

Analysis of Life Histories – A State Space Approach

Fernando Rajulton
Population Studies Centre
University of Western Ontario
London, Ontario, Canada

Abstract

The computer package LIFEHIST written by the author is meant for analyzing life histories through a state space approach. Basic ideas on which the various programs have been built are described in this paper in a nonmathematical language. Users can use various programs for multistate analyses based on Markov and semi-Markov frameworks and sequences of transitions implied in life histories. The package is under constant revision, and programs for using a few specific models the author thinks will be useful for analyzing longitudinal data will be incorporated in the near future.

Résumé

Le système d'ordinateur LIFEHIST écrit par l'auteur est établi pour analyser des événements au cours de la vie par une approche qui tient compte des états au cours du temps. Les idées fondamentales à la base des divers programmes du module sont décrites dans un langage non-mathématique. Le système LIFEHIST peut être utilisé pour des analyses Markov et semi-Markov des séquences d'événements au cours de la vie. Le module est sous révision constante, et des programmes que l'auteur compte ajouter pour l'usage de données longitudinales sont décrits.

Key Words: Life events and transitions, multistate life tables, sequences, stochastic models.

Basic Ideas used in Life History Analysis

Events and Transitions

An event is usually defined as a *qualitative* change that occurs at a specific point in time. An event can also be defined more generally as a change that places an individual in a new status that is different from the previous status the individual was in before the change took place. This general definition of an event enables us to think of events as transitions between statuses.

As events are classified as recurrent or nonrecurrent (renewable or non-renewable), so too are states. A nonrecurrent state is entered only once and a recurrent state allows multiple entries. In addition, those states from which exits can occur are called *transient* states, and those states from which no exit can be made are called *absorbing* states. In analyzing life histories in a multistate framework (see below), one can combine recurrent and nonrecurrent states to examine transitions between several states that can be either transient or absorbing.

Sequence of Events and Conditional Probability

An individual's life can be briefly described by a specific sequence of events experienced by the individual from birth to death. Some events become milestones of one's life either by choice or by chance. Events that happen by choice often place certain restrictions on future events, these restrictions imposed by moral/religious codes and social norms. Each choice reduces the number of theoretically possible sequences that can happen at random. Marriage and parenthood, for example, involve moral (and supposedly irreversible) commitments by an individual. Such commitments eliminate, or at least restrain, the occurrence of certain other events. Life decisions therefore somehow affect the number as well as the sequence of events an individual can experience.

Although the occurrence of specific events eliminates the occurrence of certain other events, from a behavioural point of view, most events comprising a sequence are *probabilistic* rather than deterministic in nature. This is because while the probability of a given event depends on the sequence of those preceding it, it is not possible to predict at a particular moment exactly which event will follow. For example, divorce can occur only after marriage, but this does not mean that divorce will be the next in sequence after marriage; other events such as births and deaths of children as well as separation or spouse's death can intervene between these two events. Thus, a study of sequences of events contrasts the unconditional probability of an event's occurrence with the conditional probability of the same event given the sequence of prior events.

The concept of *conditional probability* therefore assumes a central place in an analysis of life histories since it is an essential tool for describing and analyzing sequences of events. The fundamental idea here is that a conditional probability reduces the *uncertainty* in predicting the occurrence of an event. To see this, let us consider a simple example. Suppose there are two recurrent events A and B. We observe a sequence of these two events as follows: ABAABBAB. Both A and B have their unconditional probabilities of occurrence as 0.5 (since both occur four times out of 8 observations). However, the conditional probability of B's occurrence given that A has just occurred is $3/4 = 0.75$. That is, $P(B/A)$ at lag one = $P(AB)/P(A) = 3/4$ where the numerator indicates the frequency of joint occurrence of A and B (number of AB pairs) that is divided by the frequency of A's occurrence. Thus, while the (unconditional) probability that B will occur at any time is only 0.5, its probability is increased to 0.75 when A precedes it. B is much more likely to occur if we know that A preceded it. This reduction of uncertainty by knowledge of prior events forms the basis of analysis of sequences.

Timing of Events

Each event takes place in a time-space continuum. An individual experiences an event at a specific age, at a specific duration since an earlier event, or at a specific calendar time. Age and duration serve as proxies for many social and cultural injunctions. The calendar time is often used to describe individuals who experience a specific event, called an *event origin*, at a specific time or period. The simplest (and yet the most powerful explanatory) calendar time is the year of birth of an individual.

