POPULATION ANALYSIS AND DEMOGRAPHY

23 April 1990
Griffith Feeney
East-West Population Institute
Honolulu, Hawaii 96848

ABSTRACT

Formal demographic concepts have myriad applications both within and beyond the study of human populations. The object of this paper is to explore systematically the nature and limits of this generality. `Populations' are defined in a way that leaves the nature of the entities comprising them entirely arbitrary. It is shown that most tools and concepts of demographic analysis apply in this completely general setting. Recognition of the full generality of `demographic' analysis is a powerful tool for understanding because it reveals limitations as well as capabilities. The implications of these results for our understanding of the technique and substance of demography are discussed.

CONTENTS

POPULATION ANALYSIS AND DEMOGRAPHY

Griffith Feeney

The range of formal demographic concepts is extraordinary. One of Alfred J. Lotka's earliest contributions to mathematical demography is a 1907 paper on the population dynamics of `material aggregates'. Kenneth Boulding provided the first formal definition of `population' and applied it to the theory of capital (1934) and to the automobile population of the United States (1955). Norman B. Ryder has detailed the application of population concepts to social as distinct from demographic aggregates (1964b), a theme continued in Ford and De Jong (1970). Hartshorn (1975) has applied life table and matrix population dynamics concepts to the study of tree populations in tropical rain forests. James R. Carey has spent much of the past decade developing the discipline of— with apologies to etymological purists— `insect demography' (1982, 1983, 1984, 1985).

It would not be difficult to extend this list with more diverse and exotic applications. For most demographers, however, the important issues lie closer to home, in the application of familiar methods to new demographic situations. Life table ideas were developed for the study of human mortality, but they have long been applied in the most diverse corners of the field, including the study of marriage and divorce, birth intervals, and contraceptive use. Reflecting on this diversity, we find certain common elements. In every case we deal with one or more subpopulations of a given human population; in every case certain events signify entry to and exit from the subpopulation; and in every case the entry and exit events are formally analogous to birth and death, allowing the introduction of life table concepts.

The object of this paper is explore systematically the nature and limits of generality of formal demographic concepts. I begin by proposing a formal, set-theoretic definition of the population concept, a definition that strips away any vestige of reference to the particular nature of the entities that comprise the population. `Population analysis' is then taken to be the study of those propositions that hold for populations in general, whatever their members may be. I proceed ask how much of what is generally recognized as formal demographic analysis may be derived from this definition. The answers are surprising and instructive, both for how much of what we think of as 'demographic' analysis is in fact entirely general, and for how much more there is to demography than population analysis.

Mathematics teaches that abstraction sharpens understanding. The formalization of population analysis provides a striking case in point. The familiar concept of a population closed to migration seems an unlikely subject for intellectual (as opposed to definitional) subtlety, but its extension to subpopulations requires some unexpectedly sharp thinking. Life table concepts are centuries old, the last place one would expect to find unanswered questions, but determined abstraction yields at least one. The very definition of 'population' provides insights into the mundane but important problem of organizing demographic data.

Disciplining ourselves to think abstractly about familiar ideas enables us to think more clearly and deeply about certain aspects of them. This is not to deny that our ultimate concerns are more complex, nor that there will always be a time to move from the abstract to the particular. It is simply to say that we are likely to get further in complex matters if we have first mastered some of their simpler aspects.

POPULATIONS

People in the United States, penguins in Antarctica, automobiles in China—each of these phrases identifies a population. What do they have in common? To begin with, each consists at any given time of a certain set of entities. A population is not simply a set, however, for it has an existence in time. Individuals may enter and leave the population, and members at one time are not necessarily members at other times. Populations, we might say, are `sets whose membership changes with time,' but this turns out to be only one of several equivalent and equally useful formulations.[1]

§1 Universe, Census, and Span. The essential population concepts are membership and time. Membership is specified in two stages. We begin by asking what sort of entities are involved and regard the answer as a pool from which members of the population are drawn. We then specify a particular population with members drawn from this pool by indicating which individuals are members at which times.

A population may thus be regarded formally as a collection of pairs (u,t), where u is a member of some set U and t is drawn from an interval of real numbers representing points in time. This leads us to define a **population** as a relation between an arbitrary set U, called the **universe** of the population, and the set R of the real numbers (or a subset of this set) representing times.[2] We may denote populations by letters, such as P, and write uPt to signify that individual u is in population P at time t.

Given a population P with universe U, we may ask which individuals in U are members of P at any given time t. Abstracting from familiar usage, we call this set the **census set**, or simply the **census**, of population P at time t. In set-theoretic terms, we define a set-valued function P_C mapping the interval R of times into the power set 2^U of U by

$$PC(t) = \{u:uPt\} \tag{1}$$

We call P_C the **census function** of the population.[3]

Similarly, for any individual u we may ask for the set of all times t at which this individual is a member of population P. This leads to the set-valued function P_S, defined by

$$P_S(u) = \{t:uPt\} \tag{2}$$

which maps the universe U into the power set 2^R of R. We call $P_S(u)$ the **span** of individual u in population P, after `life span,' and P_S the **span function** of the population. Observe the necessity of referring to the span of an individual in a particular population. Individuals may belong to many different populations and have a different span in each.

The census function suggests the characterization of a population as 'a set whose membership changes with time.' The span function suggests thinking of a population as 'a bundle of life histories,' where 'life' is used in the general sense in which we may speak of the 'life' an automobile or other non-biological entity. The two conceptions are formally equivalent in the following sense. Given a bundle of life-histories, we may construct a census at time t by passing through the life histories one by one and determining, for each individual, whether or not this individual was in the population at time t. In set-theoretic terms we construct the values of the census function as

$$P_C(t) = \{u:t \in P_S(u)\} \tag{3}$$

for every time t.[4] Conversely, given a complete set of censuses for a population, we may construct a bundle of life histories by asking, for each individual, the times at which this individual was in the population. In set-theoretic terms, we construct the values of the span function as

$$P_S(u) = \{t:u \in P_C(t)\} \tag{4}$$

for every individual u.

The formal symmetry of these last two expressions is disturbed by the different cardinalities of the sets involved. For a finite population, the number of life histories will of course be finite, while the number of possible times is uncountably infinite. Clearly we will prefer to construct censuses from bundles of life histories rather than the other way around. Note however that population size in a finite population necessarily remains fixed for finite time intervals. Thus although the number of times is infinite, the number of censuses required to fully specify a population is not. Nonetheless, while a finite population may be fully characterized by a finite number of censuses, this is clearly a far less efficient way to organize population data because it requires information in each census sufficient to identify individuals between censuses.

Continuous populations, illustrated by Boulding's (1934:637) example of water in a lake, may be mentioned briefly. Mathematical formalization is readily accomplished by taking universes to be the sort of continuous mathematical space used to represent physical space. The appropriate mathematical concept is the measure space (Halmos 1950), and we shall have to require than census sets be what are called `measurable sets'. It is not immediately clear what complications would arise in such a generalization, nor what uses it might have. One might suppose that it would hold little significance for demography, but there is little basis for such a judgment without further development of the subject.

¶2 Subpopulations. A population Q is a **subpopulation** of another population P with the same universe if $Q \subseteq P$, that is, if uQt implies uPt. Population P may be called in this context the **parent** population of Q. If Q is a subpopulation of P, then the census of Q at time t is a subset of the census of P at time t for all times t. In the set-theoretic formalism of the preceding section, $Q_C(t) \subseteq P_C(t)$ for all t, which holds because $u \in Q_C(t)$ implies uQt implies uPt implies $u \in P_C(t)$.[5] By the same reasoning, the span of any individual u in Q is a subset of the span of this individual in P, i.e., $Q_S(t) \subseteq P_S(t)$. Furthermore, both of these conditions imply that Q is a subpopulation of P, for if either $Q_C(t) \subseteq P_C(t)$ for all t or $Q_S(u) \subseteq P_S(u)$ for all u, then uQt implies uPt, whence $Q \subseteq P$. The three possible definitions of subpopulation, in terms of relations, censuses, and spans, are thus equivalent.

A subpopulation is itself a population, whence anything that applies to populations in general applies to subpopulations. This apparently mundane observation accounts for much of the extraordinary scope of application of population concepts. Virtually any characteristic of individuals may be used to define a subpopulation. In human populations, for example, we have, leaving aside familiar demographic examples, medical doctors, employees of IBM, union members, students, unemployed Ph.D.s, state and federal prisoners, residents of Utah, members of congress, and the Federal Bureau of Investigation's 10 Most Wanted List. Studying Boulding's automobile populations we might be concerned with subpopulations of new cars, used cars, registered and unregistered cars, cars in manufacturers' and dealers' inventories, and foreign and domestic cars. City mayors and university presidents worry about populations of parking spaces, railroad executives about populations of engines, box cars, and railroad ties. Lists of this sort are readily extended until the diversity becomes monotonous.

§3 Birth and Death; Events Generally. An entity of any kind is **born**, and a **birth** occurs, when the entity comes into existence; it **dies**, and a **death** occurs, when it ceases to exist. Artificial as these designations may be in particular contexts, we require some general terms for coming into existence and ceasing to exist, and \`birth' and \`death' are arguably the best choices. In specific contexts, of course, we speak of \`planting' a tree, \`filing' a law suit, and so on.[6]

More generally, an **event** is anything that may happen to the individuals in some universe at a reasonably well-defined point in time. Events are associated with the individuals in some universe, not with any particular population that draws members from this universe. Associated with every population, however, are the events of **entry** to and **exit** from this population. Entry to and exit from a particular population are examples of **classes** of events, specific instances of which may occur, perhaps repeatedly, to particular individuals.

When viewed at sufficiently close range, events often turn out to be processes that develop over a period of time. In assigning them to a point in time we express a judgement that this period is sufficiently short, in context, that it may be ignored. Thus birth and death are from the medical standpoint more or less extended processes, but in demography they are appropriately regarded as events occurring at a point in time.

A class of events is **recurrent** if an individual may experience an event in this class more than once; otherwise the class is **non-recurrent**. Death is a non-recurrent event. Birth in human populations is recurrent with respect to the mother but non-recurrent with respect to the child. Marriage and divorce are recurrent events. A class of recurrent events may always be decomposed into a series of classes of non-recurrent events by ordering the events occurring to each individual chronologically and numbering them. Thus the set of all marriages may be decomposed into first marriages, second marriages, and so on.

Two classes of events, taken in order, are said to be **contingent** if an individual cannot experience an event in the second class without having first experienced an event in the first class. Contingent events generally take the form of the acquisition and loss of some status. Birth and death are contingent events, as are marriage and divorce, and entry to and exit from the labor force. It is not necessary, as these examples show, that any particular individual experience either one of a pair of contingent events.

Every pair of contingent events defines a population, the population of individuals who have, at any given time, experienced the first event but not the second. Conversely, every population defines at least one pair of contingent events, first entry to and last exit from the population.

Compound events are defined in terms of the occurrence of two or more other events. An entry to a population of employed persons is a compound event that consists either of the employment of a person not previously in the labor force or of the employment of a previously unemployed person. Dissolution of marriage is a compound event that occurs when a couple divorces or when the wife and/or husband dies. In these examples the compound event occurs with the occurrence of any one of the events of which it is compounded. Other compound events may require the simultaneous occurrence of two different events, as in the movement of a new wife into the husband's household on the occurrence of a virilocal marriage.

§4 Closure to Migration and Non-Recurrence. Given the population of a particular country, consider the subpopulation of childless women. Women enter this population at birth and leave in when they have their first child or, if they remain childless, when they die. If the given population is open to migration, however, entries to and exits from the subpopulation will occur when a childless women migrates into or out of the parent population. It evidently makes sense to say that the subpopulation is `open to migration' if this happens and `closed to migration' if it does not. This example motivates the following general definition of closure to migration.

