Arbeidsnotater

STATISTISK SENTRALBYRÅ

WORKING PAPERS FROM THE CENTRAL BUREAU OF STATISTICS OF NORWAY

IO 69/4 Oslo, 26 April 1969

PROBABILISTIC FERTILITY MODELS OF THE LIFE TABLE TYPE

by

Jan M. Hoem x)

Contents

x) Written in the Study Group for Population Models, The Central Bureau of Statistics of Norway, Oslo.

S u m m a r y

In this paper, a series of fertility models are presented as progressive extensions of the basic ideas behind the life table. Dimensions like age, parity, marital status, birth interval, and marital status duration are introduced in turn, and interrelations between the various models are indicated. The main arguments are developed in connection with a model which includes age- and parity-specificity.

The various net (influenced) and gross (partial) fertility measures are given some consideration. It turns out that retrospective fertility investigations give rise to a third kind of functions, which are denoted purged measures. These are parallel to but conceptually distinct from the gross measures. The introduction of the purged measures seems to throw some light on aspects of the theory which have previously appeared problematic.

The models are formulated in terms of transition probabilities rather than survivorship functions. In a final chapter it is argued that the latter are superfluous and even potentially harmful.

1. *I n t r o d u c t i o n*

1.A. For natural reasons the demographic literature is rich in studies of various aspects of fertility. Many of these investigations explicitly build on some model, which may be mathematically or verbally formulated. In other cases an underlying model can be inferred from the treatment given to the problems considered.

Several kinds of model are in use. One type which seems to be fairly well represented, builds on extensions in various directions of some of the basic ideas behind the life table model. (Ryder (1964, p. 453): "The prototype of statistical analysis in demography is the life-table.") It is the purpose of the present paper to study a probabilistic formulation of a set of fertility models moulded in this tradition.

1.B. The life table model, particularly as extended to cover several causes of decrement, has applications in a wide range of situations of which mortality investigation is only one. (See for instance Depoid, 1937; Grabill, 1945; Wolfbein, 1949; Sverdrup, 1967; Potter, 1967; Hoem, 1969 a. Compare also footnote 5 in Hajnal, 1958.) Two prominent features of this model as applied in demography are

(i) the treatment of individual persons as units operating independently of each other, and

(ii) the fact that only a single unspecified sex is considered.

These two features may be carried over to more complex models developed to represent the more involved types of demographic phenomena. This gives rise to a whole class of models which are characterized by (i) and (ii) and which may be described within the framework of Markov process theory. The fertility models which we shall study in this paper, belong to this class. Other examples may be found in Du Pasquier (1912/13), Chiang (1968, Chapters 4, 5, and 7), and Hoem (1968 a).

1.C. It turns out that we shall be able to throw some light on aspects of fertility theory which have previously been problematic. We quote from Ryder (1965, p. 298):

"A neglected problem concerns not the arithmetical effect so much as the consequences for observed rates[1] of the likely selectivity of mortality and migration *vis à vis* fertility[2]. It is in principle possible to maintain statistical control of those who enter or leave a cohort, if fertility histories can be obtained for such persons, but this has not been done. A similar problem plagues the interpreter of retrospective cohort fertility records, obtained by census questions addressed to women past the menopause, who are a select subset of their cohort by virtue of survival. In summary, the conventional cohort fertility history is a somewhat synthetic construction, and research is required to determine the consequences of implicit selectivity."

The question of selectivity due to mortality and migration leads to the distinction between net and gross fertility, which we will take up. Retrospective fertility investigations give rise to a third set of fertility functions, parallel to but distinct from the gross fertility measures. We shall give some consideration to this third set of functions as well.

Since Ryder refers to "observed rates" and to "statistical control", i.e. presumably to statistical estimates, we should perhaps state that results quite similar to the ones developed here for the theoretical model, hold also when real data are considered. (We shall not be systematically concerned with statistical estimation theory in this paper, but have studied such questions elsewhere (Hoem, 1968b, 1969a, 1969b).)

---

1) I.e. occurrence/exposure rates.

2) He then refers to Henry (1959).

1.D. For ease of exposition we shall proceed from a particularly simple fertility model through progressively more complex ones. In our first, introductory model (Model I), we will only take age-specificity into account. In our second model, age-parity-specificity will be introduced. Later models will contain age, parity, marital status, birth interval, and marital status duration.

Our main points will be made in connection with the age-parity-specific model (Model II). In many respects this model is too simple to give a "realistic" description of fertility. However, its very simplicity permits us to concentrate on the argument which we want to bring out without over-burdening the discussion with the details of a more complex model. As we shall sometimes indicate, the argument has general application and is independent of the disguise in which it appears. We have given a formulation of the same line of reasoning in general terms elsewhere (Hoem [12]).

We shall sometimes use the same symbol for quantities which have the same verbal interpretation in different models, but which otherwise have different properties. Thus, e.g., R_B will designate the gross reproduction rate in all models, but its properties differ somewhat from one model to another. Some comparisons of such pairs of parallel concepts will be made in *4.G* and *8.E.*

1.E. Before introducing the models which we intend to study, let us close this introductory chapter by mentioning two types of designs which we shall not consider:

(i) Fertility naturally involves parents of both sexes, and it would be nice if our models could reflect this fact. To achieve this, it would presumably be necessary to apply a "collective" treatment of the reproductive population in the sense that all persons of the population were considered simultaneously as (possibly) interacting units. To this end we would probably use models cut from the same pattern as birth-and-death processes, or more generally the theory of branching processes. This type of model will not be studied in the present paper.

(ii) We shall also leave out detailed micro-models for patterns of human reproduction, such as those studied by Sheps (1965) and her references.

2. *Some conventions*

2.A. Unless something else is explicitly stated, our presentation will be in terms of an actual cohort of parents in a closed population. It is not at all necessary to restrict oneself to this case however. For one thing the cohorts need not be closed (see *10.D*), and for another the same theoretical framework can be used in period analysis.

2.B. By a *birth* we shall mean a confinement producing at least one live child, irrespective of its sex. Thus a multiple birth will be counted as one birth in the same way as a single birth is. We shall not give any special consideration to multiple births.

2.C. To fix our ideas, we shall take the *parity* of an x year old person to be the number of births at ages up to and including age x. We shall use the symbol $N(x)$ for the parity at age x. An x year old person with parity m will be called an (x,m)-person.

2.D. The highest possible live age will be designated ω as usual. The reproductive age interval will be designated $\langle\omega_1,\omega_2\rangle$.

2.E. We will designate the probability that an (x,m)-person will have parity n *and be alive* at age y by $P_{mn}(x,y)$, for $n\geq m$, $y\geq x$. Similarly the probability that an (x,m)-person will have n-m further births and then die within age y will be designated $Q_{mn}(x,y)$.

For each member of the population, let us define a *survival variable* $D(x)$ by

$$D(x) = \begin{cases} 0 & \text{if the person is alive at age } x, \\ 1 & \text{otherwise.} \end{cases}$$

Then obviously

$$P_{mn}(x,y) = P\{N(y) = n, D(y) = 0 \mid N(x) = m, D(x) = 0\},$$

$$Q_{mn}(x,y) = P\{N(y) = n, D(y) = 1 \mid N(x) = m, D(x) = 0\},$$

when $N(y)$ is taken as the parity at death if $D(y)=1$. We also introduce

$$\pi_{mn}(x,y) = P_{mn}(x,y)+Q_{mn}(x,y) = P\{N(y) = n \mid N(x) = m, D(x) = 0\}$$

and $$P_m(x,y) = \sum_{n\geq m} P_{mn}(x,y) = P\{D(y) = 0 \mid N(x) = m, D(x) = 0\}.$$

Thus $\pi_{mn}(x,y)$ is the probability that an (x,m)-person will have n-m further births within a period of y-x years, and $P_m(x,y)$ is the probability that an (x,m)-person will survive to age y.

Since obviously $N(0) = 0$, $D(0) = 0$, we get $P_{0m}(0,x) = P\{N(x)=m, D(x)=0\}$, $Q_{0m}(0,x) = P\{N(x) = m, D(x) = 1\}$, $P_0(0,x) = P\{D(x) = 0\}$, and $\pi_{0m}(0,x) = P\{N(x) = m\}$.

