Statistical Journal of the United Nations ECE 6 (1989) 1–15
IOS

Prospects for socio-demographic risk research in Denmark

Lars Thygesen

In Denmark, official statistics on persons are based almost exclusively on administrative registers. The data base for the statistics contains longitudinal data on many socio-demographic variables for each citizen. This seems to make it possible to solve some of the problems connected with current socio-demographic risk research, e.g. in social medicine. However, there seems to be a lack of appropriate methods to make use of the new data opportunities. Multivariate statistical models should obviously be employed. But it is necessary to introcude a number of assumptions, implying an oversimplification of the real world. Will it still be possible to gain new knowledge on society by using these methods? It is believed that the new studies can be carried out and lead to useful results. Danmarks Statistik intends to initiate the studies.

1. Introduction

The purpose of this article is to formulate the demands on the analytical methods that we would like to use for the large general screening studies of the central statistical offices. What sort of questions are we seeking an answer for? And can modern statistical theory help us?

By screening studies I mean attempts at mapping how the risk of some event ("effect variable") is correlated with one or several characteristic features ("explanatory variables") of the whole population. Examples hereof are the studies of "occupational mortality" and "cancer and occupation". But the studies can also deal with unemployment, criminality or social security services.

The weakness of such studies is that only one explanatory variable at a time has been included (e.g. occupation). In order to overcome this weakness attempts have been made at using multivariate methods.

Another weakness is that it has not been possible to take into account the fact that the explanatory variables change over time. For example, the death rate of people who at a particular time were employed as painters is compared with that of farmers. Perhaps they had previously been employed in a different occupation, or they may have changed occupation during the period of observation. This can of course to a great extent disturb the comparison.

Lars Thygesen was born in 1945. He obtained a Masters Degree in economics and statistics in 1970 and has been employed in Danmarks Statistik since then. In 1979, he was appointed Chief of the register administration with responsibility for data policy and development of register-based statistics. Since 1987 he has been Director with responsibility for social and demographic statistics.

In Denmark (and in some other countries as well), the basic registers are now so advanced that a *history* can be compiled for each person with regard to the most important explanatory variables: occupation, housing conditions, marital status, etc.

This could lead to the possibility of answering questions such as:

- How risky is it to be a painter for 5 years?
- How great is the risk of cancer for painters living in a single-family house compared to the average when it has been taken into consideration that the persons have had a different occupation and other housing conditions for some part of their lives?

This article is based on the basic data material kept at Danmarks Statistik. Concrete examples of comparison of data and formulation of issues in the studies are discussed.

The problem is actually to find a method of concentrating the vast quantity of data so that they become intelligible statements.

2. Screening studies

Individual-based longitudinal studies have been used for many years in order to conduct national "screening studies" in the Nordic countries. Here it has been attempted to define a person's risk of experiencing a particular event (e.g. death) and compare this risk for various groups of the population (e.g. various occupational groups). The hypothesis is that there is some causal relation between "*the explanatory variable*" (occupation) and "*the effect variable*" (death). The purpose of the screening surveys is either to confirm or to deny this hypothesis and to obtain insight into the numerical relationship: which groups have a hight risk and which a low risk?

Examples of screening studies are the studies of "occupational mortality"[1, 8] and "cancer and occupation" [9]. But the studies can also deal with unemployment, criminality or births.

3. Methodological problems

In practice it is difficult to arrange screening studies so that tangible conclusions as to the causal relationship can be drawn. This is due to several circumstances described in the following.

3.1. Selection

A serious and presumably insoluble problem is that of selection: which people become, for example, waiters or farmers? Are they equally healthy when they begin their careers? Or do they from the beginning differ from the average as regards, for example, living habits? The problem of selection is discussed in

virtually all the previous studies (e.g. [1] and [8]). It will also have to be carefully considered in the prospective studies discussed here.

3.2. One or several explanatory variables

If one wants to analyse one explanatory variable together with an effect variable a number of "disturbing factors" can often be identified. In many of the studies conducted on population-wide data, only one explanatory variable at a time has been considered.

Some of the disturbing factors are the classic demographic variables: sex and age. In the surveys mentioned an attempt has been made to control these variables by analysing men and women separately, and by using traditional standardization techniques or suitable mathematical models for controlling the age [1].