These three aspects of timing of events, namely age, duration and calendar time, have been variously interpreted in the substantive literature on life histories. Age is simply considered as chronological age that tracks the biological process of growing old. Social time, which combines the double aspects of age and duration, marks the transition from one social role to another. And, historical time reflects the influence of a series of historical, political, economic and social events that are external to, yet act on, the individuals.

There is a close relationship between timing and sequence of events. Timing of most events examined in social research is contingent on other events preceding them and is also influential in explaining the events that will happen in the future. This is mainly because the timing of past events affects the length of "exposure" to experience a succeeding event.

This close relationship between timing and sequence of events in a life history leads to the basic premise on which a life history analysis should be developed. Both the sequence *and* timing of events are important components of an analysis of life histories. This premise implies that neither a sequence of events alone nor

the timing of events alone constitutes a life history analysis. A life ‘history’ involves both. Data that provide such information are called event-history or life-history data mainly collected either through a prospective or through a retrospective observation plan. With the complete or partially complete information on the sequence and timings of events experienced by an individual, it is possible to construct what is known as a *sample path*. A sample path traces the course of events and the different states an individual visits, with durations in each state. A set of life history data would therefore consist of sample paths of all individuals in a study. Along with the life histories, surveys often collect supplementary data on a number of covariates or socioeconomic characteristics of these individuals (either varying or not varying over time) that affect the timing and sequence of events under study.

Life History Analysis

A life history analysis involves statistical methods for examining all the three aspects of life history information, namely the order, sequence and timing of events (or transitions). As mentioned earlier, an event-history analysis true to its name should incorporate at least two of these three aspects. Otherwise, there is no history in it. Analyses that do not consider at least two aspects of life history information fall short of the ideal, however sophisticated the techniques of analysis may be. Thus, for example, one cannot claim to have done a life history analysis just by examining the timing of one specific event such as marital dissolution, data on which may have been retrieved from a life history survey. A study may analyze a single event and yet use some information on one or two preceding events as ‘covariates’ in a multivariate framework. The preceding events considered now as covariates (either as a dichotomous variable – whether the event occurred or not, or as a metric variable – the timing of these events) are assumed to have possible influence on the event being studied. According to the approach taken here, this is not a life history analysis, because neither a single event nor its timing makes up a life *history* in the absence of one of the other two aspects, namely sequence and order of events.

State Space

This paper is based on the premise that to examine both the sequence and timing of events, there is no better approach than the state space approach, which follows directly from the definition of an event as a transition between two states. A state space is a collection of all states considered for a specific analysis. The states are assumed to be discrete and are, for convenience, small in number. A sequence of events can be simply considered as shifts between successive states in the state space. For example, to analyze changes in marital status, a state space may consist of the following five states: Never Married, Cohabitation, First Marriage, Marriage Dissolution, and Second Marriage. In

this example, the states Never Married, First Marriage and Second Marriage are nonrecurrent by definition (because these states can be entered only once), while Cohabitation and Marriage Dissolution are recurrent (they can be entered more than once).

A state space can also be defined in such a way that it consists only of nonrecurrent states. In the above example, Cohabitation and Marriage Dissolution can be decomposed by the order of events such as First or Second Cohabitation. Or, a state space can be defined consisting only of recurrent states such as: Single, Married, and Cohabiting. In this definition, the Single state will include the never married, divorced, and separated. Figure 1 illustrates the direct transitions involved in a system of (a) nonrecurrent states only, (b) recurrent states only, and (c) a mix of recurrent and nonrecurrent states.

A state space will normally consist of a few transient states (that is, states from which exits are possible) and one or more absorbing states (from which no exits can be made). Depending on a specific research problem, a state that is transient in reality can be considered as an 'absorbing' state for the sake of analysis. Such situations arise when a study is not interested in what happens after this state has been reached. For example, if a researcher is interested in examining the transitions between marital states until the time of second marriage, then the state Second Marriage can be considered an absorbing state.

