, no. 26, 75-90.

Collomb, P., *La mort de l'orme séculaire. Crise agricole et migration dans l'Ouest audois des années cinquante*, Paris, 1984, INED-PUF (Travaux et Documents no. 105-106).

Costa de Beauregard, O., *Le temps déployé*, Monaco, 1988, Editions du Rocher.

Courgeau, D., «Les réseaux de relations entre personnes. Etude d'un milieu rural», *Population*, 1972, **4-5**, 641-683.
— «Effet de déclarations erronées sur une analyse de données migratoires», *Chaire Quetelet 83: Migrations internes*, 1985, Louvain-la-Neuve, 151-156.
— «Interaction between spatial mobility, family and career life cycle: a French survey», *European Sociological Review*, **1(2)**, 139-162.
— *Méthodes de mesure de la mobilité spatiale*, Paris, 1988, INED.
— «Family formation and urbanization», *Population: An English Selection*, 1989, **1**, 123-145.

Courgeau, D., Lelièvre, E., «Nuptialité et agriculture», *Population*, 1986, **2**, 303-326.
— «Estimation of transition rates in dynamic household models», in Keilman, N. *et al.* (eds.), *Modelling Household Formation and Dissolution*, Oxford, 1988, Clarendon Press, 160-176.
— *Analyse démographique des biographies*, Paris, 1989, INED.

Cox, D. «Regression models and life tables», *Journal of the Royal Statistical Society*, 1972, **8, 34**, 187-220.

De Coninck, F., Godard, F., «L'approche biographique à l'épreuve de l'interprétation. Les formes temporelles de la causalité», *Revue française de sociologie*, 1990, **XXXI**.

Duchêne, J., «Un test de fiabilité des enquêtes rétrospectives. Biographie familiale, professionnelle et migratoire», *Chaire Quetelet 83: Migrations internes*, 1985, Louvain-la-Neuve, 135-149.

Heckman, J., Singer, B., «A method for minimizing the impact of distributional assumptions in econometric models for duration data», *Econometrica*, 1984, **2**, 271-320.

Henry, L., «D'un problème fondamental de l'analyse démographique», *Population*, 1959, **1**, 9-32.
— «Analyse et mesure des phénomènes démographiques par cohorte», *Population*, 1966, **3**, 465-482.
— *Démographie: analyse et modèles*, Paris, 1972, Larousse.

Keilman, N., Kuijsten, A., Vossen, A. (eds.), *Modelling Household Formation and Dissolution*, Oxford, 1988, Clarendon Press.

Landry, A., *Traité de démographie*, Paris, 1949, Payot.

Lelièvre, E., Courgeau, D., *Approches longitudinales*, Colloque Sociologie III, 1987, Strasburg.

Léomant, C., Sotteau-Léomant, N., «Itinéraires de vie et trajectoires institutionnelles de jeunes délinquants», *Annales de Vaucresson*, 1987, no. 26, 199-222.

Menahem, G., «Trois modes d'organisation domestique selon deux normes familiales font six types de famille», *Population*, 1988, **6**, 1005-1034.

Murphy, M., «The influence of fertility, early housing career and socio-economic factors on tenure determination in contemporary Britain», *Environment and Planning*, 1984, A1B, 1303-1318.

Pitrou, A., Battagliola, F., Buisson, M., Rouffier, N., Toussaint, Y., *Trajectoires professionnelles et stratégies familiales*, 1983, ATP du CNRS, 404 p.

Pressat, R., *Principes d'analyse*, Paris, 1966, INED.

Riandey, B., «L'enquête Biographie familiale, professionnelle et migratoire (INED, 1981). Le bilan de la collecte», *Chaire Quetelet 83: Migrations internes*, 1985, Louvain-la-Neuve, 117-133.

Trussel, J., Richards, R., «Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure», in Tuma, N. (ed.), *Sociological methodology*, 1985, London, Jossey-Bass, 242-276.