3. *Model I: Age-specificity only*

3.A. Let $\mu(x)$ be the force of mortality and $\phi(x)$ the force of fertility for an x year old person, irrespective of parity, marital status, time elapsed since last birth or possible marriage, etc. This will be taken to mean the following:

We observe a person alive at age x during the age interval $\langle x, x+\Delta x\rangle$ with $\Delta x>0$. Then

(i) the probability that the person will die in the age interval without giving birth to any children, equals $\mu(x)\Delta x + o(\Delta x)$, where $o(\Delta x)/\Delta x \longrightarrow 0$ as $\Delta x \longrightarrow 0$,

(ii) the probability that the person will have exactly one birth in the age interval and survive to age $x+\Delta x$, equals $\phi(x)\Delta x + o(\Delta x)$,

(iii) the probability that the person will survive to age $x+\Delta x$ without giving birth to any children in the age interval, equals $1 - [\mu(x) + \phi(x)]\Delta x + o(\Delta x)$, and

(iv) the probability that the person has more than one birth, or has at least one birth and dies within the age interval, is $o(\Delta x)$.

We shall assume that $\mu(x)$ and $\phi(x)$ are continuous functions for $x\geq 0$ with $\mu(x)>0$ for $0\leq x<\omega$, $\mu(x) \rightarrow \infty$ as $x\uparrow\omega$, $\phi(x)>0$ for $\omega_1<x<\omega_2$, $\phi(x) = 0$ otherwise. Thus the fertile period is the age interval where $\phi(x)>0$.

3.B. The force of fertility is of course a theoretical counterpart to the common age-specific empirical fertility rate (with a number of *births* in the numerator). Similarly $\mu(x)$ is a theoretical counterpart to the age-specific empirical mortality rate. In fact the occurrence/exposure rates may be regarded as statistical estimates for the corresponding forces (Hoem, 1969b, (4.3)).

3.C. One may show (Hoem, 1969b) that in the present situation the following formulae are valid for all m:

$$P_{m,m+k}(x,y) = \frac{1}{k!}\left\{\int_x^y \phi(\xi)d\xi\right\}^k \exp\left\{-\int_x^y [\phi(\xi) + \mu(\xi)]d\xi\right\},$$

$$Q_{mn}(x,y) = \int_x^y P_{mn}(x,\xi)\mu(\xi)d\xi, \text{ and}$$

$$P_m(x,x+t) = {}_tp_x = \exp\left\{-\int_0^t \mu(x+\tau)d\tau\right\}.$$

If we define the net reproduction rate R_N to be the mean number of births that will ever be born to a new-born person (counting offspring of both sexes) then

$$(3.1) \qquad R_N = \int_{\omega_1}^{\omega_2} {}_xp_0\, \phi(x)dx.$$

If the gross reproduction rate R_B is the corresponding mean number of births provided there is no mortality before age ω_2, then

$$(3.2) \qquad R_B = \int_{\omega_1}^{\omega_2} \phi(x)dx.$$

3.D. By definition, we have

$$\phi(x) = \lim_{y \downarrow x} P_{m,m+1}(x,y)/(y-x),$$

$$\mu(x) = \lim_{y \downarrow x} Q_{mm}(x,y)/(y-x), \text{ and}$$

$$\lim_{y \downarrow x} P_{mn}(x,y)/(y-x) = \lim_{y \downarrow x} Q_{mn}(x,y) = 0$$

for all other values of m and n with n>m. It is easily verified that the formulae of *3.C* satisfy these relations.

3.E. One of the disadvantages of this kind of model is that it does not specify any largest possible parity. In fact we see from the formulae of *3.C* that the probability that an (x,m)-person will ever experience another birth is independent of m, which is obviously quite unrealistic. In the next model, this defect has been removed.

4. *Model II: Age-parity-specificity. Basic concepts*

4.A. To get a model with age-parity-specificity, we introduce generally

$$\phi_m(x) = \lim_{y \downarrow x} P_{m,m+1}(x,y)/(y-x),$$

$$\mu_m(x) = \lim_{y \downarrow x} Q_{mm}(x,y)/(y-x), \quad \text{and}$$

$$\gamma_m(x) = \lim_{y \downarrow x} Q_{m,m+1}(x,y)/(y-x),$$

without of course using the special formulae for $P_{mn}(x,y)$ and $Q_{mn}(x,y)$ of *3.C.* (Note the analogy with the formulae of *3.D.*)

Then $\phi_m(x)\Delta x + o(\Delta x)$ is the probability that an (x,m)-person will have a birth within age $x+\Delta x$ and will survive to age $x+\Delta x$. We will call $\phi_m(x)$ *the force of fertility* for an (x,m)-person. Of course $\phi_m(x) = 0$ for $x \notin \langle\omega_1,\omega_2\rangle$. Presumably one will also have $\phi_m(x) = 0$ for high values of m when x only just exceeds ω_1. We assume that there is some largest possible parity M. Then $\phi_M(x)$ will be identically equal to zero.

Similarly $\mu_m(x)\Delta x + o(\Delta x)$ is the probability that an (x,m)-person will die within age $x+\Delta x$ without having another birth. We will call $\mu_m(x)$ *the force of mortality* for an (x,m)-person.

Finally $\gamma_m(x)\Delta x + o(\Delta x)$ is the probability that an (x,m)-person will both have a birth and die within age $x+\Delta x$. Let us call $\gamma_m(x)$ *the force of childbirth mortality* for an (x,m)-person.

We will take *the forces of transition* $\phi_m(x)$, $\mu_m(x)$, and $\gamma_m(x)$ to be continuous functions of x.

For completeness we also state that

$$\lim_{y \downarrow x} P_{mn}(x,y)/(y-x) = \lim_{y \downarrow x} Q_{mn}(x,y)/(y-x) = 0$$

for all m and n with n>m, except in the cases at the beginning of this paragraph.

4.B. Formulae like the following are easily established and have obvious interpretations:

$$(4.1)\qquad P_{mm}(x,y) = \exp\{-\int_x^y [\mu_m(\xi) + \phi_m(\xi) + \gamma_m(\xi)]\, d\xi\}.$$

$$(4.2)\qquad P_{mn}(x,z) = \sum_{k=m}^{n} P_{mk}(x,y)\, P_{kn}(y,z)$$

$$(4.3)\qquad = \int_x^z P_{mm}(x,\xi)\, \phi_m(\xi)\, P_{m+1,n}(\xi,z)\, d\xi$$

$$= \int_x^z P_{m,n-1}(x,\xi)\, \phi_{n-1}(\xi)\, P_{nn}(\xi,z)\, d\xi$$

for $m<n$, $x\leqq y\leqq z$.

$$Q_{mn}(x,y) = \int_x^y P_{mn}(x,\xi)\, \mu_n(\xi)\, d\xi + (1-\delta_{mn}) \int_x^y P_{m,n-1}(x,\xi)\, \gamma_{n-1}(\xi)\, d\xi$$

for $m\leqq n$, $x\leqq y$, where δ_{mn} is a Kronecker delta.

$$(4.4)\qquad P_m(x,y) = P_{mm}(x,y) + \int_x^y P_{mm}(x,\xi)\, \phi_m(\xi)\, P_{m+1}(\xi,y)\, d\xi.$$

$$\pi_{mn}(x,y) = \int_x^y P_{mm}(x,\xi)\phi_m(\xi)\, \pi_{m+1,n}(\xi,y)\, d\xi + \delta_{m+1,n} \int_x^y P_{mm}(x,\xi)\, \gamma_m(\xi)\, d\xi$$

for $m<n$, $x\leqq y$.

4.C. The probability that an (x,m)-person will reach at least parity n within age y, is

$$\alpha_{mn}(x,y) = P\{N(y) \geqq n | N(x) = m, D(x) = 0\}$$

$$(4.5)\qquad = \sum_{k\geqq n} \pi_{mk}(x,y) = \int_x^y P_{m,n-1}(x,\xi)\, [\phi_{n-1}(\xi) + \gamma_{n-1}(\xi)]\, d\xi,$$

for $x\leqq y$, $m<n$; while $\alpha_{mm}(x,y) \equiv 1$. In particular,

$$(4.6)\qquad \alpha_{m,m+1}(x,y) = 1-\pi_{mm}(x,y).$$

For $n>m$, $\alpha_{mn}(x,y)$ also represents the expected number of n-th order births to be experienced within age y by an (x,m)-person. (Compare Karmel (1950), formula (1).)