However, other variables will also disturn the comparison. If it turns out, for example, that the mortality for waiters is three times higher than the mortality for farmers then it may of course be due to risk factors in connection with the performance of a person's occupation. But it is likely that the picture is more complicated. It should be taken into consideration that waiters have poorer housing conditions, live mainly in less healthy urban areas, etc. They may have different drinking and smoking habits. Consequently, more explanatory variables have to be included in the analysis. This has been attempted in some new studies [1], where multivariate statistical models have been employed. The studies have not all been successful, as the data seldom fit the mathematical models. But as regards this point new methods are being developed.

3.3. Explanatory variables change over time

I shall now discuss the central problem of my paper. This is the question of the *time dimension of the explanatory variable.* If it is true that it is risky to be a waiter and that it increases the death rate then the question is: for how long does one have to work as a waiter before the risk rises? Is it sufficient to be a waiter for one day? Or is it worse if one has been a waiter for 10 years? Does it affect one immediately or after some years?

In the studies of mortality and cancer mentioned the following method has been employed: the explanatory variables are measured at one specific point in time, and we observe the development in the effect variable for a fixed period (10 years) after this time. This method has to a great extent been decided by the practical possibilities, as it has only been possible to measure the explanatory variables such as occupation, housing conditions, dwelling in urban/rural districts for the whole population at times of population and housing censuses. But this contributes greatly to making the conclusions of the surveys unreliable. The persons may have changed their occupation just before or after the time of measurement and consequently they may have been exposed to totally different influences. Very mobile occupations can thus appear with misleading results if we accept the hypothesis that exposures on the job actually affect one's health. However, this problem is often underlined in the studies published [1].

4. New data possibilities

As already stated, the existing data basis has not previously made it possible in practice to include the information on course of events in the explanatory variables. During recent years this situation has changed in the Nordic countries. It has become possible to a still greater degree to collect basic information for the statistics from the administrative EDP systems on the social circumstances, etc., of the populace.

The complete 1981 Population and Housing Census (in short: Census) in Denmark was thus solely based on registers [4, 11]. And since 1980 all Census data have been collected for all persons for each year. A kind of information history has thus been collected for occupation, industry, duration of employment, income, education, housing conditions, degree of urbanization of the person's residence and much more. The data material then permits more detailed screening studies in the future. However, it is still only a few years (6 years) that have been covered. Therefore, the new knowledge we could expect to obtain today is probably very limited. But as time goes by the possibilities will be improved. The time has come to start the screening surveys of the future.

However, the problem is how to make use of the new knowledge. Are there already suitable methods? Fortunately the answer seems to be positive.

5. Occupational mortality

5.1. Earlier studies

The method employed in the studies carried out in Denmark [1] has been:

- the starting point has been taken as the population at the 1970 Census;
- the explanatory variables, particularly occupation and industry, refer to the time of the Census;
- the individuals are followed during the period of observation (5 or 10 years) or until they leave the population as a result of emigration or disappearance;
- death rates for the observation period are given by five-year birth-cohort (cf. Figure 1);
- the death rates for the individual occupations are compared to death rates of all economically active persons by means of an indirect standardization of age (Standard Mortality Ratio, SMR); it shows how many persons die in occupation No. 1 in proportion to the number that would have died if the death rates had been as for all economically active persons.

It must be added that Ref. 1, as a theoretically more satisfying alternative to the SMR, also contains an estimation of an occupational parameter in a multiplicative intensity model. The results of the two types of calculation are, however, by and large identical. It is also worth noting that the study includes several explanatory variables (housing conditions, etc.), which are analysed in a bivariate model as well as in a multivariate analysis, together with occupation.

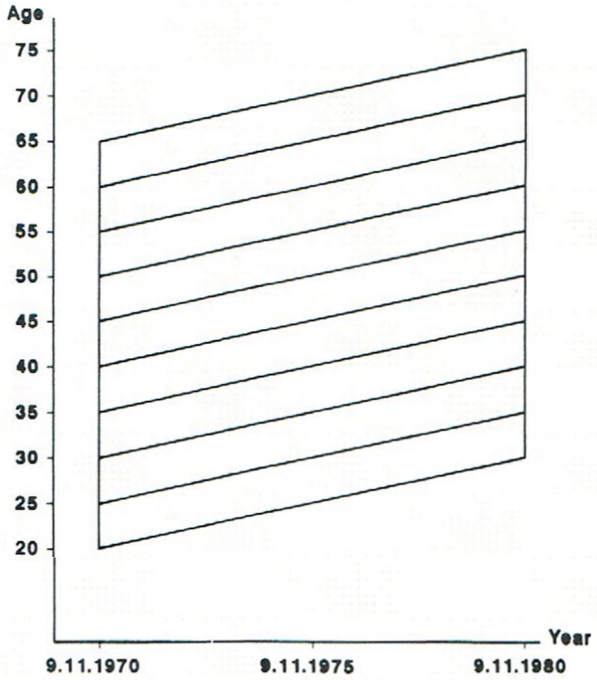

Fig. 1. Lexis diagram. Principles for the calculation of age-specific death rates.