Given any population with a particular universe, let certain classes of events that may occur to individuals in this universe be designated **proper entry** events for the population, and let certain other classes of events be designated **proper exit** events for the population. An entry to such a population that does not occur as a result of a proper entry event is then called an **in-migration**, and an exit that does not occur as a result of a proper exit event is an **out-migration**. A population that experiences no in- or out-migrations is said to be **closed to migration**, or simply **closed**. A population not closed to migration is said to be **open to migration**, or simply **open**.

By the **classical case** we mean a population in which the single proper entry event is birth and the single proper exit event is death. In the classical case, the preceding definition reduces to the familiar one: a population is closed if every entry is a birth and every exit a death. Observe that every population may be regarded as a subpopulation of a classical case; we simply extend the span of every individual who is ever a member of the population to include the interval between this individual's birth and death.

Proper entry and exit events must be specified explicitly. `Married persons with United States citizenship' does not tell us whether the proper entry event is marriage, becoming a citizen, or both. If we identify marriage as the (only) proper entry event, the naturalization of a married person represents an in-migration, but if we identify birth in the United States and naturalization as proper entry events, the marriage of a citizen represents an in-migration. Such distinctions, however unwelcome, are a necessary consequence of extending the concept of closure to migration to subpopulations.

Human populations have traditionally been defined by presence or residence in a particular geographical area. `Migration' has thus come to be associated with movement into or out of populations. It is this meaning, rather than the original meaning of movement in geographical space, that carries over into population analysis.

A population is **recurrent** if individuals may enter and exit and exit more than once; otherwise a population is **non-recurrent**. Non-recurrence does not imply closure to migration because an individual may migrate into or out of a population and still enter and exit only once. Closure to migration does not imply non-recurrence because proper entry and exit events may be recurrent. A population of married persons is likely to be recurrent, for example, whether or not it is closed, because marriage is a recurrent event. Closure to migration implies non-recurrence only when proper entry and exit events are non-recurrent, as in the classical case.

<u>§5 Age and Duration</u>. An individual's **age** at any given time is the time elapsed since this individual's birth. **Age in completed years** (or `at last birthday') refers to the greatest integer less than an individual's age. Age may be referred to as **exact age** to emphasize the distinction between age and age in completed years. `Age' in demography is often used to mean either exact age or age in completed years, the appropriate interpretation being given by context. The same terminology applies to any time unit, so that we may refer, for example, to `exact age in months' or `age in completed weeks.'

More generally, measuring time backward or forward from the time at which a particular individual experiences a particular event leads to what may be called **elapsed time** variables. We will usually think of elapsed time variables as being attached to the individual who experiences the event in question and as being defined only during the life-span of this individual. We allow negative values, however, so that, for example, time elapsed since first marriage assumes negative values before, and positive values after, first marriage.

Elapsed time variables have the property of increasing identically with an individual's age. We call any such variable **age-like**. Since every age-like variable may be thought of as an elapsed time variable corresponding to the `event' that occurs to the individual when the variable assumes the value zero, the concepts of elapsed time and age-like variables are co-extensive.

An individual's **duration** in any status is defined as the length of time an individual has spent continuously in this status. Since acquiring or losing any status may be regarded as an event occurring to the individual in question, duration in any status is the same as time elapsed since most recently acquiring this status, but only so long as the individual continues in this status. Duration variables are defined only for times at which an individual is in the status in question; they reset to zero each time an individual re-acquires this status.

§6 The Demographic Equation. Consider a non-recurrent population observed over some time period. Let P denoting the set of individuals present in the population at the beginning of the period; E and X the sets of individuals entering and exiting the population during the period, respectively; and P' the set of individuals present in the population at the end of the period. Evidently we should have[7]

$$(P \cup E) - X = P' \qquad (1)$$

To prove this set identity we have to show that any individual in $(P \cup E)-X$ is in P' and conversely. If some individual u is in $(P \cup E)-X$, then this individual is either in P or E and not in X, i.e., this individual was either present at the beginning of the period or entered during the period, and did not exit during the period. Such an individual must be in the population at the end of the period, i.e., must be in P'. Conversely, if some individual u is in P', then this individual must either have been in the population at the beginning of the period or have entered during the period, i.e., must be in $P \cup E$, and must not have exited during the period, i.e., must not be in X, which implies membership in $(P \cup E)-X$.

The corresponding numerical equation follows, assuming the sets in question to be finite, if (i) P and E are disjoint and (ii) X and P' are disjoint. These conditions raise the question of whether an individual who enters or exits a population at time t exactly is or is not a member at time t. It seems natural that an individual who enters at time t should be a member of the population at this time.

If the sets P and E are to be disjoint, however, this requires that this individual not be regarded as an entry during any time period beginning at t. This in turn implies that time periods should be regarded as intervals open on the left. If time intervals are open on the left, however, it is natural to require that they be closed on the right, for otherwise putting together successive time intervals will omit some points in time. We thus regard time intervals as left-open, right-closed intervals, written (t,t+m]. The disjointness of the sets X and P' will then require that we regard an individual who leaves a population at time t exactly as not being a member at time t. This leads us to regard spans of population membership as left-closed, right-open intervals, written [x,y). Finally, since we want the numbers of individuals in any partition of the population into age or duration groups at any given time to sum to the total population, we take age and duration groups to be left closed, right open intervals, written [x,x+n). With these endpoint conventions, (1) yields the numeric form of the demographic equation.

Formula (1) is not meaningful for recurrent populations, for in this case P and P' are sets of individuals, whereas E and X are sets of events, and the set subtraction on the left in (1) is not defined. To obtain the numerical demographic equation for recurrent populations, observe that we may derive a non-recurrent population from any population by simply disregarding the identity of individuals between successive intervals of membership.[8] This non-recurrent population has the same number of members at any given time, and the same number of entries and exits during any given period, as the given population. It differs from the given population only in that every entry is regarded as a different individual. The numerical form of the demographic equation of a recurrent population follows by observing that (1) holds for this derived population, because it is non-recurrent, and that the derived population has the same numbers of members, entries, and exits as the given population.

To obtain a set-theoretic form of the demographic equation for recurrent populations we proceed as follows. An individual is said to make a **net exit** from a population during a time period if (i) this individual was either present at the beginning of the period or entered during the period and (ii) was not present at the end of the period. An individual makes a **net entry** to a population during a period if this individual was present at the end but not at the beginning of the period. Redefining the set E to be the individuals who make a net entry during the period and the set X to be the set of individuals who make a net exit during the period makes (1) meaningful and true for recurrent as well as non-recurrent populations.

The terms of the numerical demographic equation representing entries to and exits from a population are often referred to as `components' of population change. More generally, two quantities A and B are said to be **components** of a third quantity C if either A + B = C or A - B = C holds tautologically, that is, as a logical consequence of the definitions of the three quantities. In human populations, for example, total population size equals the number of females plus the number of males, total deaths during any period equals male deaths plus female deaths, and total years lived by any person equals the sum of years lived before and after age 30. In each case the numerical identity derives from a partition of an underlying set into subsets, a set of individuals in the first case, a set of events in the second, and an interval of time in the third.

POPULATION STATISTICS

Demography projects society onto a relatively small set of numbers that describe the population of this society. The first step in this projection is the abstraction of `population' from `society'. The population consists of, and is ideally described by, a bundle of life histories of all the persons who were members of the population during some specified time period. The second step is the calculation of various aggregate statistics from these life histories. The following sections describe the main processes of aggregation, virtually all of which apply to populations in general.

§7 Cohort and Period Statistics. The relatively complex set-theoretic structure of populations results in two fundamentally different types of aggregation, known in demographic shorthand as `period' and `cohort'. We adopt a cohort perspective when we follow a situation or a group of individuals over time. We adopt a period perspective when we ask what happens during a particular time period. These ideas are extremely general, reaching beyond even purely formal population analysis. Accounting, for example, began with records of particular ventures, such as a voyage to the East Indies, and accounts were closed only when the venture was complete, a cohort perspective. As commerce developed, accounts began to be produced for successive accounting periods, a period perspective.[9]

A **cohort** is a group of individuals who experience a particular event during a given time period. It is customary to tag cohorts with the name of the event defining them, as in `birth cohort', `marriage cohort', `divorce cohort', and so on. Given any cohort, and given any population of individuals drawn from the same universe, we may form the subpopulation that consists at any given time of the individuals in the population who are members of the cohort. A population formed in this way is called a **cohort**

subpopulation. The word `cohort' is used to refer both to cohorts proper and to cohort subpopulations, the intended meaning being given by context.

Cohorts are the natural aggregates for answering certain kinds of questions about populations. Most notably, how long do indi-viduals remain in a population? What is the mean duration of stay or the distribution of length of stay? `Cohort life tables', defined below, are a device for giving systematic answers to such questions. More generally, a **cohort statistic** is any statistic that may be computed solely from information about the individuals comprising a particular cohort.[10]

For other questions, the natural aggregates are time periods. Most notably, how fast is a population growing? How many members are added or subtracted during a given period? What is the magnitude of this number in relation to total population size? A **period statistic** is any statistic that may be computed solely from information on the individuals who are members of a population during a given time period.

§8 The Principle of Exposure to Risk. Numbers of events must be considered in relation to the numbers of individuals who might have experienced these events. That deaths in China outnumber deaths in Monte Carlo tells us nothing about the level of mortality in these two countries; there are more deaths in China because there are more people in China. We must also take account of the length of time individuals are observed, for other things being equal, more deaths will be observed over a longer period than over a shorter period. It is not possible to say anything about mortality as such without taking account of these two factors. This is the **principle of exposure to risk**. With qualifications that vary by context, it applies not just to mortality, but to events generally.

Application of the principle of exposure to risk is always a matter of degree. It is not the only consideration and should rarely if ever be carried through to the bitter end. The crude birth rate is an inferior indicator of fertility, but it expresses precisely the effect of current fertility on the population growth rate, a very useful property. Age-specific birth rates will often be a sensible choice for making population projections precisely because they relate births to all women in each age group, the numbers of which are readily available. From the point of view of exposure to risk, rates excluding (say) sterilized and imprisoned women would be preferable.

§9 Life Tables. Consider a class of non-recurrent contingent events defined for the individuals in some universe. Suppose that for some group of these individuals, all of whom have experienced both events, we are given the intervals x_1, x_2, ..., x_N between the time of occurrence of the first and second events.

It is useful to think of these individuals as being members of the population defined by the given pair of contingent events. The first contingent event represents an entry to this population, the second contingent event an exit from it. The interval x_i is thought of as the life-time of the i-th individual in this population, and also as the duration at which this individual exited, duration being measured for each individual from time of entry. Given any duration x, we say that the i-th individual `survives to' or `remains at' duration x if $x_i \geq x$, i.e., if the duration at which this individual exits lies in the open duration interval $[x,\infty)$.

A **life table** consists of a more or less standard set of statistics describing the distribution x_1, x_2, ..., x_N classified into a series of duration intervals of the form $[x,x+n)$. Life tables generally give the number or proportion of individuals for which $x_i \geq x$, denoted l_x; the number or proportion of individuals who exit in the duration interval $[x,x+n)$, denoted ${}_nd_x$; and the proportions ${}_nq_x={}_nd_x/l_x$. Whether l_x and ${}_nd_x$ are to be understood as numbers or proportions is generally clear from context. Note that, according to these definitions, individuals exiting at exact duration x are counted in both l_x and ${}_nd_x$.