4.D. $\phi_m(x)$ and $\gamma_m(x)$ are measures of what we might call the "instantaneous fertility" of an (x,m)-person. The $\alpha_{mn}(x,y)$ are measures of the fertility of such a person during the age interval $<x,y]$. Another such measure would be the expected number of further births within age y. This is

$$R_m(x,y) = E\{N(y)-N(x)|N(x) = m,D(x) = 0\} = \sum_{k\geq 0} k\pi_{m,m+k}(x,y)$$

$$= \int_x^y P_{mm}(x,\xi)\ \phi_m(\xi)\ [1+R_{m+1}(\xi,y)]\,d\xi + \int_x^y P_{mm}(x,\xi)\ \gamma_m(\xi)\,d\xi$$

$$(4.7) \qquad = \sum_{n>m} \alpha_{mn}(x,y),$$

where both of the two latter formulae have obvious further interpretations.

In particular our net reproduction rate R_N, according to the verbal definition given in *3.C*, equals

$$(4.8) \qquad R_N = R_0(0,\omega_2) = \sum_m \int_{\omega_1}^{\omega_2} P_{0m}(0,x)[\phi_m(x) + \gamma_m(x)]\,dx.$$

4.E. The preceding paragraphs contain four measures of the short-time fertility of an (x,m)-person. These are

$\phi_m(x) + \gamma_m(x)$, which we could call the "total" force of fertility,

$\pi_{m,m+1}(x,x+1)$, which is the probability of having exactly one birth within age x+1,

$\alpha_{m,m+1}(x,x+1)$, which is the probability of having a birth (i.e. at least one birth) within age x+1, and

$R_m(x,x+1)$, which is the expected number of births within age x+1.

In theory, all of these are different, and they should be kept conceptually apart. In practice, however, the numerical difference between the four measures may well be negligible. By way of the mean value theorem, previous formulae give

$$\pi_{m,m+1}(x,x+1) = P_{mm}(x,\xi_x)\ \phi_m(\xi_x)\ \pi_{m+1,m+1}(\xi_x,x+1) + P_{mm}(x,\xi_x^*)\ \gamma_m(\xi_x^*),$$

$$\alpha_{m,m+1}(x,x+1) = P_{mm}(x,\xi_x')[\phi_m(\xi_x') + \gamma_m(\xi_x')]\ , \text{ and}$$

$$R_m(x,x+1) = \alpha_{m,m+1}(x,x+1) + \text{the probability of having at least two births within age } x+1,$$

where ξ_x, ξ_x^*, and ξ_x' are values in $\langle x,x+1\rangle$. Presumably the numerical difference between each of these measures and $\phi_m(x) + \gamma_m(x)$ will generally be swamped by the other causes of error involved in estimating the quantities concerned.

4.F. Ryder "has found it useful to employ a measure he has termed the parity progression ratio. The progression ratio for a particular parity is the proportion of women who, having arrived in that parity, move on to the next." (Ryder, 1958, p. 43.)

A theoretical counterpart to *the parity m progression ratio* would be the probability

$$\beta_m = P\{N(\omega_2) \geqq m+1 \mid N(\omega_2) \geqq m\}$$

for $m \geqq 0$, which we could call *the parity m progression probability*. We get

$$(4.9) \quad \beta_0 = P\{N(\omega_2) \geqq 1\} = \alpha_{01}(0,\omega_2) = 1 - \pi_{00}(0,\omega_2),$$

while for $m \geqq 1$,

$$(4.10) \quad \beta_m = \frac{P\{N(\omega_2) \geqq m+1\}}{P\{N(\omega_2) \geqq m\}} = \frac{\alpha_{0,m+1}(0,\omega_2)}{\alpha_{0m}(0,\omega_2)} = 1 - \frac{\pi_{0m}(0,\omega_2)}{\alpha_{0m}(0,\omega_2)}.$$

4.G. $R_m(x,y)$ depends on the parity m. We may also define parity-independent fertility measures, even within the present model. Let $R(x,y)$ denote the expected number of further births within age y to an x year old person. Then

$$R(x,y) = E\{N(y)-N(x) \mid D(x) = 0\} = \sum_m E\{N(y)-N(x) \mid N(x) = m, D(x) = 0\}\, P\{N(x) = m \mid D(x) = 0\},$$

which gives the formula

$$(4.11) \quad R(x,y) = \sum_m R_m(x,y) \frac{P_{0m}(0,x)}{P_0(0,x)}.$$

Thus $R(x,y)$ is a weighted average of the $R_m(x,y)$.

The probability that an x year old person will have a birth within $x+\Delta x$ and survive to age $x+\Delta x$, equals $\phi(x)\Delta x + o(\Delta x)$, where

$$(4.12) \qquad \phi(x) = \sum_m \phi_m(x) \frac{P_{0m}(0,x)}{P_0(0,x)} .$$

Note the difference between the use of the symbol $\phi(x)$ in Models I and II. In Model I, $\phi(x)$ was defined verbally in *3.A*, and the relation given in *3.D* is really its formal definition. In Model II, $\phi(x)$ is defined in (4.12), and this quantity has the same verbal interpretation as in the previous model. While the $\phi(x)$ of Model I is independent of mortality *by definition*, the $\phi(x)$ of (4.12) obviously depends on mortality via the survival probabilities.

With a similar interpretation in terms of child-birth mortality, we introduce

$$\gamma(x) = \sum_m \gamma_m(x) \frac{P_{0m}(0,x)}{P_0(0,x)}$$

Some simple manipulations with (4.2), (4.5), (4.7), and (4.12) then gives the formula

$$R(x,y) = \int_x^y P_0(0,\xi)[\phi(\xi) + \gamma(\xi)]d\xi / P_0(0,x).$$

Furthermore, (4.8) reduces to

$$R_N = \int_{\omega_1}^{\omega_2} P_0(0,x)[\phi(x) + \gamma(x)]dx.$$

(Compare this with (3.1).)

5. *Model II: Measures of fertility in the absence of mortality*

5.A. We see from the formulae of chapter 4 that the values of each of the functions $P_{mn}(x,y)$, $R_m(x,y)$, R_N, $R(x,y)$, $\alpha_{mn}(x,y)$, β_m, and even $\phi(x)$ are influenced by the mortality conditions, as they depend on the $\mu_m(x)$ and $\gamma_m(x)$. (This is the case for the $Q_{mn}(x,y)$, $\pi_{mn}(x,y)$, $P_m(x,y)$, and $\gamma(x)$ as well, of course, but that is without interest here.) Since they are all measures of fertility, such an influence is regarded as a nuisance (Henry, 1959; Hoem, 1969b, §3.2). It is customary to construct corresponding "pure" measures which are free from dependence of mortality by substituting zero for all forces of mortality in the formulae of the influenced measures. In the present context this corresponds to setting $\mu_m(x)$ and $\gamma_m(x)$ identically equal to zero. Following the terminology of Du Pasquier (1912/13) and Sverdrup (1967), we shall call the functions which then appear *partial* measures, and shall designate each of them by a bar over the symbol for the corresponding *influenced* measure. Thus, for example,

$$(5.1) \qquad \overline{P}_{mm}(x,y) = \exp\{-\int_x^y \phi_m(\xi)d\xi\},$$

$$\overline{R}_m(x,y) = \sum_{n\geq m} \int_x^y \overline{P}_{mn}(x,\xi)\, \phi_n(\xi)d\xi,$$

$$\overline{R}(x,y) = \int_x^y \overline{\phi}(\xi)d\xi, \quad \text{with}$$

$$(5.2) \qquad \overline{\phi}(x) = \sum_m \overline{P}_{0m}(0,x)\, \phi_m(x).$$

As we have already noted, the quantity $\phi(x)$ is influenced by mortality in the present model. The removal of mortality therefore makes (5.2) necessary.

5.B. By (4.5), (4.6), and (5.1)

$$\overline{\alpha}_{mn}(x,y) = \int_x^y \overline{P}_{m,n-1}(x,\xi)\, \phi_{n-1}(\xi)d\xi \quad \text{for } n>m, \text{ and}$$

$$(5.3) \qquad \overline{\alpha}_{m,m+1}(x,y) = 1 - \exp\{-\int_x^y \phi_m(\xi)d\xi\}.$$

(Compare Karmel (1950), formula (4).) It is apparently sometimes erroneously believed that in the absence of mortality, the expected number of first births to a woman equals

$$(5.4) \qquad \int_{\omega_1}^{\omega_2} \phi_0(x)dx,$$

instead of $\overline{\beta}_0 = \overline{\alpha}_{01}(0,\omega_2) = 1 - \exp\{-\int_{\omega_1}^{\omega_2} \phi_0(x)dx\}$, which is the correct answer.