5.2. The new possibilities

As mentioned, there are in Denmark new data possibilities as regards the most important explanatory variables: the duration of employment, occupation, industry, income, family and cohabitation conditions, degree of urbanization of a person's residence area.

All these data are now currently collected from administrative register systems. Most of the data are collected once a year with reference to either the past year or a time close to the turn of the year. This is not an actual registration of events, as is known from the registration of, for example, deaths.

The total statistical system is rather complicated and will not be described here; see Ref. 5. The core of the system is the Central Population Register, which contains basic information on every citizen, coded by a unique Person Number. The Person Number is used throughout the public administration as a common identifier, allowing excact matching of data from many sources. The Danish data protection legislation permits such matching for purely statistical purposes [15].

This is the basis for statistics on a total population basis, allowing combinations of all data items. The Population and Housing Censuses in Denmark are now being carried out based solely on the registers.

Generally speaking, the data quality is regarded as very good, compared to

traditional surveys. But there are measurement problems related to some items, notably occupation. Of course this may have rather serious consequences for analysis of the type discussed here. On the other hand, preparatory studies have indicated that these problems will not invalidate the whole analysis.

In the following I have concentrated on a description of the occupational dimension, although other variables of course ought to be included in the coming studies.

For each year since 1980 data have been collected on the duration of employment, industry (i.e. the kind of activity of the workplace) and occupation for each individual. The data are kept in annual registers which can be linked to form a history for each person.

The data are based on a large number of registers run by the Danish tax authorities and the Ministry of Labour, supplemented with the public employers' wage and salary systems. One of the most important source registers is the tax authorities' Salary Information Register, containing information from the employers on payments of wages and salaries for each job that a person has had in the course of the year. Here data on the employer, the workplace, the period of employment, wages and salaries and the approximate hours of work are found.

On the basis of these administrative records, data on employment are created for each person in Danmarks Statistik's register according to two principles: a Year Principle and a Date Principle.

According to the Year Principle the person is classified by the duration of employment, industry and occupation: what was the main occupation, etc., of the person during that year? The Date Principle attempts to show the situation of each person at a particular time near the end of the year.

5.3. Design of the study

The basic *hypothesis* must be that the performance of a job is an important factor that affects the death rate.

The earliest annual occupational information in Denmark stems from 1980/81. It would therefore be best to take as starting point the *population* residing in Denmark on 1.1.1981 and who were also present throughout 1980. The latter requirement is due to the fact that occupational information will be doubtful for persons immigrating in the course of the year.

The age range in the study is, according to traditional practice, 20–64 years. *Age* is defined as from 1.1.1981.

Figure 2 illustrates the available data. The figure shows the period of observation for persons aged 20–24 years on the 1st of January, 1981.

The occupational information is available according to the Year Principle for the years 1980–85, represented by the groups M_0–M_5 in Fig. 2. In accordance with the Data Principle data are available for the period from 1.1.1981 to 1.1.1986, represented by the life lines that intersect the line segments BJ, CK, DL, EM, FN and GO.

In the light of the hypothesis stated it seems expedient to define the occupational information on the basis of the *Year Principle*, which indicates the predominant connection (or lack of the same) to the labour market during the year.