In the classical case, the function defined by l_x is called a **survival function**, a usage often carried over into more general contexts. It may also be called a **retention function**. The functions defined by ${}_nd_x$ and ${}_nq_x$ may be referred to as the **deaths** or **exits function** and the **risk function**, respectively.

The values of l_x and ${}_nd_x$ for a suitable series of values of x and n are simply the empirical density function and the arithmetic complement of the cumulative distribution function, respectively, of descriptive statistics. It is the ${}_nq_x$ values that make the life table distinctive. The numerators of ${}_nq_x$ and ${}_nd_x$ are the same, the number of individuals exiting in the interval $[x,x+n)$. The denominator of ${}_nd_x$ is the total number of individuals in the group, however, whereas the denominator of ${}_nq_x$ is the number of individuals remaining at duration x. Because the number of exits in the interval $[x,x+n)$ cannot exceed the number of individuals remaining at duration x, the value of ${}_nd_x$ is dependent on, and hence influenced by, what happens in the interval $[0,x)$ as well as by what happens in the interval $[x,x+n)$. Dividing ${}_nd_x$ by l_x eliminates this influence and produces a statistic that refers solely to experience in the given duration interval.

Several elementary but useful relations follow at once from the definitions of l_x, ${}_nd_x$ and ${}_nq_x$.

$$ {}_nd_x = l_x - l_{x+n} \tag{1} $$

$$ l_{x+n} = l_x - {}_nd_x \tag{2} $$

$$ {}_nq_x \equiv {}_nd_x/l_x = 1 - l_{x+n}/l_x \tag{3} $$

$$ l_{x+n} = l_x(1-{}_nq_x) \tag{4} $$

Applying (4) recursively to a series of duration intervals starting at zero and taking $l_0 = 1$ gives

$$ l_x = \prod_{z<x} (1-{}_nq_z) \tag{5} $$

for any $x>0$, where the notation is understood to mean that the product is taken over a set of duration intervals partitioning the interval $[0,x)$.

The l_x, ${}_nd_x$ and ${}_nq_x$ `columns' of the life table, as they are often called on account of the usual tabular format, are nearly equivalent. Given the l_x column, i.e., the l_x values for a complete series of duration groups, we may compute the ${}_nd_x$ column using (1) and then the ${}_nq_x$ column using (3). Given the ${}_nd_x$ column, we may compute l_0 by summing over all duration intervals, then the l_x column using (2), and finally the ${}_nq_x$ column using (3). Finally, given the ${}_nq_x$ column and the value of l_0, we may compute the l_x column using (4) and then the ${}_nd_x$ column using (1). This last case is the exception to complete equivalence, for we need the value of l_0, called the **radix** of the life table, as well as the ${}_nq_x$ column to compute the other two columns.

Complete data may be represented in life table form by ordering the life-times x_1, x_2, ..., x_N and letting the resulting sequence define the duration intervals. In this case every duration interval $[x,x+n)$ will contain $j \geq 1$ exits at exact duration x and no exits at any other duration. Since all exits occur at the beginning of the interval, it is evidently unsatisfactory to continue to think of ${}_nd_x$ and ${}_nq_x$ as referring to the interval $[x,x+n)$. We therefore drop the interval prescripts and write d_x for the number of exits at exact age x and $q_x = d_x \backslash l_x$ for the observed risk of exit at exact duration x. The interpretation of the survival function l_x does not change, but in place of (5) we have

$$ l_x = \prod_{z \leq x} (1-d_z/l_z) \tag{6} $$

where z ranges over all durations at which one or more exits occur.

Life tables often include three further columns. The mean life-time is by definition $(\Sigma x_i)/N$. The quantity x_i may be written $\Sigma_j x_{ij}$, where x_{ij} denotes the length of the length of the j-th duration interval if x_i falls in a subsequent interval and x_i minus the beginning of the interval if x_i lies in this interval. It follows at once that

$$\Sigma_i x_i = \Sigma_i\{\Sigma_j x_{ij}\} = \Sigma_j\{\Sigma_i x_{ij}\} \tag{7}$$

The quantity $\Sigma_i x_{ij}$ is denoted by ${}_nL_x$, where $[x,x+n)$ is the j-th duration interval. The sum ${}_nL_x + {}_nL_{x+n} + {}_nL_{x+2n} + \ldots$ is denoted by T_x. These quantities are referred to as the **exposure** in the interval $[x,x+n)$ and after duration x, respectively. Exposure in the classical case is usually referred to as **person years lived**. Dividing (7) by $N=l_0$ we see that the mean life-time may be written as T_0/l_0. A similar argument shows that the mean remaining life-time for individuals remaining at duration x, denoted by e_x, is given by

$$e_x = T_x/l_x \tag{8}$$

This formula holds whether l_x represents numbers or proportions since division of l_x by N results in the division of T_x by N as well.[11]

The restriction of life tables to non-recurrent contingent events is not as severe as it may at first seem. Recurrent contingent events, such as marriage and dissolution of marriage, may be handled in either of two ways. First, we may order and number the events experienced by each individual and consider the resulting pairs of contingent events separately. Thus we consider formation and dissolution of first marriage, formation and dissolution of second marriages, and so on. Second, we may ignore the identity of individuals between successive occurrences of the events in question. We may consider a population of marriages, for example, independently of the order of marriage for the two partners.

§10 Cohort Life Tables. Life table statistics may be formally calculated for any group of individuals whatever, but the meaning of the resulting statistics depends on the way in which the group is selected. A life table computed from the ages at death of all persons who die in a population during a given year is not generally considered to be meaningful, for example, because it reflects the population age distribution as much or more than it reflects mortality.

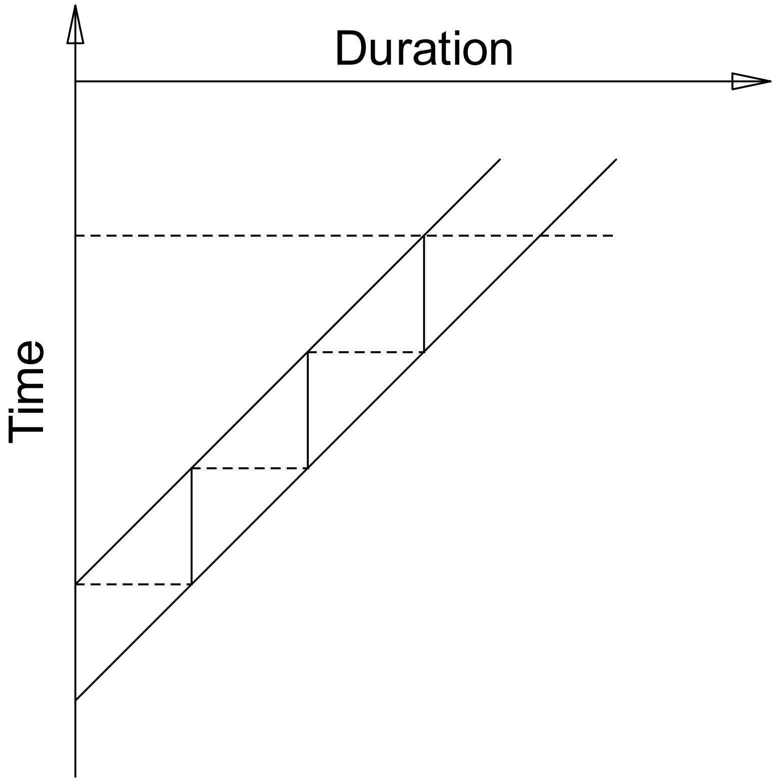
In general, life tables make sense only if the life-times represented are appropriately selected. Some selections, such as the birth cohort in the study of mortality, are generally recognized as appropriate. Others, such as the selection of closed birth intervals from retrospective fertility survey data are recognized, somewhat less widely, as faulty. Conditions for group selection at once general and precise seem never to have been formulated.[12]

A plausible condition is that the selection procedure should not require any information that becomes available only after the occurrence of the first contingent event. The rationale for this criterion is that such selection involves indirect selection of intervals according to their length: only intervals exceeding some minimal length will be included because the relevant selection information will not be available for other intervals. This condition allows the selections generally regarded as sensible and disallows many that are obviously not sensible. Selection of human life-times by birth cohort passes, for example, since we need only information on time of birth to define them, but selection of life-times that end in a given year fails because it requires information on the time of death.

A **cohort life table** is a life table in which the group of individuals represented is a cohort, i.e., in which individuals are selected on the basis of having experienced a specified event during a given time period. Cohort life tables assure reasonable group selection. They also locate the individuals represented in historical time. The life table statistic ${}_nq_x$ has special significance in this context, for division of ${}_nd_x$ by l_x has the effect of removing the influence of the experience of the cohort prior to the time when the first individual in the cohort reaches duration x. The calculation of ${}_nq_x$ values does not require data on experience of the cohort prior to duration x, an important consideration when data have limited historical depth.

When the statistics comprising a cohort life table are represented on the Lexis diagram we obtain a diagonal band, representing the cohort as a whole, divided into a series of parallelograms formed by the various duration groups, as shown by the solid lines in the following Lexis diagram.

(Next Page)



The vertical line at exact duration x represents the l_x individuals remaining at this duration. The parallelogram formed by the lines at exact durations x and x+n represent the $_nd_x$ individuals exiting in the interval [x,x+n).

An alternative disaggregation of the cohort is provided by the dotted horizontal lines, which correspond to time periods. While these lines might be drawn at arbitrary time points, they are generally chosen to mark off equal time intervals whose length is that of the time interval defining the cohort. This is the situation depicted in the figure above. Analogues of the three basic life table statistics are readily defined, and they satisfy relations precisely analogous to (1-5) of section 9 above. To see this, simply identify l_x for x>0 with the number or proportion of individuals present at time x, time being measured from the end of the time interval defining the cohort, and $_nd_x$ with the number or proportion of individuals exiting during the time interval (x,x+n], n denoting the width of the intervals.

This time-period-based disaggregation presents us with a triangle of exits at the beginning of the cohort's experience and might therefore seem less natural that the more familiar duration-based disaggregation. Population data is always truncated by the end of the observation period, however, and the usual disaggregation by duration gives a triangle at the end of each cohort's experience, excepting only those cohorts all of whose members have exited before the end of observation. What Norman Ryder (1982:16) has aptly termed "the embarrassment of triangles" is unavoidable except in historical studies.

Population projection provides the most familiar example of disaggregation of cohort experience by time periods rather than by duration groups. The **retention ratio** for a given cohort and time period is the number of individuals present in the cohort at the beginning of the period divided into the number of these individuals present at the end of the period.

Retention ratios, like ${}_nq_x$ values, operationalize the principle of exposure to risk. Life table ${}_nq_x$ values are **duration-exact** in the sense that all individuals are followed over the same duration interval. Retention ratios are **time-exact** in the sense that all individuals are followed over the same time period. Given the historical context in which life table concepts developed, with human mortality varying sharply by age but relatively constant over time, actuaries and demographers have tended to regard ${}_nq_x$ as the preferred measure of mortality. From a purely formal point of view, however, there is no reason to prefer duration-exactness to time-exactness. The variation of human death rates with age is an empirical matter. Many species exhibit approximately exponential mortality, and even where duration variation is important, time variation may be of equal or greater importance, as in the study of epidemics. The statistic ${}_nq_x$ in a cohort life table is of course time-inexact in precisely the same sense that retention ratios are duration-inexact.[13]

§11 Occurrence-Exposure Rates. A third way of operationalizing the principle of exposure to risk occurs when different individuals are followed over different lengths of time. Consider for example the kind of information provided by most vital registration systems: deaths in a population that occur between times t and t+m at an age between x and x+n. Individuals whose deaths are thus described must either have been aged x to x+n at time t or have reached exact age x between times t and t+m. Since these individuals are exposed to the risk of death for varying lengths of time during this time period and age interval, we need to relate the number of deaths to a quantity that takes account not only of the number of individuals exposed, but the time that each is exposed.