This belief presumably rests on a false analogy with (3.2). Whelpton, who has explained the nature of the fallacy (1946, p. 505), calls calculation methods based on (5.4) the "conventional procedure" (1949, p. 735). The numerical error resulting from the use of (5.4) can be considerable and can lead to absurd results, such as estimates of the expected total number of first births to a woman exceeding 1 (Whelpton, 1954, pp. 9-10). This kind of "result" has also been found in other connections, e.g. for Norwegian first marriages (Vogt, 1952, 1964).

5.C. We recognize the influenced functions as measures of net fertility, as is reflected in the name of the net reproduction rate. Similarly the partial functions are measures of gross fertility, and the gross reproduction rate, as defined verbally in *3.C*, (or more commonly in the case where offspring of both sexes are counted: the total fertility rate) is

$$R_B = \sum_m \int_{\omega_1}^{\omega_2} \overline{P}_{0m}(0,x)\, \phi_m(x)\, dx. \tag{5.5}$$

We have chosen to use the designations "influenced" and "partial" beside the more common "net" and "gross", in order to emphasize the correspondence with similar concepts in other applications (Hoem [12]). Many other names are in use for the same thing.

The origin of the idea of introducing partial measures can be traced through Makeham (1874) and Du Pasquier (1912/13).

5.D. In sections *5.A* to *5.C*, we have substituted zero for all $\gamma_m(x)$ as well as for all $\mu_m(x)$, and the partial measures resulted. Another possibility is to substitute zero only for the $\mu_m(x)$, while all $\gamma_m(x)$ are retained. This would give us a model where childbirth mortality is permitted while all other causes of death are inoperative. Such a procedure would give rise to a set of fertility and mortality measures intermediate between those which we have called influenced measures and those for which we have used the term partial measures.

6. *Model II: Retrospective fertility investigations*

6.A. In studies of the past fertility of (say) women living in a given area (e.g. a country or part of a country), data may be collected by asking the women living in the area at a certain date to give an account of their individual fertility histories. In such a case, data will be missing for the women who have died or migrated from the area before the date of observation. In Ryder's words (see the quotation in *1.C*), the women included in the investigation therefore are "a select subset of their cohort by virtue of survival". Even when we disregard all problems concerning the reliability of the information collected, there is thus a systematic difference between data obtained by retrospective fertility questions in a survey or census on the one hand, and data collected by a continuous population register, say, on the other hand (Grabill, Kiser, and Whelpton, 1958, pp. 316, 425-426). We shall now study some particulars of such retrospective fertility investigations.

6.B. Let the members of a particular cohort for which retrospective fertility information is collected, be z years old on the date of observation. Even before any data are at hand, the investigator will know that $D(z) = 0$ for each member investigated from this cohort. All probability statements must take this into account. Thus for instance the probability that one of the persons studied who had parity m at age x, will have had parity n at a subsequent age y $(m \leq n, \quad x \leq y \leq z)$, equals

$$P\{N(y) = n \mid N(x) = m, D(z) = 0\}.$$

Similarly the expected number of further births within age y to such a person equals

$$E\{N(y) - N(x) \mid N(x) = m, D(z) = 0\}.$$

In other words: In the present situation all statistical functions appear as measures conditional upon the event "$D(z) = 0$".

6.C. We introduce some additional notation: For $m \leq n$, $x \leq y \leq z$, let

$$P^{\star}_{mn}(x,y,z) = P\{N(y) = n \mid N(x) = m, D(z) = 0\}$$

$$= \frac{P\{N(y) = n, D(z) = 0 \mid N(x) = m, D(x) = 0\}}{P\{D(z) = 0 \mid N(x) = m, D(x) = 0\}}$$

$$= \frac{P_{mn}(x,y)\ P_n(y,z)}{P_m(x,z)}. \tag{6.1}$$

Then $P^{*}_{mn}(x,y,z)$ is the quantity which in the present situation corresponds to the $P_{mn}(x,y)$ of chapter 4 and the $\overline{P}_{mn}(x,y)$ of chapter 5.

6.D. Corresponding to $\phi_m(x)$ we have

$$(6.2) \qquad \phi^{*}_m(x,z) = \lim_{y \downarrow x} P^{*}_{m,m+1}(x,y,z)/(y-x) = \phi_m(x)\frac{P_{m+1}(x,z)}{P_m(x,z)} .$$

Here $\phi_m(x)$ is the force of fertility for an (x,m)-person as it would appear in a continuous registration system, and $\phi^{*}_m(x,z)$ is the corresponding force in a retrospective fertility investigation.

$P_m(x,z)$ and $P_{m+1}(x,z)$ are survival probabilities. It is intuitively obvious that these two functions will be equal if $\mu_{m+k}(y)$ is independent of k for $k \geq 0$, $x \leq y \leq z$, i.e. if mortality is independent of parity exceeding m in the age interval $[x,z]$, and it is possible to prove this mathematically (Hoem [12], § D). Thus the two sets of forces of fertility (for all m and all $x \leq z$) are identical if mortality is independent of parity. In real life, however, mortality is probably apt to increase with parity, so that

$$(6.3) \qquad P_{m+1}(x,z) < P_m(x,z).$$

One must expect, therefore, that $\phi^{*}_m(x,z)$ *is usually smaller than* $\phi_m(x)$.

Grabill, Kiser, and Whelpton (1958, pp. 425-427) would seem to take the $\phi^{*}_m(x,z)$ as the basic functions and the $\phi_m(x)$ as a kind of substitutes. They write, e.g.: "Ideally, a cumulative birth rate for the women in a cohort who live to a specified age would be computed by adding the annual rates for *these women* during younger ages, rather than the rates for all the women living at each younger age." (P. 425. Their italics.) For model-building purposes it has been natural to reverse the situation and take the $\phi_m(x)$ as the basic functions and the $\phi^{*}_m(x,z)$ as derived ones.

Comparing fertility rates calculated from cohort data with corresponding ones calculated from census data, Grabill, Kiser, and Whelpton (1958, p. 431) find that "a substantial majority of the cumulative birth rates from the cohort tables are slightly larger than those from the census data." The numerical effect is very small (p. 316).

The authors explain such differences as there are in terms of registration errors and different treatment of the data from the two kinds of sources. Can it be possible that the *systematic* differences that should be present, can have contributed to the *observed* differences, which are in the "correct" direction? Or is the systematic difference between the $\phi_m(x)$ and the $\phi^{*}_m(x,z)$ so small that it is completely swamped by other effects?

6.E. We see from (6.1) that $P^*_{mn}(x,y,z)$ is generally different from $P_{mn}(x,y)$. In particular,

$$(6.4) \qquad P^*_{mn}(x,y,y) = P_{mn}(x,y) \,/\, P_m(x,y) > P_{mn}(x,y)$$

and

$$(6.5) \qquad P^*_{mm}(x,y,z) = P_{mm}(x,y) \frac{P_m(y,z)}{P_m(x,z)} > P_{mm}(x,y).$$

It is important to note that *$P^*_{mn}(x,y,z)$ also generally differs from $\overline{P}_{mn}(x,y)$.* This is so because the $\overline{P}_{mn}(x,y)$ can be uniquely constructed from the forces $\phi_m(x)$, and the $P^*_{mn}(x,y,z)$ can be uniquely constructed in exactly the same way from the different forces $\phi^*_m(x,z)$. In particular (5.1) holds for $\overline{P}_{mm}(x,y)$, while

$$(6.6) \qquad P^*_{mm}(x,y,z) = \exp\{-\int_x^y \phi^*_m(\xi,z)\, d\xi\}.$$

(A proof of (6.6) is given in the appendix.)

We also see that $\phi^*_m(x,z)$ and $P^*_{mn}(x,y,z)$ generally depend on mortality.

6.F. Corresponding to the $R_m(x,y)$ of *4.D* we introduce (for $x \leq y \leq z$)

$$R^*_m(x,y,z) = E\{N(y) - N(x) \,|\, N(x) = m, D(z) = 0\} = \sum_{k \geq 0} k\, P^*_{m,m+k}(x,y,z)$$

$$= \sum_{n \geq m} \int_x^y P^*_{mn}(x,\xi,z)\, \phi^*_n(\xi,z)\, d\xi$$

$$= \sum_{n \geq m} \int_x^y P_{mn}(x,\xi)\, \phi_n(\xi) \frac{P_{n+1}(\xi,z)}{P_m(x,z)}\, d\xi.$$

Even in the absence of childbirth mortality (i.e. even when $\gamma_m(x) \equiv 0$), $R^*_m(x,y,z)$ may therefore be different from $R_m(x,y)$. (Cfr. (4.5) and (4.7).)