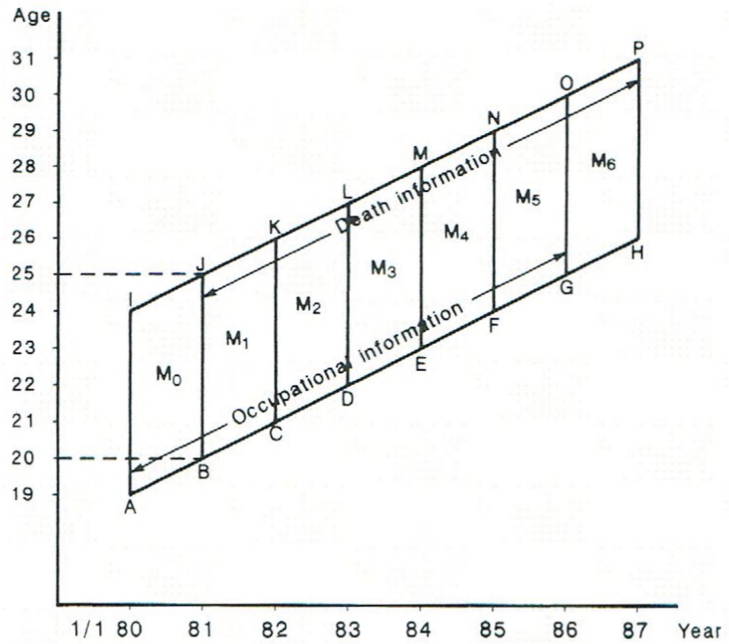

Fig. 2. Lexis diagram. Persons aged 20–25 years.

5.4. The mortality measurement

The population is kept under observation for the *effect variable* (death) during the period 1.1.1981 to 1.1.1987 corresponding to the groups $M_1 - M_6$ in the figure.

A person leaves the population in the event of death, emigration or disappearance.

For each person (i) a *risk time* (r_i) is calculated from 1.1.1981 until the exit from the population as a result of death, emigration or disappearance; if the peron has not left the population the risk period is calculated until the end of the observation period, 1.1.1987.

For each person group defined by age (x) and "occupational history" (u) an age-specific death rate M_{ux} can be calculated:

$$M_{ux} = \frac{D_{ux}}{R_{ux}} \tag{1}$$

where D_{ux} is the number of deaths, whereas R_{ux} is the sum of the risk periods of all persons in the group:

$$R_{ux} = \sum_{\substack{(\text{age} = x) \\ (\text{history} = u)}} r_i \tag{2}$$

5.5. The occupational dimension

The problem is of course that the variable "occupational history" is complex. A value of u could be:

"Was employed as a baker 1980–82, unskilled worker in the textile industry 1983–84 and left the population in 1985."

In previous studies an occupational variable has been available with generally speaking only one dimension. It has been common to operate with about 350 occupational groups. It has been natural to ask questions such as:

A: "How risky is it to be a baker?"

The answer required an estimate of an index of mortality for bakers compared to the average of economically active persons.

The type of question which the new kinds of studies seem to be able to answer could be as follows:

B: "How much does it increase the death risk after 1 year, 2 years, etc., to be employed as baker for 2 years as a 24–25-year-old man, when one has left this occupation and has been employed in the textile industry as an unskilled worker?"

To illustrate question B, one could imagine that the death rates for the selected group of occupational histories are to be compared to the death rates for persons of the same age who had the same history apart from being employed as a baker for the first two years.

It is obvious that such comparisons will be unmanageable.

If the ordinary variable "occupation" for each year has n potential values and the survey covers t years then u will have

$$\frac{n\,(n^t-1)}{n-1}$$

potential values. This will result in a vast number if, for example, one operates with 100 industrial groups. The number of observations in most groups will be zero or close to zero.

It is also a great problem that the type of statement that could be put forward would be unintelligible to the human brain. We still demand to be informed of how risky it is to be employed as a baker, and not about 1 million different assertions of the risks of the baker's trade at various ages and in connection with several other trades.

We demand a simplification in order to make the results useful.

5.6. A simplified model

In order to get a general view of the situation, it is necessary to produce a model with simplified preconditions.

M_{ux} is considered to be an estimate of the *death intensity* μ_{ux} for the group in question:

$$M_{ux} \approx \mu_{ux} \tag{3}$$

It is the death intensities that we want to analyse. The first simplification that I will propose is an assumption that a possible excess risk for a group manifests itself proportionally for all age groups. That is, all the death intensities are a product of an *age parameter* λ_x and an *occupational history parameter* θ_u:

$$\mu_{ux} = \theta_u \cdot \lambda_x \tag{4}$$

θ_u now indicates the excess mortality associated with exactly the occupational history u.

This type of model is widely used in epidemiological studies [3, 6]. It has been tested in some of the previous studies ressembling the one described here, e.g. Ref. 1, and has unfortunately proved untenable in most cases.