The obvious choice is the sum over all individuals of the time spent by each individual in the given age group during the given time period. In general, the **exposure** of a population in any set A in the Lexis plane is defined as the sum over all individuals whose life-lines intersect A of the number of years spent by these individuals at time and duration points contained in A. This definition is consistent with that given above for exposure during a duration interval in a cohort life table, for which the set A is the parallelogram in the age-time plane of the Lexis diagram defined by the given cohort and duration interval.

It is useful to think of individuals as `coming into observation' for the first time, with respect to the set A, when their life line in the Lexis diagram first touches A and `leaving observation' for the last time when it last touches A. If the population is non-recurrent and the set A is convex, which in practice it usually is, the individual will be exposed from the time of first coming into observation to the time of last leaving observation. If the population is recurrent, the lengths of any intervals of non-membership in the population by the individual will be subtracted from this interval.

Given any population and any class of events experienced by individuals in this population, the **occurrence-exposure rate** for any set A in the Lexis plane is defined as the number of events in the given class that occur in A divided by the exposure of the population in A. When A is the band in the Lexis plane corresponding to a time period, this is the **crude rate** of the event for this period. When A is the rectangle formed by the intersection of a time period and a duration or age group it is the **duration- or age-specific rate** for the population for this duration or age group and time period.

The sweeping generality of occurrence-exposure rates is seductive but misleading. We may define rates for births of any order, for example, but since only women who have already had i children can have an (i+1)st birth, these rates violate the principle of exposure to risk in a way that age-specific death rates do not. Age-specific death rates are analogous to life table ${}_nq_x$ values in that a proper relation of events and exposure to risk is preserved by removing deceased individuals from the denominator; the removal occurs immediately in the case of age-specific death rates and at age x+n in the case of ${}_nq_x$, but it occurs in both cases. This removal of individuals from the denominator of the rate is not possible in the case of birth rates because birth is a recurrent event. Age-specific birth rates are analogous to the ${}_nd_x$ values in the life table in the sense that their denominator does not change when a woman has a birth. Under the often maintained assumption that mortality is negligible during the reproductive ages, the denominators of age-specific birth rates for a birth cohort will be identical and the rates will describe the frequency distribution of births by age, just as the ${}_nd_x$ values in a cohort life table describe the frequency distribution of deaths by age.

To compute birth rates analogous to the ${}_nq_x$ values of the life table we must relate i-th order births to women with i-1 children, the only women capable of having an i-th birth. This was done in one way by Whelpton (1954), in which rates are taken specific for age and parity, and in another way by Henry (1953/1980), in which rates are taken specific for parity and time elapsed since last birth or, in the case of first births, duration of marriage or age. Both approaches generalize to any recurrent event, the essential idea being to disaggregate by order and to relate events of each order only to individuals who have experienced the events of the previous orders. For recent applications and discussion of these and related issues of fertility measurement see Feeney and Yu (1987), Lutz (1989), Luther, Feeney and Zhang (1990), and Feeney and Lutz (1990).

§12 Partitioned Populations. Individuals in a population are frequently classifiable, at any given time, into one of a finite set of mutually exclusive and exhaustive categories or `states'. Any such classification defines a **partition** of the population into mutually exclusive and exhaustive subpopulations. These subpopulations will in general be open to migration and recurrent even if the given population is closed and non-recurrent.

Partitioning by age or duration is most familiar and arguably most fundamental, but partitions may be based on any attribute whatever. Biologists sometimes classify by stage of growth in place of age, as in Lefkovitch (1965) and Caswell (1989). Classifications may be compounded of two or more attributes. The study of populations partitioned by residence and age group, for example, was pioneered by Rogers (1968, 1975; see also Feeney 1970, 1973). Enzer (1964) and Matras (1967) had earlier considered the same formal model in connection with, respectively, partitioned populations of capital goods and social mobility. Female populations partitioned by age, parity, and/or duration in parity are used in the study of fertility (Murphy 1965; Oechsli 1975; Feeney 1983). Tuljapurkar (1983) gives a continuous formulation of what is essentially the same model for yeast cell populations.

Life table ideas may be extended to describe the movement of individuals between states during their life-time. Consider first what happens over a single duration interval [x,x+n). Let $l_x(j)$ denote the number of individuals present and in category j at duration x, S_j the number of these $l_x(j)$ individuals remaining at duration x+n, S_{ij} the number of these S_j individuals who are in category i at duration x+n, and $l_{x+n}(i)$ the number of individuals present and in category i at duration x+n. Tautologically,

$$S_{ij} = \frac{S_{ij}}{S_j} \, \frac{S_j}{l_x(j)} \, l_x(j) \qquad (1)$$

for all i and j, because the S_j and $l_x(j)$ terms cancel out. Also, since the individuals in category i at duration x+n may be disaggregated according their category at duration x,

$$l_{x+n}(i) = S_{i1} + S_{i2} + \ldots + S_{in} \qquad (2)$$

for all i. Combining (1) and (2) gives

$$l_{x+n}(i) = \Sigma_j \; m_{ij} s_j l_x(j) \qquad (3)$$

for all i, where the transition proportions p_{ij} and the survival proportions s_j are the quotients in (1).

Formula (3) describes the redistribution of individuals between categories in the duration interval [x,x+n). It may be written in matrix form as

$$\mathbf{l}_{x+n} = {}_n\mathbf{M}_{xn}\mathbf{S}_x\mathbf{l}_x \qquad (4)$$

where $\mathbf{l}_x$ and $\mathbf{l}_{x+n}$ denote vectors giving the distribution of individuals among categories at durations x and x+n, respectively; ${}_n\mathbf{M}_x$ is the matrix of redistribution proportions p_{ij}; and ${}_n\mathbf{S}_x$ is the diagonal matrix containing the value s_j in the j-th position. Formula (4) may be regarded as a generalization of the life table relation $l_{x+n} = l_x(1-{}_nq_x)$. The entries of the diagonal matrix $\mathbf{S}_j$ are $1-{}_nq_x$ values for the individuals in each subpopulation at exact duration x. Analogues of the other life table relations given in §9 are readily derived.

The definition of exposure and expectation statistics for partioned populations is simple enough in principle. We define mean total time spent in any given state, for example, mean time spent continuously in a state, condition these means on the attainment of a particular duration x, and so on. Calculation of such statistics would be simple if we knew the exact duration at which every individual experienced every transition between categories. This detail is frequently unavailable, however, at least in demographic applications, and various complications and perplexities arise in estimation from grouped data. For further development, including the introduction of discrete and continuous time Markov chain models, see Land and Rogers (1982) and Schoen (1988).

An equation identical to (3) applies when cohort experience is disaggregated by time periods rather than by duration groups. In this case the entries in the diagonal matrix represent retention ratios and the entries in the redistribution matrix represent change in the composition of the cohort during a particular time period rather than over a particular duration interval. These statistics are used in connection with population projection, discussed further below.

A particularly ingenious and important application of “multistate” life tables, as they are sometimes called, is given by Bongaarts (1987), who uses them to derive distributions of women by number of surviving children under given schedules of fertility and mortality. Classifying women by number of surviving children is easy enough, but tracing changes in the number of surviving children requires the introduction of mortality. Since mortality in childhood varies sharply by age, we need to know not only the total number but the distribution of the surviving children by age, and this quickly leads to an unmanageably large number of categories. Bongaarts ingenious device of `partial family status life tables' involves classifying women by the number of surviving children they will have at a particular future age and computing one such table for every age to obtain the desired result (1987:192-193).

MODELS

A “model” is any construct—verbal, mathematical, computation-al, or physical—that may be manipulated so as to resemble reality. This generality compounded with that of the population concept makes any sort of review of `models in population analysis' quite out of the question. The areas discussed in the following sections have been chosen because they meet some combination of the following conditions: (i) they are important to demography; (ii) they are widely generalizable within and beyond demography; and (iii) they are difficult and more or less unsolved.

Population projection and synthetic survival functions, the first two models discussed below, use the elements of population structure to pose and answer various “what if” questions about numbers and rates. Though both are well known and widely used, their generality has not (I contend) been sufficiently appreciated. The life table probability model, discussed next, will seem terribly pedestrian to many groups of readers who have gone far beyond it. The rationale for including it is to differentiate clearly the model probability expressions from the corresponding observed statistics, a distinction that demographic terminology and usage tend to muddle.

Models for unobserved heterogeneity move us onto more difficult and less familiar territory. Important work has been accomplished, but we are still far from routine and confident application in data analysis. “Period Versus Cohort Statistics” takes up the general question of the impact on period statistics of shifts in the timing of events within cohorts. More progress has been made in this area, principally by Norman Ryder, than is generally known or applied in the field. The two-sex problem is perhaps the most basic area in which we have no satisfactory solutions, perhaps because the formal demographic apparatus of rates and decompositions that works so well in other areas is fundamentally inappropriate.[14]

§13 Population Projection. The essential idea of population projection is to ask what would happen to a population with a given initial distribution if it experienced prescribed rates of entry, exit, and redistribution in future periods. In the simplest case, we simply apply a growth rate observed in some base period to the ending population to obtain a first `projected' population, then to this first projected population to obtain a second projected population and so on. The so-called `component method' disaggregates the population by age and sex and introduces suitable schedules of rates to project the age-sex distribution forward period by period.

More generally, consider a closed, non-recurrent population observed during some time period. Proceeding as in §12 above we may write

$$\mathbf{P}_1' = \mathbf{MSP} \tag{1}$$

where **P** denotes the vector giving the distribution of individuals by category at the beginning of the period; **S** a diagonal matrix of retention proportions for each category; **M** a matrix of transition proportions between categories; and $\mathbf{P}_1'$ the vector giving the distribution by category of the individuals in the population at the end of the period who were present at the beginning of the period. Similarly,

$$\mathbf{P}_2' = \mathbf{M}_E\mathbf{S}_E\mathbf{E} \tag{2}$$

where **E** denotes the vector giving the distribution by category of the individuals entering the population during the period; $\mathbf{S}_E$ the diagonal matrix giving the corresponding retention proportions; $\mathbf{M}_E$ the transition matrix for movement between categories between entry and the end of the period; and $\mathbf{P}_2'$ the distribution by type of the individuals in the population at the end of the period who entered during the period. Finally, since $\mathbf{P}_1' + \mathbf{P}_2'$ gives the distribution by category of all individuals in the population at the end of the period, we have

$$\mathbf{P}' = \mathbf{MSP} + \mathbf{M}_E\mathbf{S}_E\mathbf{E} \tag{3}$$

an equation given in slightly different form in Feeney (1973).

Component projection consists of using equation (3) to generate a series of **projected** distributions for a population given (i) a distribution of the initial population by type, (ii) a distribution of entries by type for each successive time period, and (iii) the rate parameters contained in the matrices $\mathbf{S}$, $\mathbf{M}$, $\mathbf{S}_E$ and $\mathbf{M}_E$ for each successive time period.

When categories are age or duration groups with width equal to the length of the time period, (3) reduces to the familiar equations for component projection. In the classical case, entries are births and are computed by applying birth rates to the population age distributions. Non-classical cases include Boulding's study of the dynamics of populations of automobiles (1955) and the application given in Feeney (1985) to women with a given number of children ever born, both of which involve exogenously specified entries. Equation (3) incorporates a multitude of particular cases, many of which have been noted already in §12 above.