The above discussion makes it clear that *the first step in a life history analysis is to correctly specify the types of states in a state space*. This mainly depends on how complete or incomplete is the available information and on the scope and extent of a research problem. The way a state space is specified has critical consequences for the inferences that will be made later. An analysis of marital status changes done with three (recurrent) states such as Single, Married and Cohabiting – as in Figure 1b – would obviously lead to conclusions that are different from an elaborate analysis done with six (nonrecurrent) states of Never Married, First Cohabitation, First Marriage, First Separation, First Dissolution and Second Marriage – as in Figure 1a. The research problem, the hypotheses to be tested, the available techniques of analysis, and computer facilities and costs are critical elements in selecting the number and types of states in a state space.

Multistate Analysis of Event Dependencies

A fundamental assumption in life history analysis through the state space approach is that a specific stochastic process generates the observed events, and that the available data can be appropriately analyzed and tested for the validity of this assumption. This fundamental assumption implies that from an observed distribution, the underlying stochastic process or probabilistic laws governing the occurrence of events can be inferred. For example, to say that the number of

Figure 1a
Flow Diagram of Transitions in a System of Marital States

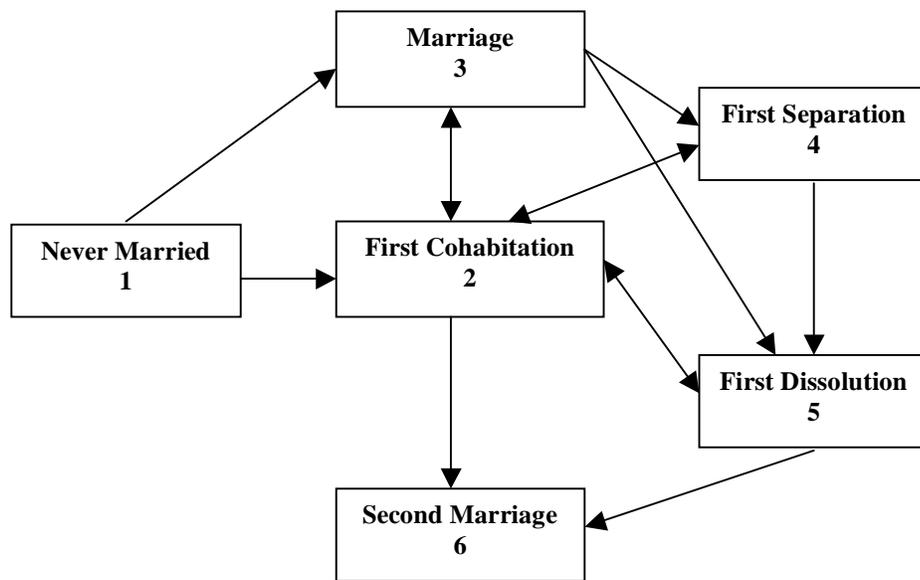


Figure 1b
Transitions in a System of Recurrent States Only

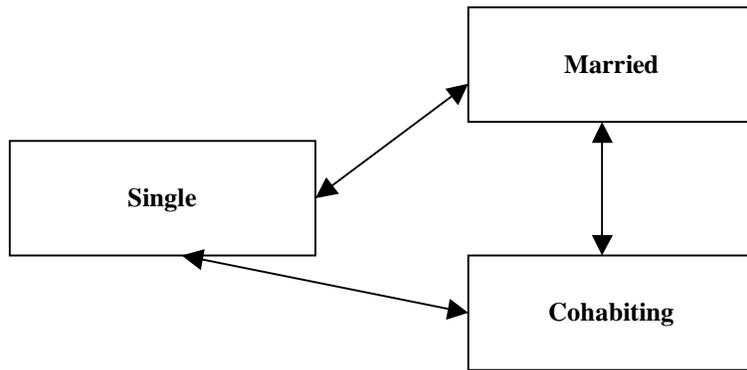
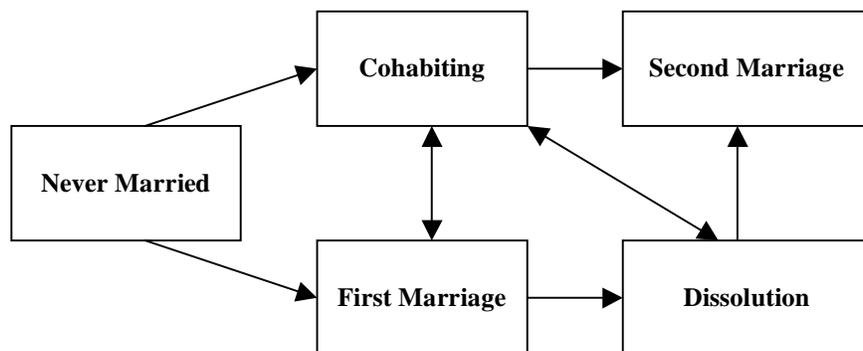