In particular

$$R^* = R^*_0(0,\omega_2,\omega_2) = \sum_{k \geq 0} k\, P_{0k}(0,\omega_2) \,/\, P_0(0,\omega_2)$$

$$(6.7) \qquad = \sum_{m \geq 0} \int_{\omega_1}^{\omega_2} P_{0m}(0,x)\, \phi_m(x) \frac{P_{m+1}(x,\omega_2)}{P_0(0,\omega_2)}\, dx.$$

This is the reproduction rate (i.e. the mean number of births ever experienced) calculated for a person who has actually survived the reproductive period. Alternatively it may be interpreted as the mean number of births ever to be experienced by a new-born person who will survive to age ω_2. *This quantity should not be confused with the gross reproduction rate.* Since $\overline{P}_{0m}(0,x)$ will in general be different from $P_{0m}(0,x)\ P_{m+1}(x,\omega_2)\ /\ P_0(0,\omega_2)$, a comparison of (6.7) and (5.5) shows that R^* and R_B are generally different.

Nor should R^* be confused with Woofter's completed generation reproduction rate (1947), which is simply our R_N calculated on a cohort basis.

It may be useful to have a name for R^*, and we shall call it *the purged reproduction rate.* Similarly $\phi^*_m(x,z)$ and $P^*_{mn}(x,y,z)$, as well as the quantities $\alpha^*_{mn}(x,y,z)$ and β^*_m defined below, will be called *purged* measures. The idea behind this terminology is that data of the kind collected in a retrospective fertility investigation have the same properties as the data that would result from purging a continuous population register of all information concerning members deceased or emigrated (i.e. removing such information).

6.G. As a counterpart to the $\alpha_{mn}(x,y)$ of *4.C* we introduce

$$\alpha^*_{mn}(x,y,z) = P\{N(y) \geq n | N(x) = m, D(z) = 0\}.$$

We get for $n>m$

$$\alpha^*_{mn}(x,y,z) = \begin{cases} \sum_{k=m}^{n-1} P_{mk}(x,z)\ \alpha_{kn}(z,y)\ /\ P_m(x,z) & \text{for } x \leq z < y, \\ \\ \sum_{k \geq n} P_{mk}(x,y)\ P_k(y,z)\ /\ P_m(x,z) & \text{for } x \leq y \leq z. \end{cases}$$

In particular, $\alpha^*_{m,m+1}(x,y,y) = 1 - P_{mm}(x,y)\ /\ P_m(x,y)$.

The purged parity m progression probability will be

$$\beta^*_m = P\{N(\omega_2) \geq m+1 | N(\omega_2) \geq m, D(\omega_2) = 0\}$$

$$\text{(6.8)} \qquad = \frac{P\{N(\omega_2) \geq m+1 | D(\omega_2) = 0\}}{P\{N(\omega_2) \geq m | D(\omega_2) = 0\}} = \frac{\alpha^*_{0,m+1}(0,\omega_2,\omega_2)}{\alpha^*_{0m}(0,\omega_2,\omega_2)}\ ,$$

with $\beta^*_0 = \alpha^*_{01}(0,\omega_2,\omega_2)$ in particular.

6.H. As examples of the use of the purged measures, we refer to Welpton (1954, Table A) and Ryder (1958). In Welpton's Table A, the column headed "total births" contains estimates of

$$R_0^*(0,x,x) = E\{N(x)|D(x) = 0\},$$

and the subsequent columns contain estimates of

$$\alpha_{0m}^*(0,x,x) = P\{N(x) \geqq m|D(x) = 0\}$$

for m=1,2,etc. The corresponding parity distribution can be calculated from the formulae

$$P\{N(x) = 0|D(x) = 0\} = 1 - \alpha_{01}^*(0,x,x),$$

$$P\{N(x) = m|D(x) = 0\} = \alpha_{0m}^*(0,x,x) - \alpha_{0,m+1}^*(0,x,x) \text{ for } m \geqq 1.$$

Ryder (1958) uses (6.8) to calculate time series of estimates for the β_m^*.

6.I. By way of summary, then, we repeat that there is a definite conceptual difference between the purged and the partial fertility measures. Thus in particular the purged measures account for survival to a given age, say z, but they do *not* generally result from the assumption that there is no mortality, as seems commonly believed. (Compare, e.g., Ryder (1960), footnote 3.)

7. *Model II: Mean age concepts*

7.A. In the life table model, there appears a quantity $\mathring{e}_{x:\overline{n|}}$ representing the expected (or mean) lifetime up to age x+n of an x year old person. In the present chapter we shall generalize this concept to the situation described by Model II.

7.B. For an (x,m)-person the mean lifetime in parity n within age y is

$$e_{mn}(x,y) = \int_x^y P_{mn}(x,\xi)d\xi$$

for $m \leqq n$, $x \leqq y$. In particular $e_{mm}(x,\omega)$ is the expected future time to be spent with parity m. The period during which an (x,m)-person continues to have parity m may end in one out of two ways: either by a birth (including possible death in childbirth) or by the death (other than in childbirth) of the person in question. Thus $e_{mm}(x,\omega)$ is not the expected waiting time up to the next birth. The latter quantity will be introduced in *7.C* below.

7.C. For an (x,m)-person who ever reaches parity n>m, let Y_n be the age at which the person experiences the n-th birth. The corresponding mean age is

$$A_{mn}(x) = E\{Y_n | N(x) = m, D(x) = 0, N(\omega_2) \geq n\}$$

$$= \int_x^{\omega_2} y\, P_{m,n-1}(x,y)\left[\phi_{n-1}(y) + \gamma_{n-1}(y)\right] dy \,/\alpha_{mn}(x,\omega_2). \tag{7.1}$$

In particular, $A_{m,m+1}(x) - x$ represents the expected waiting time up to the next birth for an (x,m)-person who will have another birth.

For an (x,m)-person who ever experiences a further birth, let $Y_{m+1},\ldots,Y_{m+L}$ represent the ages at which the future births arrive, and let $Y = \frac{1}{L}\sum_{\ell=1}^{L} Y_{m+\ell}$ be the observed mean age. The expected mean age at future births of an (x,m)-person who will experience another birth then is

$$A_m(x) = E\{Y | N(x) = m, D(x) = 0,\ N(\omega_2) > m\}$$

$$= \int_x^{\omega_2} y \sum_{n \geq m} P_{mn}(x,y)\left[\phi_n(y) + \gamma_n(y)\right] dy \,/\, R_m(x,\omega_2)$$

$$= \sum_{n>m} A_{mn}(x)\, \alpha_{mn}(x,\omega_2) \,/\, R_m(x,\omega_2). \tag{7.2}$$

7.D. We may motivate the introduction of the $A_{mn}(x)$ and the $A_m(x)$ as follows: Consider a closed group of K (x,m)-persons. Suppose that all births and deaths in the group during the age interval $[x,\omega_2]$ are recorded. Let $B_{kn} = 1$ if person no. k ever reaches a given parity n>m, and let $B_{kn} = 0$ otherwise, for $k = 1,2,\ldots,K$. If $B_{kn} = 1$, let Y_{kn} be the age at which person no. k reaches parity n, and let $Y_{kn} = \omega$ otherwise. Then $B_n = \sum_k B_{kn}$ is the number of persons ever to reach parity n, and the empirical mean age at n-th birth of the group would be calculated as

$$W_{Kn} = \sum_{k=1}^{K} Y_{kn} B_{kn} / B_n,$$

provided $B_n > 0$. If $B_n = 0$, set W_{Kn} arbitrarily equal to ω.

We now introduce

$$F_{mn}(x,y) = P\{Y_{kn} \leq y | N_k(x) = m, D_k(x) = 0, N_k(\omega_2) \geq n\},$$

where $N_k(\cdot)$ and $D_k(\cdot)$ are the values of $N(\cdot)$ and $D(\cdot)$ of chapter 2 which correspond to person no. k. Then

$$f_{mn}(x,y) = \frac{\partial}{\partial y} F_{mn}(x,y) = \frac{P_{m,n-1}(x,y)[\phi_{n-1}(y) + \gamma_{n-1}(y)]}{\alpha_{mn}(x,\omega_2)}$$

for $x \leq y \leq \omega_2$, and $A_{mn}(x)$ is the mean of this distribution. It is possible to prove that as $K \to \infty$, $\frac{1}{K} \sum_{k=1}^{K} Y_{kn} B_{kn}$ converges to the integral in (7.1) with probability 1, and B_n/K converges to $\alpha_{mn}(x)$, similarly with probability 1. Thus, with probability 1, W_{Kn} converges to $A_{mn}(x)$.