§14 Synthetic Survival Functions. Suppose that the information on life times required to construct a life table is incomplete because of the out-migration of a single individual at exact duration z. This out-migration confounds the calculation of life table statistics in several ways. We may compute l_x for $x<z$, and ${}_nd_x$ for $x+n<z$, as before, for example, but how should we proceed for $x>z$? We may compute ${}_nq_x$ as before if $x+n<z$ or $x>z$, but what if z lies between x and x+n?

We might simply drop the out-migrating individual from the calculation, reducing the open-to-migration to the closed-to-migration case, but this is often undesirable. That the individual in question remained in the population for z years without experiencing the exit event evidently provides useful information. Additionally, there are many situations, as in historical studies of village populations, in which discarding out-migrants would entail serious information loss.

An alternative approach that utilizes this information is to break the duration interval that includes z into the subintervals [x,z) and [z,x+n). For these intervals, the calculation of ${}_nq_x$ is unproblematic because the out-migration occurs at the boundary between them. Thus we compute ${}_{z-x}q_x$ as the total number of individuals present at exact duration x divided into the number of these individuals who make a proper exit during the interval [x,z), which excludes the out-migrant; and ${}_{x+n-z}q_z$ as the number of individuals present at duration z, excluding the out-migrant, divided into the number of individuals who make a proper exit during the interval [z,x+n).

Now consider the general case of a population open to migration. Whatever the series of duration intervals for which a life table is to be constructed, we refine the series so that every exact duration at which an in- or out-migration occurs represents an interval boundary. We then compute ${}_nq_x$ values for these intervals, including or excluding in- and out-migrants as appropriate. Finally, we use formulas (4) and then (1) of section 9 to compute l_x values and then ${}_nd_x$ values. Should the additional detail provided by the subintervals introduced into the calculation not be desired in the final result, it may be suppressed by deleting l_x values as desired and then computing ${}_nd_x$ and ${}_nq_x$ values for the remaining intervals from formulas (1) and (3) of section 9, respectively.

In calculating these synthetic l_x values we chain together ${}_nq_x$ values that may represent very different groups of individuals. This synthetic survival function tells us what would happen in a cohort of individuals that experienced, in successive duration intervals, the observed ${}_nq_x$ values. This may be a sensible question or not, according to context, and this is why we regard the synthetic survival function as a model rather than as merely another population statistic.

Synthetic l_x schedules may be invoked whenever a group of individuals come into observation after the occurrence of the relevant entry event and/or leave observation before the occurrence of the relevant exit event. This may come about either because information that we would like to have is missing, or because we choose to focus on certain life history segments. In studies of village populations, for example, we may exclude experience outside the village because we have no information about it, or because we want statistics that represent village conditions unadulterated by experience elsewhere.

A particularly fundamental application of synthetic survivorship schedules is the familiar **period life table**, defined by regarding every individual present at the beginning (end) of a specified time period as coming into (leaving) observation at this time. The familiar actuarial techniques for life table construction provide an approximation based on grouped data.

Multiple decrement (competing risk) theory provides another important application. The discovery of vaccination in the eighteenth century raised the question of what would happen to survival in a population in which smallpox was eliminated. The simplest way to answer this question is to calculate a synthetic survival function in which individuals dying of smallpox are treated as passing out of observation. Traditional actuarial techniques, based on the force of mortality function, reflect the necessity of dealing with grouped data (Jordan 1952: Chapter 15; Schoen 1988: Chapter 2). The formulation described here is noted in Tuma and Hannan (1984: Section 3.2.9). The same ideas apply in any situation in which exits from a population are classified into two or more mutually exclusive and exhaustive categories and we want to know how the survival function would change if a particular category of exits were eliminated. We may ask what proportion of individuals would ever marry in the absence of mortality, for example, or how long marriages would last in the absence of divorce.

The multiple decrement application illustrates how far beyond literal description the synthetic survival calculation may go. The hypothetical elimination of a particular cause of death or exit assumes implicitly that this is not inconsistent with leaving the risks for remaining causes unchanged. Schoen calls this `the vitality assumption' (1988: Section 2.3.1). This will be plausible in some circumstances and not in others.

§15 Probability Models. Much of what we see in small populations may best be attributed to natural random fluctuation. Probability models and statistical theory help decide what should be dismissed in this way and what should be subject to further interpretation. Traditional demographic analysis paid little attention to probability models and statistical estimation because population numbers were usually so large as to render random fluctuations negligible, either in absolute terms or in relation to such systematic errors as age heaping or differential under-enumeration by age. When dealing with small populations or samples, however, random fluctuation should not be ignored. The following discussion is restricted to the simplest probability model for life tables and shows how this model relates to the life table statistics already defined.

Suppose we calculate a complete data life table for a birth cohort in a small human population. Though the statistics in this table do literally and (we may suppose) accurately describe what happened to the individuals in the cohort, we will probably be unwilling to interpret the absence of deaths in any particular small age interval as indicating that the risk of death was somehow suspended during this interval. We are more likely to think that the risk of mortality varies smoothly with age, that the precise ages at which particular deaths occurred was a matter of chance, and that we should try to eliminate this chance element in assessing the observed life table statistics. This is in fact part of what we usually understand by `mortality' as distinguished from `death'.

These considerations lead us to impose probability models and statistical estimation procedures on the data. In doing so we refuse to take the data at face value, beyond a certain point, and impose upon it independently formed conceptions of what it represents. We exercise volition and judgement in deciding to treat the data in this way because we decide that this imposition is appropriate.

The simplest probability model for the life table assumes that life-times are independent, identically distributed random variables. The probability that any individual will remain at duration x may then be defined by s(x), where s is a monotonically declining function defined on the interval $[0,\infty)$ with $s(0)=1$. Remaining at duration x is equivalent to exiting in the interval $[x,\infty)$. Since this interval is the disjoint union of $[x,x+n)$ and $[x+n,\infty)$, we have

$$s(x) = d[x,x+n) + s(x+n) \tag{1}$$

where d[x,x+n) denotes the probability that any individual exits in the duration interval [x,x+n). Consequently

$$d[x,x+n) = s(x) - s(x+n) \tag{2}$$

for any $x\geq 0$ and $n>0$. If the occurrence of the exit event is certain, s(x) and d[0,x) go to zero and one, respectively, as x goes to infinity.

The conditional probability that the exit event will occur during the interval [x,x+n), given that it has not occurred by duration x, will be denoted by q[x,x+n). This conditional probability is by definition the probability of exit in the interval $[x,x+n)\cap[x,\infty) = [x,x+n)$ divided by the probability of exit in the interval $[x,\infty)$, or

$$q[x,x+n) = d[x,x+n)/s(x) \tag{3}$$

for any $x\geq 0$ and $n>0$.

These formulas are analogous to those of the life tables section above, but their terms are probabilities rather than observed proportions. Observed life table statistics may be regarded as estimators of the corresponding probabilities under the probability model just defined. There is an extensive and diverse literature on statistical estimation theory for life table data, most of it outside demography. Synthetic survival functions for complete data were originated by Kaplan and Meier (1958:461) and are referred to as `product-limit' or `Kaplan-Meier' estimates of s(x). Chiang (1968) gives an early and readable account emphasizing the use of grouped data. Cox (1972) originated the statistical analysis of covariates into life table analysis. Subsequent work includes the books of Elandt-Johnson and Johnson (1980), Kalbfleisch and Prentice (1980), Lee (1980), Miller (1981), Lawless (1982), Cox and Oakes (1984), and Mode (1985). There is also a specialized literature on reliability testing (Gertsbak and Kordonsky 1969; Barlow and Proschan 1975; Nelson 1982; Martz and Waller 1982).

Given the probability model defined by some function $s(x)$, the limit

$$\lim_{\Delta\to 0} \frac{s(x) - s(x+\Delta)}{s(x)\Delta} \tag{4}$$

reduces, assuming s differentiable, to $-s'(x)/s(x)$. This is the **risk function** corresponding to the function s, known in the classical case as the **force of mortality function**. Its values are most simply interpreted by observing that they approximate ${}_nq_x/\Delta$ for small Δ. The risk function is a continuous, probabilistic analogue of the ${}_nq_x$ values of the life table. Early actuarial treatment of the force of mortality function works from a conceptually <u>ad hoc</u> smoothing of observed l_x schedules (Smith 1948).

Given an integrable function h, we may ask for the function of which h is the risk function, that is, for the function s satisfying

$$-s'(x)/s(x) = h(x) \tag{5}$$

Solving this differential equation with the boundary condition $s(0)=1$ gives

$$s(x) = \exp\{\int_0^x h(z)dz\} \tag{6}$$

The risk function is readily generalized to different causes of exit, and the risk all causes combined equals the sum of the risks for the various individual causes,

$$\mu(x) = \mu_1(x) + \mu_2(x) + \ldots + \mu_n(x) \tag{7}$$

$\mu_i(x)$ denoting the risk of exit at exact duration x for the i-th cause of exit. The survival function that would result from eliminating a particular cause of exit is obtained by putting the corresponding risk function to zero and applying the exponential formula (6). This is the traditional actuarial approach to the multiple decrements (Schoen 1988: Chapter 2).

<u>§16 Unobserved Heterogeneity</u>. The simple probability model described above makes explicit an assumption, frequently made implicitly in working with life table statistics, that a statistic computed for the aggregate may be interpreted as indicating the

level of risk experienced by an individual. When the individuals represented in a life table experience the same (different) risks of experiencing the relevant exit event we say that the group is **homogeneous** (**heterogeneous**) with respect to this risk. Heterogeneity presumably exists to some degree in every situation, so the relevant questions are its extent and consequences.[15]

There are various situations in which it is natural to suspect heterogeneity without having observed it directly. `Unobserved' heterogeneity is necessarily an hypothesis, but it may be a compelling one. Leridon (1977:18) presents ${}_nq_x$ values for progression from marriage to first subsequent birth that peak around the ninth month following marriage and then decline by about one half over the following year. The rapidity of the decline appears to rule out both declining physiological ability to bear children and declining sexual activity as explanations. A more plausible explanation is that women are heterogeneous with respect to the risk of having a birth. The tendency for higher risk women to experience a birth before lower risk women would lead to an increasing concentration of low risk women at higher durations. This in turn would result in rates for the group as a whole that decline with increasing duration.[16]

Ideally, we would follow up such a speculation by securing information on the characteristics of women that result in their having higher or lower risks of having a birth. The group could then be disaggregated by these characteristics and a life table calculated for each of several subgroups.[17] In this way we would hope to establish the fact and distribution of heterogeneity, show that risks for women in any homogeneous subgroup are indeed relatively constant in the several years following marriage, and establish the extent to which the duration pattern of birth rates for the group as a whole is due to the progressive removal of higher risk women.

This direct proceeding is often impossible. In the example just given, a woman's risk of birth appears to depend on behavioral and physiological characteristics that cannot be adequately ascertained in most demographic surveys. More generally, we may face the necessity of analyzing whatever data are at hand, without the luxury of being able to go out and collect more. We may collect additional data, finally, only to find that it does not capture the heterogeneity we expected or hoped it would.

Faced with an impasse of this sort we may assume that the individuals represented in a life table are distributed according to some distribution $f(z)$ with respect to an attribute Z that determines survival probabilities $s_z(x)$, so that the survival function for an individual chosen at random from the group is

$$s(x) = \int s_z(x)f(z)dz \qquad (1)$$

We will not be able to make any inferences about the distributions on the right without imposing further structure, but by parameterizing either or both terms we may hope to estimate both the distribution of heterogeneity and the survival function for individuals with any given value z of the attribute Z.