Figure 1c
Transitions in a System of Recurrent and Nonrecurrent States



events at time t depends on the number of events that already occurred implies a contagion or diffusion process. To say that events fall into different categories, each with the same process but with different parameters, points to the problem of heterogeneity. Or, to say that the parameters themselves are functions of time suggests a non-stationary process.

Since in many (social) processes, random variables at different times are mutually dependent, relationships among these random variables play a prominent role in stochastic analytic procedures. In fact, the theory of stochastic processes itself is nothing but a body of principles describing different types of relationships among random variables.

Once the state space has been clearly defined according to a selected research problem, the next step in analyzing life histories is to specify the type of dependency between the observed events (or equivalently, transitions between states). This section describes the concepts underlying the types of dependencies implied by Markov, Semi-Markov and Non-Markov processes. These three processes have been selected for two reasons: i) They use the information on the number, timing and sequence of events; and, ii) Their applications yield results which closely resemble life tables, the unifying framework in demographic analysis. Their applications however are not restricted to the field of demography.

Markov Model \Leftrightarrow Only the Preceding Event Matters

A stochastic process most commonly used to study the timing and sequence of events is the Markov process. This process implies a simple dependency of events:- *the occurrence of an event of interest depends directly on the occurrence of the preceding event, and only on it.* In state-space terminology adopted here, it means that a transition from one state (called the origin state) to another state (called the destination state) depends only on the origin state. A Markov process, therefore, ignores the manner in which the preceding event occurred or the manner in which the origin state was reached.

Although a simple dependency of events as implied by a Markov process rarely describes reality well, it has been applied to various social phenomena such as labour mobility (Blumen et al., 1955), change in attitudes (Coleman, 1964) and collective violence (Spilerman, 1970), and in more recent times to the development of the Multistate Demography (Rogers, 1975; Schoen and Land, 1979; Willekens and Rogers, 1978; Hoem and Jensen, 1982; Schoen, 1988).

A Markov model ignores past experiences and examines the effect of the most recent experience. This does not mean that the information on the sequence of events becomes useless if a Markov model is adopted in a life history analysis. The initial versions of multistate analyses mainly used only cross-sectional

information gathered through censuses and registration systems (that cannot provide any knowledge on the sequence of events). But in a multistate framework that can analyze life history data, the information on the sequence of events is preserved in the transitions from one state to another (in their recorded sequence) even though the probabilities of a specific transition depend only on the immediately preceding state. Thus, a Markov-based multistate analysis of life history data provides more realistic and more meaningful results than a Markov-based multistate analysis of cross-sectional data.

More generally, by ignoring the past in an observed sequence of events and by focussing on the last observed experience, a Markov model highlights the effects of timing or, more specifically, the age or duration at which a specific transition takes place. As mentioned earlier, these two aspects of timing capture the influence of social norms. In life history analysis, therefore, the Markovian assumption still serves as a powerful tool because social norms dictate how individuals ought to behave at different ages. If a group of individuals involved in a specific process are all of the same age at the start of the process, then for the first transition it would not matter whether age or duration is considered. For subsequent transitions, however, an analysis needs to differentiate between age and duration since individuals will differ in their timing as well as sequence of transitions. Since a consideration of duration aspect "relaxes" the strict Markovian condition in some respects (see the next section), I suggest using a Markov model mainly as a tool for examining the age-dependence of transition probabilities.

Semi-Markov model \Leftrightarrow Duration Makes a Difference

Of particular interest in analyzing life histories is the influence of duration in each state on making a transition – a duration dependence that is distinct from age dependence. Social norms dictate not only at what age certain behaviour is acceptable but also that certain events should occur after (or should not occur before) a certain length of time since another event. Births, remarriages, job promotions, educational qualifications, immigration and naturalization are all dependent on duration or the length of stay in an origin state.