The empirical mean age at any future birth of the group of K persons (with K finite) would be calculated as

$$W_K = \sum_{n>m} \sum_{k=1}^{K} Y_{kn} B_{kn} / B,$$

where $B = \sum_{n>m} B_n$ is the total number of future births observed. We see that

$$W_K = \sum_{n>m} W_{Kn} \frac{B_n/K}{B/K} .$$

With probability 1, B/K converges to $\sum_{n>m} \alpha_{mn}(x,\omega_2) = R_m(x,\omega_2)$. Utilizing this and previous results, we see that W_K converges with probability 1 to the quantity in (7.2), which equals $A_m(x)$.

7.E. It is possible to introduce quantities $\bar{e}_{mn}(x,y)$, $e^*_{mn}(x,y,z)$, $\bar{A}_{mn}(x)$, $\bar{A}_m(x)$, $A^*_{mn}(x,z)$, and $A^*_m(x,z)$ in analogy with our account in chapters 5 and 6. It is quite obvious how this is done, however, so we will leave out this part of the theory.

8. *Model III: The introduction of marital status*

8.A. Our next extension is to include marital status. For our purposes it does not matter much exactly how the various marital statuses are defined, so let any finite set of marital states $1,2,\ldots,J$ be given, with marital state 1 corresponding to the never-married status. (For examples, see Hoem (1968a).) For a given person, let $I(x)$ be the marital state at age x. (If $D(x)=1$, let $I(x)$ be the marital state at death.) In replacement of the transition probabilities $P_{mn}(x,y)$ and $Q_{mn}(x,y)$ we introduce

$$P_{mn}^{ij}(x,y) = P\{N(y)=n,\ I(y)=j,\ D(y)=0 \mid N(x)=m,\ I(x)=i,\ D(x)=0\},$$

$$Q_{mn}^{ij}(x,y) = P\{N(y)=n,\ I(y)=j,\ D(y)=1 \mid N(x)=m,\ I(x)=i,\ D(x)=0\}.$$

Similarly let

$$\phi_{im}(x) = \lim_{y\downarrow x} P_{m,m+1}^{ii}(x,y)/(y-x),$$

$$\mu_{im}(x) = \lim_{y\downarrow x} Q_{mm}^{ii}(x,y)/(y-x), \text{ and}$$

$$\gamma_{im}(x) = \lim_{y\downarrow x} Q_{m,m+1}^{ii}(x,y)/(y-x),$$

be continuous functions with interpretation as forces of fertility, mortality, and childbirth mortality, respectively. We also introduce forces of change of marital state (i.e. forces of primary nuptiality, of divorce, of remarriage, etc.) by the definition

$$\lambda_{ijm}(x) = \lim_{y\downarrow x} P_{mm}^{ij}(x,y)/(y-x) \quad \text{for } i\neq j.$$

If it is impossible to move direct from marital state i to marital state j at age x, the corresponding $\lambda_{ijm}(x)$ will be equal to zero. Thus for example $\lambda_{i1m}(x)$ will be identically zero for all $i>1$.

For simplicity we shall assume that it is impossible to change parity and marital status simultaneously, and also to change marital status and die at the same time. Formally this is included in the assumption that

$$\lim_{y\downarrow x} P_{mn}^{ij}(x,y)/(y-x) = \lim_{y\downarrow x} Q_{mn}^{ij}(x,y)/(y-x) = 0$$

for all m, n, i, and j apart from the four cases above and apart from the case i=j, m=n.

We shall also need the quantities

$$\pi_{mn}^{ij}(x,y) = P\{N(y) = n, I(y) = j | N(x) = m, I(x) = i, D(x) = 0\} =$$

$$= P_{mn}^{ij}(x,y) + Q_{mn}^{ij}(x,y), \text{ and}$$

$$P_m^i(x,y) = P\{D(y) = 0 | N(x) = m, I(x) = i, D(x) = 0\} = \sum_{n\geq m} \sum_j P_{mn}^{ij}(x,y).$$

A large part of the theory of Model II can be reproduced within the present model. Since this mostly does not really produce interesting new problems or results, we shall mainly be content to merely indicate some of the possibilities.

8.B. We have

$$P_{mn}^{ij}(x,z) = \sum_\nu \sum_{k=m}^{n} P_{mk}^{i\nu}(x,y)\, P_{kn}^{\nu j}(y,z)$$

for $x<y<z$ and $m\leq n$, while for $m<n$, $x<y$,

$$P_{mn}^{ij}(x,y) = \sum_\nu \int_x^y P_{mm}^{i\nu}(x,\xi)\, \phi_{\nu m}(\xi)\, P_{m+1,n}^{\nu j}(\xi,y)\, d\xi$$

$$= \sum_\nu \int_x^y P_{m,n-1}^{i\nu}(x,\xi)\, \phi_{\nu,n-1}(\xi)\, P_{nn}^{\nu j}(\xi,y)\, d\xi.$$

8.C. Consider a person who at age x is alive and has parity m and marital status i. We will designate by $\alpha_{mn}^{ij}(x,y)$ the probability that such a person will reach parity n *while in marital state j* within age y, and by $R_m^{ij}(x,y)$ the number of births expected to such a person while in marital state j within age y. Then

$$\alpha_{mn}^{ij}(x,y) = \int_x^y P_{m,n-1}^{ij}(x,\xi)\, [\phi_{j,n-1}(\xi) + \gamma_{j,n-1}(\xi)]\, d\xi \quad \text{and}$$

$$R_m^{ij}(x,y) = \sum_{n>m} \alpha_{mn}^{ij}(x,y).$$

(8.1) $$R_m^{i}(x,y) = \sum_j R_m^{ij}(x,y)$$

will be the corresponding expected total number of births, irrespective of marital status at birth. We also have

(8.2) $$R_m^i(x,y) = \sum_j \sum_{k\geq 1} k\, \pi_{m,m+k}^{ij}(x,y).$$

When comparing (8.1) and (8.2) one should not draw the erroneous conclusion that $R_m^{ij}(x,y)$ equals

$$\sum_{k\geq 1} k\, \pi_{m,m+k}^{ij}(x,y).$$

The latter quantity is

$$E\{[N(y)-N(x)]\ \delta_{jI(y)} \mid N(x) = m, I(x) = i, D(x) = 0\},$$

which may be quite different from $R_m^{ij}(x,y)$. ($\delta_{jI(y)}$ is a Kronecker delta, i.e. $\delta_{jI(y)} = 1$ if $I(y) = j$, $\delta_{jI(y)} = 0$ otherwise.)

8.D. Purged measures similar to those of chapter 6 build on transition probabilities of the form

$$P_{mn}^{*ij}(x,y,z) = P\{N(y) = n, I(y) = j \mid N(x) = m, I(x) = i, D(z) = 0\}$$

$$= \frac{P_{mn}^{ij}(x,y)\ P_n^j(y,z)}{P_m^i(x,z)}.$$

Purged forces of transition will appear as

$$\phi_{im}^{*}(x,z) = \lim_{y\downarrow x} P_{m,m+1}^{*ii}(x,y,z)/(y-x)$$

(8.3) $$= \phi_{im}(x)\ \frac{P_{m+1}^i(x,z)}{P_m^i(x,z)}$$

and as

$$\lambda_{ijm}^{*}(x,z) = \lim_{y\downarrow x} P_{mm}^{*ij}(x,y,z)/(y-x) = \lambda_{ijm}(x)\ \frac{P_m^j(x,z)}{P_m^i(x,z)}.$$

Thus there is an extra set of purged forces, just as there is an extra set of "original" forces, because of the extra dimension introduced (viz. marital status). (8.3) shows that we can easily carry over to the present case what was said as a commentary to formula (6.2).

8.E. In *4.A* we defined a quantity $\phi_m(x)$ and gave a verbal interpretation of it. In the present model, a quantity with the same verbal interpretation would be defined by the formula

$$(8.4) \qquad \phi_m(x) = \sum_i \phi_{im}(x) \frac{P_{0m}^{1i}(0,x)}{P_{0m}^{1}(0,x)},$$

where $P_{0m}^{1}(0,x) = \sum_j P_{0m}^{1j}(0,x) = P\{N(x) = m, D(x) = 0\}$

and where marital state 1 still corresponds to the never-married.