In order to simplify the occupational history parameter it is so far my opinion that it is necessary to presuppose that the effects of the various parts of the occupational history are mutually *independent*; e.g. being employed as a baker for 2 years contributes equally to the death risk irrespective of a person's future or previous occupational history.

Also I am compelled to presuppose that the influence on the death risk resulting from being employed as a baker will appear immediately and thereafter remain "forever". This influence may be assumed to depend exponentially on the time of exposure.

These statements can be formalized so that the occupational history parameter is the product of parameters γ_v for the individual *occupations* v of which the history consists:

$$\theta_u = \prod_t \prod_v \gamma^{j_{vt}} = \prod_v \gamma_v^{a_v} \tag{5}$$

where $j_{vt} = 0$, if the history does not comprise occupation v in year t;
$j_{vt} = 1$, if the history comprises occupation v in year t.
a_v is the total number of years employed in occupation v,

$$\text{i.e. } a_v = \sum_t j_{vt}$$

γ_v means the excess mortality effect of one year employed in occupation v.

The occupation in the year in which a person departs from the population (death or emigration) constitutes a special problem. It would be possible to include a person's occupation in this year in proportion to the time he has lived, so that the sum of a_v is equal to the total risk time. Another possibility is that the year of withdrawal is not included in the occupational history.

The vector of the number of years employed in an individual occupation is called α:

$$\alpha = (a_v;\ v = 1,2 \ldots, n)$$

Persons with the year vector α have the death intensity:

$$\mu_{\alpha x} = \gamma_x \prod_v \gamma_v^{a_v} \tag{6}$$

If the model is applicable, we would be able to answer the question:

C: "How much does it increase the death risk to be employed as a baker for one year, compared to all economically active persons?"

5.7. Is the model reasonable?

There is no doubt that the preconditions that I have put forward represent an oversimplification. The question is whether it makes the model inapplicable in practice and whether improvements can be suggested.

I must emphasize that here I am only aiming at getting a broad picture of the structure in occupation-related risks. Such a picture must be based on an oversimplification. It will still be necessary to carry out intensive analyses of the conditions of the individual occupations and exposures where assumptions are made more realistic and based on concrete knowledge of the conditions for exactly these occupations.

As mentioned, it is common practice to assume in risk studies that deviations in the death intensities for the various ages are proportional. But a test of this part of the model has in many instances shown that the assumption does not hold.

A weakness of the model is that it does not distinguish between different causes of death. Presumptions as well as experience make it reasonable to assume that the multiplicative model for the death intensity should hold if a particular cause of death was considered. If the model holds for the individual specific causes of death it will (normally) not hold for the sum of all causes of death as a whole.

The assumption that a particular occupational exposure works immediately and with a particular strength, which will never change, is to a certain degree at variance with generally accepted assumptions. For example, working environmental exposure to a well-established carcinogenic effect (e.g. silicon dust) is known to affect the cancer risk with a delay of some 20 years. Conversely, it must be expected that some exposures have a strong effect during and immediately after the performance of an occupation, and that an occupational-related excess risk will thereafter diminish over time. However, it is difficult to say if these conditions make the model inapplicable in practice.

Is it possible to improve the model in this respect? Maybe it would strengthen it if an assumption was added that the exposure from occupation v changed in proportion to the time elapsed since the exposure, for example, with a factor $1 + r_v$ per year; r_v could be positive (e.g. silicon dust) or negative. But such an assumption is probably not easy to handle.

There is reason to be sceptical of the assumption that the different occupational exposures are independent and that the order in which different occupations appear is unimportant.

The assumption that the risk exposures involved in being employed in occupation v increase exponentially in proportion to the time employed in the occupation seems immediately fairly reasonable. However, with today's limited

period of observation (6 years) it might be a bit tricky to use it: we do not know which occupations the person had before 1980. The best guess might be that previous occupations were similar to those in 1980. This would imply that the occupation in 1980 should have a greater a_v, especially for those age groups that have been on the labour market for a long time. Another possible assumption is that the previous occupation corresponds to "the average" of the experience of the past 6 years. This would mean that old persons have been exposed to a particular risk for longer than the 6 years – and longer than young persons. If this is right the exponents a_v should actually have been higher for old persons.

The problem is insoluble as long as we only have data for a short period such as 6 years. The accumulation of the registrations will help us over time. Until then it is perhaps a good idea to supplement the survey with historical data on the population's employment data at the last traditional Census in Denmark in 1970.