In a semi-Markov model, changes in states depend on the state of origin (which is true also in a Markov model) as well as on the state of destination (unlike in a Markov model). Thus, the event dependency in a semi-Markovian framework is a modified version of a Markovian one: The occurrence of an event of interest depends *on both the preceding and succeeding events, and on the length of duration between the two events.*

Another distinct feature of a semi-Markov process is that it allows direct *self-transitions* or transitions to the same origin state (a Markov process does not allow self-transitions). Although some analysts view self-transitions to be

unrealistic, they cannot be dismissed outright. Admitting self-transitions depends on how a state space is defined. For example, if one defines a state space to consist of regions such as urban and rural in a migration study, then self-transitions are possible – an individual can move from one urban area to another. The same applies to a state space consisting of occupational statuses such as skilled and unskilled labour.

In analyzing life histories, besides duration in a state, the age at entry into the state can also influence the probability of transition. A minor variation regarding the effect of age at entry into a state can be introduced in a semi-Markov model. If an event does not depend on the age at which the previous event occurred, then the semi-Markov model is called a *homogeneous* or age-independent model. If it does, then the model is called a *nonhomogeneous* or age-dependent model (Mode, 1982, 1985). For example, in analyzing second births, one obviously considers the duration (t) since first births. In addition, it may be argued that age at first birth (x) also significantly influences the timing of second birth. An age-dependent semi-Markov model then would also incorporate age at first birth as one of its intrinsic dimensions; for example, its probability would be specified as $p(x,t)$, rather than simply as $p(t)$.

Whether dependent on age or not, a semi-Markov model ignores the number of events that have already occurred; that is, the Markovian condition is still valid.

Non-Markov model \Rightarrow Past History is Influential

The Markovian and semi-Markovian frameworks use the information on the timing (and sequence, though not explicitly) of events. Suppose that a study has the objective of finding out in what way a specific sequence of events influences a future event. Then, the third aspect of life history information, namely the *order* of events should be included in the analysis. The order of events makes the sequence aspect of life history information explicit, and a model that considers the history of events becomes non-Markovian.

Nobody doubts that history influences social or individual behaviour. No social system is memoryless, and no individual behaviour is free of repercussions of past behaviour. Neglecting the past in an analysis of life histories, therefore, can only produce an incomplete (if not biased) evaluation of social or individual behaviour.

The past becomes more relevant in analyzing life histories when it is recognized that in real situations, life histories simultaneously encompass two or more systems like educational and occupational careers, marital, maternal/paternal and professional statuses. Events occurring in one specific system are often dynamically dependent on events occurring in the same system or in another system. It is indeed rare that events occur sequentially and entirely within one

system alone. Moving from one parity to the next, for example, can depend on changes in marital and occupational statuses. Similarly, changes in levels of education and employment status can be mutually dependent. A realistic analysis therefore goes for a simultaneous consideration of two or more dependent subsystems rather than one system alone. Only a non-Markovian framework can handle such dependencies.

However, it is generally inconvenient to build models on non-Markovian lines. In practice, when attempts are made to include the past, a non-Markovian scheme is usually reduced to several Markovian or semi-Markovian schemes. The program for non-Markov analyses included in LIFEHIST makes use of the same algorithm used for a semi-Markov model but preserves the different sequences of events already experienced in computing the probability of experiencing a succeeding event.

Multivariate Analysis: Parametric Analysis of Socioeconomic Covariates

The procedures outlined in the last section have the purpose of examining the *nature* of a specific social or behavioural process by considering the order, sequence and timing of events experienced by a group of individuals. These procedures may¹ leave untouched the problem of heterogeneity existing in the group. Many empirical studies on life histories, however, are interested not so much in understanding the nature of a process as in examining whether a process differs among groups of individuals with different personal characteristics or exposed to varying contextual or institutional influences.