A peculiarity of this proceeding is that nothing on the right hand side of (1) is ever observed directly, neither the attribute Z, nor its distribution f(z), nor the retention functions $s_z(\cdot)$. Obviously we will be able to rule out some combinations of distributions on the grounds that (1) does not correspond sufficiently well to an observed survival function. Good fits may in general result from many different combinations, however, whence observed survival functions will not usually help us discriminate possibilities. Further evidence and argument will be required to go beyond the weak conclusion that the stipulated distributions might, but need not, represent what is going on in the data.

Vaupel, Manton and Stallard (1979) present a model for mortality in which `frailty' is gamma distributed and individuals experience mortality risks that differ by a constant factor. Choosing plausible parameter values, they provide several useful examples of what Vaupel and Yashin (1985) call the `deviant dynamics' of death in heterogeneous populations. They conclude provocatively, suggesting that standard life table calculations over-estimate life expectancy by failing to take account of heterogeneity. A subsequent paper by the same authors fits specific models to data for the United States and Sweden to ascertain what the extent of heterogeneity would be if these models are correct (Manton, Stallard and Vaupel 1981). While these results suggest that heterogeneity is important to the study of mortality, they also show that estimates of its extent are worryingly non-robust against alternative parameterizations of the force of mortality.

Vaupel (1988) shows how useful heterogeneity models may be despite the difficulty of establishing the empirical validity of a particular model. The relatively weak correlation between life spans of parents and children is puzzling because one expects a common influence of both genetic and environmental factors. Using the model of Vaupel, Manton and Stallard (1979), Vaupel shows that "even in the extreme case in which sons (or daughters) perfectly inherit their fixed frailty from their fathers (or mothers), ... [the correlation] between fathers' and sons' life spans will be close to zero" (Vaupel 1988:278). A similar result for waiting times to conception is given by Sheps and Menken (1973:73-74).

Heterogeneity in the examples just given is intrinsically interesting, but there is a literature in which it arises more as a nuisance parameter in the estimation of covariate effects on life-times. Heckman and Singer's finding that radically different estimates of covariate effects may result from different parametric assumptions about the distribution of heterogeneity (Heckman and Singer 1982: Section 3) led to their non-parametric procedure for handling heterogeneity (Heckman and Singer 1984). Trussell and Richards undercut this method to some extent by showing that it can be sensitive to the parametric form assumed for the risk function (1985).

§17 Period Versus Cohort Statistics. Wars, economic down-turns, or other circumstances may lead to births being planned for or avoided in particular years. Period fertility statistics fluctuate accordingly, but cohort statistics are effected only if the total number of children born to the cohort is effected.[18] Recognition of such effects and of their importance for the analysis of fertility trends goes back at least to Hajnal (1947).

A sustained shift to earlier ages at childbearing exerts an upward influence on period fertility measures because the same number of births tends to be compressed into a shorter time period. Similarly, a sustained shift to later ages at childbearing stretches out the same number of births over a longer time period and so pushes period fertility measures down.[19] Our understanding of this phenomenon is due almost entirely to the fundamental work of Norman B. Ryder. The potential magnitude of the effect is indicated by Ryder's translation model, first proposed in (1964a) and further developed and applied in Ryder (1980; 1983). See also Ryder (1951/1980; 1982) and Foster (1990). The name derives from the simplest case assumption that distributions of events within cohorts are translated back and forth on the age or duration axis. The simplest result states that, under suitable assumptions, the period total fertility rate exceeds or falls short of the cohort total fertility rate by a factor of 1-m, where m denotes the annual rate of change of the mean age at marriage within cohorts (Ryder 1982:741).

The distortions in period fertility indicators that may be occasioned by changes in the timing of childbearing constitute a powerful argument in favor of cohort statistics in the analysis of fertility. Cohort statistics become available only after the history of the cohort is complete, however, which means that they represent primarily the experience of several decades past. There is no alternative to period statistics if we are want to analyze recent trends. At the same time, an understanding of the relations between period and cohort fertility statistics is essential to the analysis of trends in period statistics. An observed trend in period fertility may be generated by a movement toward smaller families, meaning a decline in family size for cohorts, but it may also be generated simply by a movement toward later ages at childbearing with no change in family size.

This tension between cohort and period fertility statistics generalizes widely. Most obviously, a tendency to shorter or longer birth intervals will effect various period fertility statistics in the same way as similar trends in overall age at childbearing. To take a rather different example, suppose a change in divorce laws reduces the time between filing for and receiving a divorce. This is likely to result in a sudden but temporary surge in divorces, as earlier filings under the new law pile up on top of delayed filings under the old law. The essential statistic here is evidently the proportion of filings that result in divorce. It would obviously be fallacious to read into such a rise in divorce rates any tendency to increased marital breakup; the rise and fall in period numbers and rates of divorce is essentially spurious.

Consider finally the analysis of delayed registration of vital events. The relevant cohort statistic here is the proportion of events occurring in any given period that are eventually registered. The period statistic is the total number of events registered in any given period divided by the total number of events that occurred in this period. In the absence of delayed registration, the latter statistic represents completeness of registration and will nearly always be less than one. Delayed registration creates a more complex situation. A decline in intervals between occurrence and registration, for example, will cause registrations of events occurring in several prior years to `pile up', as in the divorce example. This will lead to an apparent increase in registration completeness and may lead registrations exceeding occurrences. A period-cohort analysis of marriage registration statistics in Japan shows a remarkable cancellation (Feeney and Saito 1985: 4-11). Cohort completeness of registration was around 50 percent for marriages that occurred in 1950 and rose to about 90 percent for marriages that occurred in 1980. Despite this increase, the period ratio of marriage registrations to marriages remained very close to one during these years. The increase in cohort ever-registration was counteracted by a concurrent decline in the delay between the occurrence and the registration of marriages.

§18 The Two-Sex Problem. Rates of all kinds embody the idea that the incidence of events depends first of all on the corresponding exposure to risk, that whatever other influences there may be should be considered only after exposure has been accounted for by calculating a suitable rate. Because marriage requires two partners, however, we ought to consider the sex composition as well as the number of marriageable persons in assessing the incidence of marriage. How to do this is, very briefly, the two-sex problem. Though clearly recognized as long ago as Knibb's remarkable Appendix to the 1911 Census of Australia (1917), it has thus far defied effective solution. See Schoen (1988: Part III) for a recent discussion and references. The following draws primarily on Feeney (1972b).

Consider crude marriage rates in a population of single persons of both sexes. These rates take no account of sex composition, just as crude death rates (say) take no account of age, so we attempt to refine them. Crude death rates are refined by disaggregating deaths and exposure by age groups and calculating a rate specific to each age group. The analogous operation in the case of crude marriage rates evidently involves computing a schedule of rates corresponding to different sex compositions of the single population, sex composition being represented by the proportion of males or females. Age is an attribute of individuals, however, whereas sex composition is an attribute of the population. We can calculate age-specific death rates directly from population data for any given period, but we cannot follow suit for a schedule of marriage rates because the data available for any one period represents a single sex composition.

A schedule of marriage rates by sex composition, unlike a schedule of age-specific death rates, cannot be calculated directly from any observable data. To make inferences about it we will need to observe marriage rates for numerous populations and/or time periods, with a suitable range of sex compositions, and making suitable controls for other influences on marriage. While we need not be discouraged by this prospect, which is after all a more or less standard paradigm for causal inference in social science, it is different in kind from the direct observation and calculation that characterizes most demographic analysis.

Attempts to determine empirically how marriage rates respond to changing sex composition necessarily begin by considering what might be the shape of this schedule of marriage rates in relation to sex composition. The rate will certainly be zero in the absence of either males and females, and we will probably be willing to assume that it will rise smoothly from zero at either of these extremes to a maximum value somewhere near equal numbers of marriageable males and females. If the maximum value is low, less than 10 marriages per thousand single population (say) we might be willing to assume that the schedule will be approximately flat so long as the imbalance between the sexes is not too extreme, with at least 20 percent (say) of the sex in shorter supply. For this range of sex imbalances, then, the schedule reduces precisely to the crude marriage rate for both sexes combined.

This is evidently the simplest possible model, and it is this feature more than any other that recommends it to our attention. Though its utility depends on the empirical conditions being met, it is likely that they will be met in at least some applications.

It might appear that the model achieves nothing of interest because we have assumed away any effect of sex imbalance on the marriage rate, but this is true only of the rate for both sexes combined. The marriage rates for males and females implied by the model vary in the expected way, with the sex in shorter supply experiencing a higher rate and the sex in greater supply a lower rate. Interestingly, the effect results not from changes in the number of marriages in the numerator, but from changes in numbers of marriageable males and females in the denominator.

Empirical application even in this simplest case poses both a practical and conceptual problem. The principle of exposure to risk dictates that we should define as `marriageable' only those persons actually exposed to the risk of marriage. At the very least, we will want to exclude persons below some sensibly selected minimum age. It will not do to set this age too high, because marriages of persons in the middle teens do occur with some frequency in many populations. If we set the minimum low enough to incorporate these cases, however, we will find that the numbers of marriageable males and females are dominated by large numbers of very young persons, many of whom are probably not in fact very much exposed to the risk of marriage. This happens because the number of single persons at older ages decreases sharply because large proportions of these persons have already married.

The general approach to this problem is evidently to down-weight younger persons in some way that reflects their lessor exposure to risk. A simple empirical approach, one that would probably be justifiable in, for example, East Asian populations, would be to exclude persons in school from the marriageable population. Alternatively, one might construct model weights of some sort, specific for age and sex, to be applied to each age-sex group of the marriageable population. In this case we would have not only to define the weights, but to consider whether, and if so how, they should change over time. A more elaborate implementation of the same general idea would disaggregate the marriageable population into persons in and not in a `marriage pool', using an estimated age-specific rates of movement of into the pool. The marriage pool idea was introduced in Coale (1971) and modeled in Feeney (1972a; see also Coale and McNeil 1972).

We have thus far been considering the effect of the sex distribution of marriageable persons on numbers of marriages, age entering in only to identify appropriately how many persons ought to be taken as `marriageable'. What of the effect of the age distribution of marriageable men and women on numbers of marriages? If men and women were indifferent to the ages of prospective spouses, we might argue that these age distributions will not influence the total number of marriages, though of course they will effect the distribution of these marriages by age of bride and age of groom.[20] For such age preferences to effect numbers of marriages, they must be sufficiently strong that men and women fail to marry in significant numbers if they cannot find spouses at suitable ages.

This brings us to the issue of the strength of preferences for marriage partners of given ages. Consider first the case in which preferences, though perhaps present, are not strong enough to result in non-marriage: men and women prefer spouses of certain ages, perhaps, but settle for what they can get in this regard. Numbers of marriages are determined by numbers of marriageable men and women and are not influenced by the corresponding age distributions. The distribution of these marriages by age of bride and/or groom will of course depend on the age distributions, of course, but an important element of simplification is effected. We may for example consider whether sex imbalances among marriageable persons are leading to a `marriage squeeze' without worrying about age distribution except to define the pool of marriageable persons.

Alternatively, suppose that every individual insists on a spouse in a particular age group: if a spouse of the preferred age is unavailable, they don't marry. In this case it seems reasonable to suppose that numbers of marriages of men aged x and women aged y will depend only on the numbers of marriageable men and women at these ages. The rationale is that numbers of marriageable men and women at other ages will influence the number of marriages of men aged x and women aged y only if men at other ages compete with men aged x for women aged y, and women at other ages with women aged y for husbands aged x. If preferences are rigid, there will be no such competition. In this case total marriages are obtained as the sum of marriages of men aged x and women aged y, which in turn depend on the numbers of marriageable men and women at these ages. The functional relationship between for marriages of men aged x and women aged y may be handled in the same way as that between all marriageable men and women. Here also simplification is effected, though not so radically as in the first case.