My conclusion here is that the model is presumably not sufficiently realistic. New ideas are needed of how to approach the matter. However, we are going to start the study using a model like that described here, being prepared to modify it along the way.

5.8. Can the parameters be estimated?

As earlier studies have shown [3], there are no theoretical problems in the estimation procedure. It seems, however, that the methods have not previously been used on huge amounts of data materials like those considered here. This may bring about serious data-processing problems.

6. Fertility and occupation

6.1. Issues

The purpose of a study of occupational fertility is to throw light on who has many children and who has few or no children. In the light of the dramatic changes in fertility during the last 20 years it would also be interesting to throw light on the groups that have particularly been affected by the decline of fertility during the 1970s.

Some of the most important background factors of such an analysis are, apart from the ages of the potential parents, their occupations. However, no major studies have been conducted on the relationship between occupation and fertility in Denmark since the turn of the century [12], nor am I acquainted with such studies from other countries.

The reason is probably, among other things, that the methodological problems are very great. The question of whether there was a fall in the fertility in the 1970s or merely a postponement of births to a later time in women's lives has been very difficult to answer.

It is also quite clear that occupation in relation to fertility can only in a very indirect way be considered to be an "explanatory variable".

Nevertheless, the question of such a survey is gradually becoming urgent. This is not only due to the existence of the many data possibilities (cf. section 4). The task is immense and will require outside financial support.

6.2. A cohort study

Like the mortality study, the fertility study must be a cohort study. All persons aged 15–49 years must be followed, not only those who have children.

The analysis must deal with the fertility of both men and women. In Denmark this will be something new, as so far births have been analysed only in relation to women. In this connection it will be necessary to supplement the Birth Registers with data about the fathers of about 40 per cent of the children who were born out of wedlock. This can be done by means of the Central Population Register (CPR) as long as the father is known and registered, which is the case for some 98 per cent of the children.

6.3. Individuals or families?

An important question is which *objects* are to be followed: individuals or families (couples).

Individuals are "stable" and unambiguous objects which are easy to follow. The problems in this connection are similar to those described in the preceding sections for mortality studies. Entry into and exit from the population can easily be described. Occupational histories, etc., can be collected. The situation will be similar to that illustrated in Fig. 2, above.

Questions to be answered could be:

D: "How likely is it for a male baker to have a child in the course of t years compared to the average man?"

E: "How much does it increase the probability of having a child during t years if a person has been employed as baker for one year?"

These questions aim at a simplified model like that proposed in section 5.

However, it would be natural to see the risk (or probability) of births as a function of attributes and attitudes among those *couples* who may become parents or who refrain from becoming parents. These attributes consist partly of attributes of the two parties, partly of attributes appearing by way of an interaction. For this reason we wish to follow the object "the couple" instead of the persons.

For practical reasons it is necessary to concentrate on couples who have a certain degree of permanence and whose existence can be recorded in the registrations. In practice, this means married couples living together and persons cohabiting, which can be considered to be "a kind of marriage" either because they have joint children or because they have made some legal declaration.

The background data on couples must be historical data on the couple's existence as a couple (including family circumstances) and the two parties'

occupational history, etc. How are all these data simplified into data that can be managed?

A complication in connection with couples is the dissolution and establishing of new couples. The date of the wedding can of course be recorded, but in other instances it can be difficult to determine the beginning and end of a relationship. Change of address must also be taken into consideration, as it is reasonable to require that married couples must live together in order to be included in the "risk population" of couples. Furthermore, it must be borne in mind that the establishing of a "permanent" relationship between two persons (either married or unmarried) could be caused by a birth as well as the reverse.

It could also happen that one of the persons of a couple has a child with a third person.

In spite of all these complications, it is still my view that an effort should be made to include couples as objects in the study of occupational fertility.

6.4. The fertility measurement

The fertility is measured by fertility rates (f) for each group of objects defined by means of the background variables:

$$f_{ux} = \frac{l_{ux}}{R_{ux}} \tag{7}$$

where u signifies the background variable, e.g. occupational history, x is the age, l is the number of live births and R is the risk period.

6.5. The occupational dimension

The hypothesis in the study does not require a very detailed occupational classification, but a fairly rough socio-economic classification of around 20 groups is considered. This means that the number of movements between the groups will be more limited. However, movements into and out of occupations must be carefully observed.