While $\phi_m(x)$ of Model II was independent of mortality by definition, $\phi_m(x)$ of the present model generally depends on all forces of transition, including mortality, via the survival probabilities $P_{0m}^{1i}(0,x)$. We recognize the same kind of reasoning as in our commentary to formula (4.12).

When we introduced the partial measures in Model II by replacing all $\mu_m(x)$ and all $\gamma_m(x)$ by zero while keeping all $\phi_m(x)$ constant, the value of $\phi(x)$ had to be replaced by a quantity $\overline{\phi}(x)$, which generally is different. Similarly, if in Model III we replace all $\mu_{im}(x)$ and all $\gamma_{im}(x)$ by zero and keep all $\phi_{im}(x)$ constant, the values $\phi_m(x)$ will generally change unless we make some compensatory changes in the $\lambda_{ijm}(x)$. Thus in Model III, constant values of $\phi_m(x)$ is inconsistent with the very idea of the partial measures, which in the present context appear through the removal of mortality while all other basic forces of transition are kept fixed. We see, therefore, that the properties of quantities like these depend very much on the model in which they appear, even in cases where the verbal interpretation of two quantities may be independent of the model. (The same kind of ideas appear, e.g., in Stolnitz and Ryder (1949).)

In this connection there does not seem to be any model which is "really fundamental" and in which the "real" properties of the verbal concepts are brought out. One could keep adding new dimensions to the fertility model, and this would lead to a hierarchy of progressively complex and "more basic" models. We *have* included the dimensions age, parity, and marital status. We could go on and include social status, income, residence, religious denomination, health, status and age of spouse (for marital fertility), and so on. (Compare Murphy (1966), Chapter VI.) There is no apparent end to this process. In practice, therefore, the investigator must make a deliberate choice of model, and this very choice will in itself influence his findings.

9. *Model IV: The introduction of birth intervals and marital status duration*

9.A. We shall not make the extensions suggested at the end of *8.E*, although most of this could actually have been done with only minor changes in notation. Instead we shall briefly indicate how important variables like marital status duration and birth interval can be included in the analysis.

9.B. For an (x,m)-person in marital status i, let T(x) be the time elapsed since last birth for $m \geq 1$, and let T(x)=x for m=0. Similarly, let U(x)=x if i=1 (marital state 1 still corresponds to the never-married), and let U(x) be the time elapsed since the last change of marital status if i>1. We may then call T(x) *parity duration* at age x, while U(x) similarly is *the duration of the present marital status* at age x, i.e. the duration of I(x). (Other definitions of "marital status duration" may be relevant, for instance in the case of a remarried person, but U(x), as defined above, suffices in bringing out the basic ideas of this kind of concept.)

9.C. The durations T(x) and U(x) are brought into the analysis by the introduction of transition probabilities of the form

$$F_{mn}^{ij}(x,y,s,t,u,v) =$$

$$P\{N(y)=n,\ I(y)=j,\ D(y)=0,\ T(y)\leq t,\ U(y)\leq v \mid N(x)=m,\ I(x)=i,\ D(x)=0,\ T(x)=s,\ U(x)=u\},$$

and

$$G_{mn}^{ij}(x,y,s,t,u,v) =$$

$$P\{N(y)=n,\ I(y)=j,\ D(y)=1,\ T(y)\leq t,\ U(y)\leq v \mid N(x)=m,\ I(x)=i,\ D(x)=0,\ T(x)=s,\ U(x)=u\}.$$

Forces of transition have the form

$$\phi_{im}(x,s,u) = \lim_{y\downarrow x} F_{m,m+1}^{ii}(x,y,s,\infty,u,\infty)/(y-x),$$

$$\mu_{im}(x,s,u) = \lim_{y\downarrow x} G_{mm}^{ii}(x,y,s,\infty,u,\infty)/(y-x),$$

and so on. The whole apparatus of the previous models can be reproduced in the present context, but the formulae will obviously reach a new level of complexity.

9.D. The models of the present paper appear as progressive extensions of each other. Once Model IV has been reached, one can produce other models by removing one or more of the dimensions included there. There are five such dimentions in all in Model IV (viz. age, parity, marital status, parity duration, and marital status duration) so there are many possibilities. The literature abounds with models which can be produced in this way.

An example of reasoned suggestions for this kind of procedure can be found in Ryder (1965, pp.295-296).

10. *Survivorship functions. In- and out-migration*

10.A. In the life table model, ${}_tp_x$ is the probability that an x year old member of the population survive to age x+t. A survivorship function ℓ_x is defined by specifying some radix ℓ_0 and defining

(10.1) $\ell_x = \ell_0 \; {}_xp_0.$

One of the many simple relations which can then be established, is

(10.2) ${}_tp_x = \ell_{x+t} / \ell_x.$

A rigorous proof of this formula goes as follows: If T denotes the total lifetime of a member of the population considered, then

$${}_tp_x = P\{T>x+t \mid T>x\} = \frac{P\{T>x+t\}}{P\{T>x\}} = \frac{{}_{x+t}p_0}{{}_xp_0} ,$$

which by (10.1) implies (10.2).

No doubt the survivorship function is a valuable tool when properly understood, although the one-to-one relationship (10.1) with the survival probabilities makes it quite dispensable. Unfortunately, to use the survivorship function also has its dangers, as we shall point out in the following paragraphs.

10.B. The form of formula (10.2) invites an interpretation in terms of a number of favourable cases divided by a number of possible cases. As a mnemotechnical device and an aid to intuition, such an interpretation may have its value, but it is important to be aware that it is not rigorously correct. (The misconception that the interpretation is correct, or even that (10.2) *defines* ${}_tp_x$, is often quite apparent. See, e.g., Menge and Fisher (1965, p.11), Saxer (1955, p.9), Zwinggi (1945, p.22).)

ℓ_x is *not* a number of possible cases, and similarly ℓ_{x+t} is *not* a number of favourable cases. Probabilistically ℓ_x is interpreted as the expected number of survivors to age x of a closed cohort of ℓ_0 new-born persons, and ℓ_x need not even be an integer. Similarly for ℓ_{x+t}.

An additional weakness of a formula like (10.2) is that it may substantiate the frequent novicial misconception that a probability must necessarily occur as a ratio between two "numbers of cases".

10.C. Although this has not proved necessary, we could introduce survivorship functions, e.g. in Model II. This might be done in analogy with (10.1) by choosing some radix ℓ_0 and letting

$$(10.3) \qquad \ell_x^{(m)} = \ell_0 \, P_{0m}(0,x).$$

$\ell_x^{(m)}$ would then be the expected number of survivors with parity m at age x of a closed cohort of ℓ_0 new-born persons. (Whelpton's Table C (1954) contains estimates for numerous values of $\ell_x^{(m)}$.) We would then get

$$P_{00}(x,x+t) = \ell_{x+t}^{(0)} \,/\, \ell_x^{(0)},$$

in complete analogy with (10.2). For all other values of m and n ($m \leq n$), $P_{mn}(x,x+t)$ does *not* equal $\ell_{x+t}^{(n)} / \ell_x^{(m)}$ (not even for n=m), as one might believe at first thought. Utilizing (10.3) we see that, on the contrary,

$$\frac{\ell_{x+t}^{(n)}}{\ell_x^{(m)}} = \frac{P\{N(x+t) = n,\ D(x+t) = 0\}}{P\{N(x) = m,\ D(x) = 0\}} =$$

$$= \frac{\sum_{\nu=0}^{n} P\{N(x) = \nu,\ D(x) = 0,\ N(x+t) = n,\ D(x+t) = 0\}}{P\{N(x) = m,\ D(x) = 0\}}$$

$$> \frac{P\{N(x) = m,\ D(x) = 0,\ N(x+t) = n,\ D(x+t) = 0\}}{P\{N(x) = m,\ D(x) = 0\}}$$

$$= P\{N(x+t) = n,\ D(x+t) = 0 \mid N(x) = m,\ D(x) = 0\},$$

so that

$$P_{mn}(x,x+t) < \ell_{x+t}^{(n)} / \ell_x^{(m)} \quad \text{unless } n=m=0.$$

In order to achieve the flexibility due to the use of the transition functions $P_{mn}(x,y)$, one would have to include a whole family of survivorship functions, and even if this were done, it would not give any new insight.

10.D. (In- and out-migration.) The survivorship function concept is connected with a cohort closed to in-migration, but out-migration may easily be taken care of. What we have called forces of mortality, can be taken to represent out-migration as well as mortality. Alternatively, each force of mortality may be split into two or more forces of decrement (i.e. decrement from the population), one of which may represent out-migration. (Compare Ryder (1964, p.455).)