Given the difficulty of dealing effectively with the relatively simple extreme cases, it might be better to opt for one or the other and proceed to empirical exploration before attempting a more realistic model of slightly inflexible but not wholly rigid age preferences. As between the two, the flexible preferences model is evidently the more plausible. People may be rigid about certain characteristics of prospective spouses, but age is not likely to be one of them. While tables of marriages by age of bride and age of groom show a clear correlation, they also show a very considerable dispersion. Fortuitously, the no-age-preference model is considerably simpler because we need only a single set of parameters defining the relation between marriages and marriageable men and women. The rigid preference model requires one set of parameters for every combination of marriageable ages.

CONCLUSION

We set out to explore the limits of generality of formal demographic concepts. What have we learned? The set-theoretic definition of the population concept given in §1 allows the members of a population to be any kind of entities whatever, living or inanimate, tangible or abstract. This generality is so sweeping that we gain insight better by asking negative than positive questions. What is <u>not</u> a population? Where are the ideas of population analysis <u>not</u> likely to be useful?

For finite (but not continuous) populations we require a reasonable distinctness of individual members. This renders `clouds in the sky' and other such vaguely defined entities problematic. Populations that are mere sets hold no interest, so it is natural to restrict attention to populations that experience some turnover. On reflection, however, one concludes that this excludes God in monotheistic religions and little if anything else.

The generality embodied in population analysis is of a particular kind, focusing attention on individuals and their characteristics. In doing so it tends to draw attention away from such matters as relationships between individuals, groups such as families and households, and the environment the individuals inhabit. Demography has suffered from neglect of these areas, and it could be argued that this reflects a failure to appreciate sufficiently the way in which the highly abstract and very successful methods of population analysis can restrict intellectual vision.

Age and age-like variables play a fundamental role in population analysis. If age is unimportant, either because empirical variation by age is negligible or because age is <u>a prior</u> not meaningful, only the simplest results are likely to be of any use. The length of time a one has possessed a particular dollar bill has no economic significance, for example, whence we expect the significance of population analysis for monetary theory to be minimal. The perspective of central bank officials concerned with replenishing the physical supply of money is of course very different: the life-time of bills and coins is an essential parameter.

Perhaps the most striking finding of this exegesis of formal demography is the extent to which the fundamental concepts and tools can be developed without reference to any concept of reproduction. The set-theoretic characterization of the population concept was adopted because it seemed natural, and because a narrower definition would have excluded many applications. The extent to which familiar demographic concepts follow from these bare, set-theoretic bones is gratifying, but it is also surprising and disconcerting. Reproduction is arguably the most fundamental population concept of all, for without some way to generate new members, we have no population. Yet age-specific birth rates are virtually the only thing we cannot define without specific reference to birth in the biological sense.

How can it be that so much of demographic analysis may be developed without reference to reproduction? One answer is simply `that's how it is.' Most of the concepts and tools of demographic analysis are simply not `demographic' in any substantive sense of the word. They are pure population analysis, applied to human populations. On the other hand, many concepts we tend to associate with the study of mortality, such as life tables, Lexis diagrams, and stationary populations, turn out to be equally relevant to the study of fertility. The lack of any direct embodiment of reproduction in these concepts is thus somewhat misleading.

Another notable result is the absence of any reference to spatial distribution. Demography is generally regarded as including the study of migration and spatial distribution, and properly so. The distribution of human populations in space is arguably as fundamental and as demographic as their distribution by age. Population analysis is inherently aspatial, however, and demographic analysis, consisting largely of population analysis applied to human populations, has tended to follow suit. This does not mean that demography should exclude spatial analysis. It means that population analysis is insufficient for demography, that we need concepts and tools for spatial analysis in addition to those provided by population analysis.

The extreme generality of population analysis brings demography proper into sharper focus. If anything is clear from the preceding exposition, it should be that there is far more to demography, no matter how narrowly conceived, than formal population concepts. Demography is inconceivable without them, to be sure, but demography is defined by its extensive, subtle, and complex empirical content— from model life tables representing human mortality patterns to effects of breast-feeding on fertility— quite as much as it is by formal population concepts. Demography is knowledge of particular human societies concretely located in historical time and geographical space.

The division of demography into `formal demography' and `population studies,' going back at least to Lorimer (1959), is evidently unsatisfactory. Truly formal demography, being devoid of demographic content, dissolves into population analysis. Population studies, on the other hand, has always served more to blur distinctions (sometimes a very useful thing to do) than to make one. It covers far too much territory to sustain intellectual coherence, a point made by Ford and De Jong (1970:iii-iv) and doubtless by others.

Population analysis touches a wide range of practical affairs. Letters mailed but not yet delivered constitute a population. The effectiveness of the postal system is reflected to a large extent in the survivorship curves for this population. Members of Congress constitute a population. Concern has recently been expressed about survivorship in the form of high rates of re-election of incumbents. Case backlogs are a serious problem of the criminal justice system. They relate among other things to prison populations, in which sentences and parole policies define life spans and survivorship, and in which population pressure of a distinctly non-Malthusian kind is observed. Marketing makes the distinction between capturing and retaining a customer, and statistics of achievement in these areas are very different, reflecting what we know in demography as period and cohort indicators. The demographer is equipped with powerful tools for the analysis of many kinds of social phenomena not substantively demographic.[21]

Population analysis alone will not take us very far in any field, but it nonetheless provides invaluable services. The Lexis diagram enables us to think systematically and precisely about population data entirely detached from the idiosyncracies of particular data sources. The distinction between period and cohort statistics identifies two fundamentally different modes of aggregation of raw population data. Life tables, occurrence-exposure rates, and other population statistics suggest what we might want to look at in any given situation. We may then work back from the statistics desired to the data required to compute them. Models provide general approaches for making data and statistics address various kinds of fundamental questions.

Demography has a reputation in the social sciences of being atheoretical and data-bound. This is in no small part because the classical demographic data sources tend to face the demographer with a *fait accompli*. In demography as elsewhere, however, available data often has serious shortcomings, not simply of accuracy, but of conception and structure. A clear conception of what we would like to know is essential to deciding whether or not, and if so how, existing data can be made to address our concerns. It is equally essential to the successful creation of new data sources. Thinking about the data we would like to have is an essentially imaginative enterprise. The highly abstract concepts of population analysis are eminently practical tools for disciplined demographic imagination.

NOTES

[1] The set-theoretic development given in the following sections is essentially that given in Feeney (1968), a mimeograph that received limited circulation at the time.

[2] Here and throughout, bolded words and phrases are defined by the sentence in which they appear. A **relation** between two sets A and B is simply a subset of the set of all pairs (a,b), with a a member of the set A and b a member of the set B. The exposition in this section, the following section, and §6 utilizes a few basic concepts and notations from set theory that will be explained briefly in these end notes. For a systematic exposition see the first few chapters of Halmos (1960).

[3] The expression $\{u:uPt\}$ illustrates a general notation, rading in this case "the set of all individuals u that satisfy the condition uPt". The **power set** of a given set is the set of all subsets of this set.

[4] The expression $t \in P_S(u)$ illustrates another general set-theoretic notation, reading in this case "t is a member of the set $P_S(u)$".

[5] The notation $P_C(t) \subseteq Q_C(t)$, read "the set $P_C(t)$ is contained in the set $Q_C(t)$", means that every member of the set $P_C(t)$ is also a member of the set $Q_C(t)$.

[6] Compare Boulding's characterization of the population concept (1934:650), in which birth is identified with entry to, and death with exit from, a population. This leads to confusion when dealing with subpopulations, where entry and/or exit are often defined by events other than birth and death.

[7] If A and B are any sets, $A \cup B$ denotes the set that consists of all elements that are either in A or in B, and $A-B$ denotes the set that consists of all elements in A that are not also elements of B.

[8] In Lexis diagram terms, this corresponds to moving the initial point of the life-line segment representing each interval of membership to the time axis. Alternatively, we may think of moving from age or time since first entry to time since most recent entry.

[9] One occasionally encounters references to `accounting identities' in demography. It would be more appropriate to refer to `population identities' in accounting. Accounting involves populations to the extent that it involves inventories, but is primarily concerned with valuation, the assignment of money value to assets and liabilities.

[10] The word `statistic' here is used, as in the field of statistics, to mean `a function of the data'.

[11] There appear to be two reasons why expectations of life are not calculated using the familiar formula of elementary statistics. The early development of life tables by actuaries occurred long before the advent of mechanical calculators, and when calculating by hand it is advantageous to replace multiplication by addition. The actuarial calculation is also more efficient when calculating e_x for all values of x.

[12] The statistical literature sometimes refers to 'random samples' of life-times, which begs the question: random samples of what?

[13] Synthetic survival calculations, discussed below, may be used to compute synthetic ${}_nq_x$ values that are both time and duration exact.

[14] Stable population theory is perhaps the best known demographic model, and an important one. It has been treated so thoroughly in so many other places, however, that no purpose would be served by reviewing it once again here. See for example Fisher (1958: chapter 2), Keyfitz (1968), Coale (1972), Impagliazzo (1980), Song, Tuan and Yu (1985), and Caswell (1989). The generalizations of Preston and Coale (1982) and Arthur and Vaupel (1984) ought logically to be included here. They require the use of the continuous formalism, however, which deserves a systematic development, and their inclusion here would expand the enterprise beyond reasonable limits.

[15] It should perhaps be noted that heterogeneity as used in this context excludes individual variation in the time elapsed since the entry event. This is of course already accounted for in the life table statistics.

[16] For more on heterogeneity models in the study of fertility see Brass (1958) on completed parity distributions, Potter and Parker (1964) on the time required to conceive, and the books by Sheps and Menken (1973), Leridon (1977), and Mode (1985).

[17] Characteristics that change with duration may be accommodated by using synthetic survival schedules. If small numbers make disaggregation into subgroups impractical we may be able to introduce a statistical model that accomplishes the same purpose without disaggregation, e.g., the life table regression model of Cox (1972). This will of course be at the cost of some maintained assumptions about the nature of the heterogeneity.

[18] It is probably safe to say that an event cannot be postponed

without decreasing the likelihood of its eventual occurrence, if only because of mortality risks. Because confounding factors are invariably present, however, the significance of this purely qualitative observation depends on the magnitude of the effects involved. The suggestion is that it will in many cases be quantitatively insignificant.

[19] Because the effect operates on the numerators of birth rates, it is not controlled for by making them specific for marital status. The components of change calculations used to assess the contribution of age distribution, marital fertility, and marriage patterns to changes in crude birth rates do not take account of this effect.

[20] `Might' argue because the observed correlation could result from prospective marriage partners moving in social circles that tend to consist of persons of their own age. If this is the case, and if persons choose partners without regard to age, aggregate marriages by age of bride and age of groom will reflect a correlation between age of bride and age of groom. This idea is formalized in Henry (1972); see also Schoen (1988: Section 6.4.2). Since the social circles involved are unlikely to be readily observable, the idea of introducing models for unobserved heterogeneity naturally suggests itself, but there has been no work along these lines so far as I am aware.

[21] The role of the demographer in such matters will evidently be that of a statistical consultant. One could argue on logical grounds that population analysis is a branch of statistics. Statisticians have not to have done so, however, leaving the field to demographers.

REFERENCES

Arthur, W. Brian, and James W. Vaupel. 1984. "Some General Relationships in Population Dynamics. Population Index 50(3):214-226.

Barlow, Richard E., and Frank Proschan. 1975. Statistical Theory of Reliability and Life Testing. New York: Holt Rinehart and Winston, Inc.