To examine simultaneously the influences of personal/institutional characteristics together with the dependencies existing among random variables leads to many practical problems. On the one hand, dealing with stochastic processes inevitably results in detailed analyses of timing and sequence of events. Generating processes for different heterogeneous groups results in small "numbers at risk" of experiencing a specific event and thus leads to erratic estimates of conditional probabilities. On the other hand, an (over)emphasis on the problem of heterogeneity and on examining the effects of individual and societal characteristics ignores the number-sequence aspects of life histories. It leads to complacency with the "up and down" interpretation of effects of several socioeconomic characteristics. Scientific enquiries, particularly those used for framing social policies, should ideally combine both types of analysis. "Policy suggestions" based exclusively on one type of analysis soon become meaningless and irrelevant.

To date, multivariate analyses applied to life history data are mostly based on dynamic modeling procedures, in which there is a function of timing that is adjusted by the effects of observed heterogeneous covariates (called often as

'parametric models'). However, the function of timing itself is ignored in many studies, since it is difficult to specify it in many research contexts. Since the time when the most popular proportional hazards model was introduced by Cox in statistical analysis, this neglect has unfortunately become a standard practice (although the results from the Cox's model can be used to estimate the underlying function). Somehow researchers have tended to look at the function of timing not only as something of a 'nuisance' but also as something that is irrelevant. The main advantage of doing first a dynamic analysis before going for examining heterogeneity is to discover the 'unknown' function of timing. This is the main purpose, as well as the contribution, of all the multistate programs included in LIFEHIST.

Further, to address the problem of heterogeneity, the programs included in LIFEHIST try to abide by the basic premise of life history analysis discussed earlier: Even while examining heterogeneity, at least two of the three aspects of life history information should be preserved in an analysis. This job is not an easy one. Some methods of doing a multivariate analysis based on this fundamental premise will be incorporated in the package in the future.

How to use LIFEHIST

Preparing the Data File

A data file for analyzing life histories through the state space approach consists of specific events selected for analysis, timings of these events, and the necessary information on censoring (in what state, and at what time) if experiences are curtailed at the time of survey. Since the retrieval of the necessary information on timings and sequences of events is a complex and difficult task, LIFEHIST requires as minimum effort as possible on the part of users in preparing the data file. Users need to provide only the *ages* at which the events under consideration occur. This will not be a difficult task since many surveys record the ages at occurrence of events. LIFEHIST processes this information for finding the sequence of events and the censoring times for different applications. All the programs in LIFEHIST are so general that they can be used for many phenomena for which life history data have been collected.

The data file should be a free-formatted ascii text file, with spaces between the ages at different transitions. Preparing the data file however should be done with some care, keeping in mind the research problem as well as the definition of the state space that flows from it. Users should take special care especially when many states are involved. The package has no restrictions on the number of states or the number of transitions; any limitations will be imposed only by the available memory on the PC. A small change introduced at a later stage for refining a state space may require a complete reorganization of the data. The

importance of correctly defining a state space before going ahead with any analysis should never be undermined, especially if an analysis includes some recurrent events. The following steps may be helpful in preparing the data file.

Step 1: Define the state space and identify the transient and absorbing states.

Step 2: Once the state space is properly defined, assign integer values to each state, first the transient states and then the absorbing states. For example if there are 6 states, of which 4 are transient and 2 are absorbing, then the transient states are assigned integers 1 through 4 and the absorbing states are assigned integers 5 and 6. Absorbing states should come together after transient states. The order within the two groups does not matter.

Step 3: Prepare the data file consisting of individual sample weight (if any), age at the time of survey, and ages at which all the events considered for analysis occur. If no sample weights are used for analysis, then age at the time of survey is given in the first column. If an individual does not experience an event, then assign a missing value for the corresponding age at occurrence. Missing values can be any large value beyond the maximum value taken by ages at transitions, such as 99 or 98.4, etc. In the Parameter Specification window (see below), the *lower limit* of these missing values should be specified so that any value larger than this would be considered as missing by the program.

Since an analysis of life histories through the state approach usually involves more than three states, the data file on timings of events will generally be huge with hundreds and thousands of data points. Users should prepare this data file through any available means at disposal such as data handling through SPSS or SAS. The number of columns in the data file depends on whether sampling weights are used and whether the states admit re-entries (that is, at least one state is a recurrent state). It is easy to prepare the file when the state space contains only nonrecurrent states; the ages when events occur are given in *the same order in which the states were assigned integer values* in Step 2. Extra effort will be needed to prepare the data file with one or more recurrent states (for lack of space, details are not given here). Once the file is ready, add a title on the first line to identify the data file.