In-migration, on the other hand, is not easily included into arguments based upon survivorship functions. The present writer has never seen anyone succeed in doing so, and it seems doubtful that it is at all possible.

To this the formulation in terms of transition probabilities offers a striking contrast. If we look through the arguments of chapters 2 to 9 once more, we see that in-migration poses no problem. All arguments are given in terms of individual persons about whom it is known that at age x they are alive and have some parity m (and possibly also some marital status i and given parity duration and marital status duration), and nothing apart from this is said about their behaviour prior to age x. As long as they have the same fertility and mortality etc. as the rest of the population, they may have migrated into and out of the population several times for that matter. (If the migrants differ from the original population in respect of mortality, fertility, etc., their data should be analysed separately.)

To be sure, the models of the present paper do not take in-migration explicitly into account. Due to the concentration on the individual rather than on the population as an aggregate of individuals, this has not been necessary. To incorporate in-migration as an explicit part of the model would require a macroanalytic approach with a "collective" treatment of the individuals within at least one population.

10.E. In summary, then, we have refrained from using survivorship functions in our fertility models because they are superfluous, because they may be harmful since potentially misleading, and because they may hamper the introduction of in-migration.

Appendix

Proof of (6.6): By (6.5),

$$\text{(A.1)}\qquad \frac{\frac{\partial}{\partial y} P^{*}_{mm}(x,y,z)}{P^{*}_{mm}(x,y,z)} = \frac{\frac{\partial}{\partial y} P_{mm}(x,y)}{P_{mm}(x,y)} + \frac{\frac{\partial}{\partial y} P_{m}(y,z)}{P_{m}(y,z)} \ .$$

Differentiation in (4.4) gives

$$\frac{\partial}{\partial x} P_m(x,y) = P_{mm}(x,y)\ [\mu_m(x) + \phi_m(x) + \gamma_m(x)] - P_{m+1}(x,y)\ \phi_m(x)$$

$$+ \int_x^y P_{mm}(x,\xi)\ [\mu_m(x) + \phi_m(x) + \gamma_m(x)]\ \phi_m(\xi)\ P_{m+1}(\xi,y)\, d\xi$$

$$= P_m(x,y)\ [\mu_m(x) + \phi_m(x) + \gamma_m(x)] - P_{m+1}(x,y)\ \phi_m(x).$$

Introducing this in (A.1), we get

$$\frac{\frac{\partial}{\partial y} P^{*}_{mm}(x,y,z)}{P^{*}_{mm}(x,y,z)} = - \phi^{*}_m(y,z).$$

Since $P^{*}_{mm}(x,x,z) = 1$, (6.6) immediately follows. □

R e f e r e n c e s

[1] Chiang, Chin Long (1968). *Introduction to Stochastic Processes in Biostatistics.* John Wiley & Sons, Inc.

[2] Depoid, Pierre (1937). Tables d'extinction des mariages et des couples suivant la durée de l'union. *Congrès Internat. Popul. Paris* 5 51-60.

[3] Du Pasquier, L.G. (1912/13). Mathematische Theorie der Invaliditätsversicherung. *Mitteil. Verein. schweizer. Versicherungsmath.* 7 1-7 and 8 1-153.

[4] Grabill, W.H. (1945). Attrition life tables for the single population. *J. Am. Statist. Ass.* 40 364-375.

[5] Grabill, W.H.; C.V. Kiser; and P.K. Whelpton (1958). *The Fertility of American Women.* John Wiley & Sons, Inc.

[6] Hajnal, John (1958). The study of fertility and reproduction: A survey of thirty years. Pp. 11-37 in *Thirty Years of Research in Human Fertility: Retrospect and Prospect.* Milbank Memorial Fund.

[7] Henry, Louis (1959). D'un problème fondamental de l'analyse démographique. *Population* 14 (1) 9-32.

[8] Hoem, Jan M. (1968a). A probabilistic approach to nuptiality. Pp. 14-30 in Working Paper IO 68/18 from the Central Bureau of Statistics of Norway.

[9] Hoem, Jan M. (1968b). Time-continuous Markov chain estimation techniques in demographic models. Pp. 31-50 in Working Paper IO 68/18 from the Central Bureau of Statistics of Norway.

[10] Hoem, Jan M. (1969a). A probabilistic model for primary marital fertility. To appear in the *Yearbook of Population Research in Finland.*

[11] Hoem, Jan M. (1969b). Fertility rates and reproduction rates in a probabilistic setting. To appear in the *Biométrie-Praximétrie.*

[12] Hoem, Jan M. Purged and partial Markov chains. To appear in the *Skandinavisk Aktuarietidskrift.*

[13] Karmel, P.H. (1950). A note on P.K. Whelpton's calculation of parity adjusted reproduction rates. *J. Am. Statist. Ass.* 45 119-124.

[14] Makeham, W.M. (1874). On an application of the theory of the composition of decremental forces. *J. Inst. Actuaries* 18 317-322.

[15] Menge, W.O. and C.H. Fisher (1965). *The Mathematics of Life Insurance.* Second Edition. Macmillan.

[16] Murphy, E.M. (1966). *A Generalization of Stable Population Techniques.* Ph.D. dissertation, Department of Sociology, University of Chicago.

[17] Potter, Robert G. (1967). The multiple decrement life table as an approach to the measurement of use effectiveness and demographic effectiveness of contraception. Pp. 869-883 in International Union for the Scientific Study of Population: *Contributed Papers. Sidney Conference, Australia, 21 to 25 August 1967.*

[18] Ryder, N.B. (1956). La mesure des variations de la fécondité au cours du temps. *Population* 11 (1) 29-46.

[19] Ryder, N.B. (1958). An appraisal of fertility trends in the United States. Pp. 38-49 in *Thirty Years of Research in Human Fertility: Retrospect and Prospect.* Milbank Memorial Fund.

[20] Ryder, N.B. (1960). The structure and tempo of current fertility. Pp. 117-131 in National Bureau of Economic Research: *Demographic and Economic Change in Developed Countries.* Princeton University Press.

[21] Ryder, N.B. (1964). Notes on the concept of a population. *Am. J. Sociol.* 69 (5) 447-463.

[22] Ryder, N.B. (1965). The measurement of fertility patterns. Pp. 287-306 in Mindel C. Sheps and Jeanne Clare Ridley (Eds.): *Public Health and Population Chenge.* University of Pittsburgh Press.

[23] Saxer, W. (1955). *Versicherungsmathematik. Erster Teil.* Springer-Verlag.

[24] Sheps, Mindel C. (1965). Application of probability models to the study of patterns of human reproduction. Pp. 307-332 in Mindel C. Sheps and Jeanne Clare Ridley (Eds.): *Public Health and Population Change.* University of Pittsburgh Press

[25] Stolnitz, G.J. and N.B. Ryder (1949). Recent discussion of the net reproduction rate. *Population Index* 15 114-128. Reprinted as pp. 147-161 in J.J. Spengler and O.D. Duncan (eds., 1956): *Demographic Analysis. Selected Readings.* The Free Press, Glencoe.

[26] Sverdrup, Erling (1967). Statistiske metoder ved dødelighetsundersøkelser. Second Edition. Mimeographed. Institute of Mathematics, University of Oslo.

[27] Vogt, Johan (1952). Befolkningsutvikling og ekteskapsstruktur i Norge. *Statsøkonomisk Tidsskrift (J. Econ. Ass. Norway).*

[28] Vogt, Johan (1964). Component parts of the number of births. *Statsøkonomisk Tidsskrift (J. Econ. Ass. Norway)* 1964 (4) 287-307.

[29] Whelpton, P.K. (1946). Reproduction rates adjusted for age, parity, fecundity, and marriage. *J. Am. Statist. Ass.* 41 501-516.

[30] Whelpton, P.K. (1949). Cohort analysis of fertility. *Am. Sociol. Rev.* 14 (6) 735-749.

[31] Whelpton, P.K. (1954). *Cohort Fertility: Native White Women in the United States.* Princeton University Press.

[32] Wolfbein, S.L. (1949). The length of working life. *Population Studies* 3 (3) 286-294.

[33] Woofter, T.J. (1947). Completed generation reproduction rates. *Human Biology* 19 133-153.

[34] Zwinggi, Ernst (1945). *Versicherungsmathematik.* Verlag Birkhäuser, Basel.