Bongaarts, John. 1987. "The Projection of Family Composition over the Life Course with Family Status Life Tables." In Family Demography: Methods and Their Applications, edited by John Bongaarts, Thomas K. Burch, and Kenneth W. Wachter. Oxford: Clarendon Press.

Boulding, Kenneth. 1934. "The Application of the Pure Theory of Population Change to the Theory of Capital. Quarterly Journal of Economics 48:645-666.

------. 1955. "An Application of Population Analysis to the Automobile Population of the United States." Kyklos 2:109-124.

Brass, William. 1958. "The Distribution of Births in Human Populations." Population Studies 12(1):51-72.

Carey, James R. 1982. "Demography and Population Dynamics of the Mediterranean Fruit Fly. Ecological Modelling 16:125-150.

------. 1983. "Practical Application of the Stable Age Distribution: Analysis of a Tetranychid Mite (Acari: Tetranychidae) Population Outbreak. Environmental Modelling 12:10-18.

------. 1984. "Host-Specific Demographic Studies of the Mediterranean Fruit Fly Ceratitis capitata. Ecological Entomology 9:261-270.

------. 1985. Insect Demography: A Mathematical Framework. Lecture Notes for Entomology 205, Department of Entomology, University of California at Davis.

Caswell, Hal. 1989. Matrix Population Models. Sunderland, Massachusetts: Sinauer Associates, Inc.

Chiang, Chin Long. 1968. Introduction to Stochastic Processes in Biostatistics. New York: John Wiley & Sons.

Coale, Ansley J. 1971. "Age Patterns of Marriage." Population

Studies 25(2):193-214.

------. 1972. The Growth and Structure of Human Populations. Princeton: Princeton University Press.

------, and Donald R. McNeil. 1972. "The Distribution by Age of the Frequency of First Marriage in a Female Cohort." Journal of the American Statistical Association 67(340):743-749.

Cox, D. R. 1972. "Regression Models and Life Tables." Journal of the Royal Statistical Society, B34:187-220.

Cox, D. R., and D. Oakes. 1984. Analysis of Survival Data. London: Chapman and Hall.

Elandt-Johnson, Regina C., and Norman L. Johnson. 1980. Survival Models and Data Analysis. New York: John Wiley and Sons.

Enzer, Herman. 1964. "On a Useful Capital Growth Matrix." Econometrica 34(1):183-192.

Feeney, Griffith. 1968. "Abstract Populations." Mimeograph, Claremont, California.

------. 1970. "Stable Age by Region Distributions." Demography 7:341-348.

------. 1972a. "A Model for the Age Distribution of First Marriage." Paper No. 23. Honolulu: East-West Population Institute.

------. 1972b. "Marriage Rates and Population Growth: The Two-Sex Problem in Demography." Ph.D. dissertation, Department of Demography, University of California at Berkeley.

------. 1973. "Two Models for Multiregional Population Dynamics." Environment and Planning 5:31-43.

------. 1983. "Population Dynamics Based on Birth Intervals and Parity Progression." Population Studies 37(1):75-89.

------. 1985. "Parity Progression Projection." In International Population Conference, Florence 1985, vol. 4, International Union for the Scientific Study of Population.

------, and Yasuhiko Saito. 1985. "Progression to First Marriage in Japan: 1870-1980. NUPRI Research Paper Series No. 24. Tokyo: Nihon University Population Research Institute.

------, and Jingyuan Yu. 1987. "Period Parity Progression Ratio Measures of Fertility in China." Population Studies 41:77-102.

------, and Wolfgang Lutz. "Distributional Analysis of Period Fertility." In Future Demographic Trends in Europe and North America: What Can We Assume Today?, ed. Wolfgang Lutz. London: Academic Press.

Fisher, Ronald A. 1929/1958. The Genetical Theory of Natural Selection. First published 1929. Second revised edition 1958. New York: Dover Publications.

Ford, Thomas R., and Gordon F. De Jong, eds. 1970. Social Demography. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Foster, Andrew. 1990. "Cohort Analysis and Demographic Translation: A Comparative Study of Recent Trends in Age Specific Fertility Rates from Europe and North America." Population Studies 44(2):287-315.

Gertsbakh, I. B., and Kh. B. Kordonskiy. 1969. Models of Failure. New York: Springer-Verlag.

Hajnal, J. 1947. "The Analysis of Birth Statistics in the light of the recent International Recovery of the Birth-Rate." Population Studies 1(2):137-164.

Halmos, Paul R. 1950. Measure Theory. Princeton, New Jersey: D. Van Nostrand Company, Inc.

--. 1960. Naive Set Theory. Princeton: D. Van Nostrand.

Hartshorn, Gary S. 1975. "A Matrix Model of Tree Population Dynamics." Pages 41-51 in Tropical Ecological Systems: Trends in Terrestrial and Aquatic Research, edited by Frank. B. Golley and Ernesto Medina. Ecological Studies 11. New York: Springer-Verlag.

--. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration data." Econometrica 52(2): 271-320.

Heckman, James J., and Burton Singer. 1982. "Population Heterogeneity in Demographic Models" In Land and Rogers (1982), pp. 567-599.

------. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for

Duration Data." Econometrica 52(2):271-320.

Henry, Louis. 1953/1980. Fertility of Marriage: A New Method of Measurement. Population Studies Translation Series No. 3. New York: United Nations Economic and Social Commission for Asia and the Pacific, 1980.

------. 1972. "Nuptiality." Theoretical Population Biology 3:135-152.

Impagliazzo, John. 1980. Deterministic Aspects of Mathematical Demography. Berlin: Springer-Verlag.

Jordan, Chester Wallace. 1952. Life Contingencies. Chicago: Society of Actuaries.

Kalbfliesch, J. D. and R. L. Prentice. 1980. The Statistical Analysis of Failure Time Data. New York: John Wiley.

Kaplan, E. L., and Paul Meier. 1958. "Nonparametric Estimation from Incomplete Observations." Journal of the American Statistical Association 55:457-481.

Keyfitz, Nathan. 1968. Introduction to the Mathematics of Population. Reading, Massachusetts: Addison-Welsey Publishing Company.

Knibbs, George H. 1917. The Mathematical Theory of Population, of its Character and Fluctuations, and of the Factors which Influence them .... Appendix A of the Census of the Commonwealth of Australia, April 1911.

Land, Kenneth C., and Andrei Rogers, eds. 1982. Multidimensional Mathematical Demography. New York: Academic Press.

Lawless, Jerald F. 1982. Statistical Models and Methods for Lifetime Data. New York: John Wiley.

Lee, Elisa T. 1980. Statistical Methods for Survival Data Analysis. Belmont, California: Lifetime Learning Publications.

Lefkovitch, L. P. 1965. "The Study of Population Growth in Organisms Grouped by Stages." Biometrics 21:1-18.

Leridon, Henri. 1977. Human Fertility: The Basic Components. Chicago: The University of Chicago Press.

Lorimer, Frank. 1959. "The Development of Demography." In The Study of Population: An Inventory and Appraisal, edited by

Philip M. Hauser and Otis Dudley Duncan. Chicago: The University of Chicago Press.

Lotka, Alfred J. 1907. "Studies on the Mode of Growth of Material Aggregates." The American Journal of Science XXIV: 199-216.

Luther, Norman Y., Griffith Feeney, and Weimin Zhang. 1990. "One-Child Families of a Baby Boom? Evidence from China's 1987 One-per-Hundred Survey." Population Studies 44:341-357.

Lutz, Wolfgang. 1989. Distributional Aspects of Human Fertility: A Global Comparative Study. London: Academic Press.

Manton, Kenneth G., Eric Stallard, and James W. Vaupel. 1981. "Methods for Comparing the Mortality Experience of Heterogeneous Populations." Demography 18(3):389-410.

Martz, Harry F., and Ray A. Waller. 1982. Baysian Reliability Analysis. New York: John Wiley & Sons.

Matras, Judah. 1967. "Social Mobility and Social Structure: Some Insights from the Linear Model." American Sociological Review 32(4):608-614.

Miller, Rupert G. 1981. Survival Analysis. New York: John Wiley & Sons.

Mode, Charles J. 1985. Stochastic Processes in Demography and Their Computer Implementation. Berlin: Springer-Verlag.

Murphy, E. M. 1965. A Generalization of Stable Population Techniques. Doctoral dissertation, Department of Sociology, University of Chicago.

Nelson, Wayne. 1982. Applied Life Data Analysis. New York: John Wiley & Sons.

Oechsli, Frank. 1975. "The Parity and Nuptiality Problem in Demography." Theoretical Population Biology 7:229-245.

Potter, R. G., and M. P. Parker. 1964. "Predicting the Time Required to Conceive." Population Studies 18:99-116.

Preston, Samuel H., and Ansley J. Coale. 1982. "Age Structure, Growth, Attrition, and Accession: A New Synthesis." Population Index 8(2):217-25.

Rogers, Andrei. 1968. Matrix Analysis of Interregional Population Growth and Distribution. Berkeley, California: University of California Press.

------. 1975. Introduction to Multiregional Mathematical Demography. New York: John Wiley & Sons.

Ryder, Norman Burston. 1951/1980. The Cohort Approach: Essays in the Measurement of Temporal Variations in Demographic Behavior. Ph.D. Dissertation, Princeton University, 1951. Reprinted in 1980 by Arno Press, New York.

------. 1964a. "The Process of Demographic Translation." Demography 1:74-92.

------. 1964b. "Notes on the Concept of a Population." American Journal of Sociology LXXIX(5):447-463.

------. 1980. "Components of Temporal Variations in American Fertility." In R. W. Hiorns, ed., Demographic Patterns in Developed Societies. London: Taylor & Francis Ltd.

------. 1982. "Progressive Fertility Analysis." World Fertility Survey Technical Bulletin No. 8. Voorburg: International Statistical Institute.

------. 1983. "Cohort and Period Measures of Changing Fertility." In Rodolfo A. Buletao and Ronald D. Lee, eds., Determinants of Fertility in Developing Countries, Volume 2. New York: Academic Press.

Schoen, Robert. 1988. Modelling Multigroup Populations. New York: Plenum Press.

Sheps, Mindel C., and Jane A. Menken. 1973. Mathematical Models of Conception and Birth. Chicago: The University of Chicago Press.

Smith, F. C. 1948. "The Force of Mortality Function." American Mathematical Monthly LV(5):277-284.

Song, Jian, Chi-Hsien Tuan, and Jing-Yuan Yu. 1985. Population Control in China: Theory and Applications. New York: Praeger.

Trussel, James, and Toni Richards. 1985. "Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure." Pages 242-276 in Sociological Methodology 1985, ed. Nancy Brandon Tuma. San Francisco: Jossey-Bass.

Tuljapurkar, Shripad. 1983. "Transient Dynamics of Yeast Populations." Mathematical Biosciences 64:157-167.

Tuma, Nancy Brandon, and Michael T. Hannan. 1984. Social Dynamics: Models and Methods. Orlando: Academic Press.

Vaupel, James W. 1988. "Inherited Frailty and Longevity." Demography 25(2):227-287.

Vaupel, James W., Kenneth G. Manton, and Eric Stallard. 1979. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." Demography 16(3):439-454.

Vaupel, James W., and Anatoli I. Yashin. 1985. "The Deviant Dynamics of Death in Heterogeneous Populations." In Sociological Methodology 1985, edited by Nancy Brandon Tuma, pp. 179-211. San Francisco: Jossey-Bass.

Whelpton, Pascal K. 1954. Cohort Fertility: Native White Women in the United States. Port Washington, New York: Kennikat Press.