LIFEHIST uses the information on age *at the time of survey*, along with the missing information, to find at what time (age or duration) an individual is (right) censored and in which state. At the moment, the package can deal only with right censoring. Some users of the package communicated to me the problem they face with left censoring in their data sets. I am planning to modify the programs for handling left censoring as well

A suggestion regarding the ages at transitions. LIFEHIST needs to know *only the ages* when events occur. Depending on the type of analysis, it transforms this variable into duration since entry into a previous state or duration since an earlier event or finds the sequences of events. It is important therefore that these ages are given with as much precision as possible. Some surveys record the year and the month when an event occurred, which though not very precise is good enough for many analyses. Other surveys however record only the year. This seriously jeopardizes a life history analysis, for example, while sequencing the events. If two events occur in the same year, the sequencing will be unsatisfactory and sometimes may be inaccurate unless extra care is taken in giving proper instructions to LIFEHIST about admissible transitions (see below). When ages at transitions are grouped or rounded, only discrete procedures can be used for analysis, provided no events are systematically omitted. Users should carefully check the sequencing output produced by LIFEHIST.

Table 1 presents an example of the data file prepared for a marital history analysis with six nonrecurrent states with the associated integers: 1-Never Married, 2-First Cohabitation, 3-First Marriage, 4-First Separation, 5-First Dissolution, and 6-Second Marriage. The relevant data are taken from the Canadian Fertility Survey (CFS) 1984. This survey has sampling weights for respondents from different provinces, and records most timings of events in years and months, excepting first cohabitation. The analysis starts either in the Never Married State at age 15 and ends with second marriage. Therefore, the state Second Marriage is considered absorbing. Table 1 gives the marital histories of seven women. The first woman, who has been assigned a weight of 0.84, is aged 38 at the time of survey. This is the woman's age *in completed years*. More precise age at the time of survey can be obtained since the CFS has recorded also the respondents' date of birth and the date of interview, both in year and month. Such precise information would be useful for finding the time of censoring. In the case of the seventh woman, for example, age at first separation and age at the time of survey are so close that censoring time will be poorly estimated.

The timings of various events (with 99.99 as missing values) are ordered in the same way as the states have been ordered in the state space [the titles for columns are given for illustration purpose only, and not to be included in the data file]. For the first woman, the age at first marriage has been recorded as 13.42, which is earlier than the assumed age (15) at the start of marital history in the Never Married state. If any such inconsistencies are observed in a data file, LIFEHIST will produce a warning and replace the observed age with the assumed age at the start of history.

Besides arranging the data, users need also give some instructions to the package as to how to handle the data. This set of instructions is given on a window titled Parameter Specification. Again, some care is needed in having these parameter values ready before starting to run the program. To change the

Table 1
An Example of Data File Containing Only Non-recurrent States

Weight	Age at Survey	Never Marr	First Cohab	First Marr	First Separ	First Divor	Second Marr
0.84	38.00	15.00	99.99	13.42	99.99	99.99	99.99
0.84	36.00	15.00	99.99	27.08	31.33	34.08	99.99
0.84	37.00	15.00	15.00	22.75	25.42	25.83	27.50
0.79	35.00	15.00	99.99	99.99	99.99	99.99	99.99
0.84	38.00	15.00	99.99	22.50	99.99	99.99	99.99
0.87	35.00	15.00	18.00	23.50	99.99	99.99	99.99
0.84	36.00	15.00	99.99	28.58	35.95	99.99	99.99

default values of these parameters, double click on each and make changes. The number of parameters varies from program to program, but for most of the multistate programs in the package, the following parameters are specified:

- whether weights are used (Yes/No)
- lower limit of missing values
- total number of states
- number of transient states
- any recurrent states (Yes/No; if yes, another parameter specification window will appear)
- total number of admissible transitions, and pairs of origin and destination states