A difficulty as regards births is that the parents are often young and their occupational career has perhaps hardly begun. Many of them may still be receiving education. It can be assumed that young persons will have some fairly definite expectations of their future occupations which will influence the fertility level in advance. This could lead us to use as an explanatory variable the occupational history completed later, rather than the history that precedes the birth. It is thus possible to speak of the fertility of groups that later become, for example, managerial employees.

However, it must be considered that the causal relationship may work in two directions: not only is fertility dependent on occupation, but the choice of occupation may also be influenced by birth history.

6.6. Other explanatory variables

In addition to the occupational histories it should be natural to include the variables that demographers agree to consider important background factors. Age and type of cohabitation (married or unmarried) have already been mentioned. Furthermore, the number of previous live births ("parity") should be mentioned. This piece of information could in practice be obtained from the CPR system for both men and women, at any rate for children who had not died before 1968.

It could be an alternative to use as a background variable the number of children living together with the parents.

6.7. How to obtain a general view

As has appeared, the problems of the fertility study are even more difficult to handle than those of the mortality study.

The occupational histories are equally multitudinous and if we look at couples we even have two occupational histories. This problem could be tackled by means of a simplified model similar to than in section 5.6. But we cannot disregard parity. It is comforting to note that this problem will be easier to solve when the study period becomes longer, because the birth histories will eventually comprise all children ever born to the women in the study.

Perhaps some rough groups could be separated, at least for an introductory comparison. A question to be answered could thus be:

F: "Do couples made up of managerial employees have a greater tendency not to have any children at all than couples made up of one managerial employee and one outside the labour force?"

But here as in other studies of fertility the answers will be uncertain due to the interaction between effects of generation and calendar year. Will the differences later be adjusted? Have births been postponed or precipitated?

7. Concluding remarks

In spite of the massive problems we have now decided to continue with the studies outlined in this paper. The first project will be the fertility study. This may seem surprising considering that the methodological problems are undoubtedly worse in this study. The reason is that knowledge of fertility behaviour in Denmark is so very sparse, and the demand for new knowledge is correspondingly strong.

Danmarks Statistik hopes to get research funding for the project, which will be very experimental: many different methods will be employed and most of them will probably be rejected.

It is our hope that results will emerge and be published early in 1991.

References

[1] Otto Andersen, Dødelighed og erhverv 1970–80. Danmarks Statistik, Copenhagen (1985).
[2] Otto Andersen, Occupational Mortality in the Nordic Countries 1971–1980. Statistical Reports of the Nordic Countries, Copenhagen (1988).
[3] N.E. Breslow and N.E. Day, Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. Journal of Chronic Diseases 28 (1975).
[4] Danmarks Statistik. Folke- og boligtællingen 1. januar 1981. Ll Landstabelværk. København (1984).
[5] Poul Jensen and Lars Thygesen, Linkage of records of different kinds. Methodological problems and practical experience. Bulletin of the ISI, Vol. LI. (Amsterdam, 1985).
[6] Richard Kay, Cox's Regression Model, in: Encyclopedia of Statistical Sciences (John Wiley, 1982) Vol. 2.
[7] L. Kristoffersen, Yrke og dødelighet. Statistisk Sentralbyrå, Oslo (1979).
[8] Elsebeth Lynge. Dødelighed og erhverv 1970–75. Danmarks Statistik, Copenhagen (1979).
[9] Elsebeth Lynge and Lars Thygesen, Use of surveillance systems for occupational cancer: data from the Danish National System. International Journal of Epidemiology (1988).
[10] Philip Redfern, Which countries will follow the Scandinavian lead in taking a register-based census of population? Journal of Official Statistics, Stockholm (1986).
[11] Philip Redfern, A study of the future of the census of population: alternative approaches. Eurostat Theme 3, Ser. C. European Communities, Luxembourg (1987).
[12] Marcus Rubin and Harald Westergaard, Ægteskabsstatistik, paa Grundlag af den Sociale Lagdeling efter Folketællinger og Kirkebøger i Danmark (P.C. Philipsen, København, 1980).
[13] H. Sauli, Yrke och dødlighet. Statistikcentralen, Helsinki (1979).
[14] Statistiska Centralbyrån, Dødsfallsregister 1961–70. Stockholm (1981).
[15] Lars Thygesen, Data protection in a national register-based statistical system. Eurostat Theme 9, Ser. C. European Communities, Luxembourg (1986).