We have already discussed about these except the last. 'Admissible transitions' means theoretically possible transitions among the states considered in the state space. Special attention should be paid to specify this parameter. Let us consider the example seen above, the six marital states, and use the integers associated with them. From State 1, two transitions are possible: 1-2 and 1-3. From State 2, four transitions are possible: 2-3, 2-4, 2-5 and 2-6. Similarly, from the other states, we have as possible transitions: 3-2, 3-4, 3-5, 4-2, 4-5, 4-6, 5-2, 5-6, thus to a total of 14 admissible transitions. Once this number 14 is given as the total

number of admissible transitions, a window will appear asking to specify these 14 pairs of origin-destination states. Then, users should give the values 1-2, 1-3, 2-3, etc. (without the hyphens). LIFEHIST uses these specifications to check for possible errors in the data on transitions. It may happen that some individuals do not give correct information on ages at transitions, and it is possible that the data show a direct transition from first marriage to second marriage without passing through divorce. LIFEHIST checks every recorded transition and only if it is admissible (as specified in this list) will consider it for analysis; otherwise, it will simply ignore it. Users therefore should check carefully before specifying these two parameters. At the same time, users also should feel free to experiment with this specification. For example, a direct transition from First Separation to Second Marriage is not theoretically feasible as specified above (4-6). But, it may happen, because of historical circumstances (before the Divorce Law came into effect), people were getting remarried without legal divorce. If there are serious inconsistencies between these specifications and the raw data, LIFEHIST will produce warnings and error messages in the output. One can expect a few warnings because no data file is perfect. If there are too many such warnings and error messages, however, users are strongly recommended to check the raw data file for these discrepancies.

For lack of space, detailed outputs from LIFEHIST for Markovian, semi-Markovian and nonMarkovian types of analysis are not shown in this paper. Readers can refer to some of the published papers or to the Manual (in preparation). I would strongly urge users of LIFEHIST to get first the descriptive outputs by running the relevant programs under the Descriptives Menu and check for observed multistate transitions for any possible defects in the data before going for multistate analysis proper.

Other than Multistate Analysis

As far as possible, all the programs in LIFEHIST follow the same procedures for preparing the data files. Some specific programs need much less effort on the part of users, some others may need more, depending on the type of analysis desired. Apart from the multistate analysis of life histories, LIFEHIST includes also a few other specific programs that are often used by researchers working with longitudinal data or that I think will be of great use in future research. My main aim of writing the package LIFEHIST is to offer researchers working with longitudinal data an opportunity to do those types of analysis that are not available in the commonly used statistical packages. It is not my intention to duplicate programs for analysis that can be done with the available software like SPSS, SAS, S-PLUS, STATA, etc.

However, mainly for the sake of Canadian researchers, I have also included a

few other programs in LIFEHIST such as those programs for constructing Single Decrement, Multiple Decrement Life Tables as well as for doing a Cox Proportional Hazards Analysis (for time-invariant and time-varying covariates). I say 'for Canadian researchers' because most Canadian survey data come with sample weights attached to each individual and these weights have to be used for proper analysis. Unfortunately, most statistical packages do not allow using fractional weights, and using discrete weights in whatever modified form in these packages makes the standard errors meaningless. There is a way of correcting the standard errors produced by these packages (see Zhao et al., 1995 for details) but most researchers may not be able to follow such procedures.

In these extra programs included in LIFEHIST, I have considered possible improvements to the existing procedures. For example, in constructing single decrement life tables, a common procedure in all the packages is to assume linearity or uniformity of events in a specific time interval. LIFEHIST includes in the output Chiang's coefficients for average lifetime that can show whether the assumption is valid or not for each interval, and uses these coefficients rather than the linearity assumption in constructing the life tables. Thus, the LIFEHIST output may differ from the outputs produced by other packages. Similarly, in the Cox proportional hazards model output, users will find some useful information on the rank of the timing vector and on ties. Few researchers nowadays have time to worry about all these finer details on the assumptions on which analytical techniques are based, without being aware that their inferences are not justified when these assumptions are not met. Working with longitudinal data calls for a serious attention to many such assumptions with which we do our research.

Besides these common types of analysis, LIFEHIST will include in the near future a few other specific variations like handling left censoring with longitudinal data. I am also planning to incorporate a few specific models like beta-logistic model for studying diffusion processes. Questions and suggestions from users are always welcome.

End Note

1. I say '*may* leave untouched', because I think, a nonMarkovian scheme introduces the heterogeneity *par excellence*, that is, the influence of the past transitions. This is based on the view that many characteristics, except the ones which are called background variables such as gender or race, that are usually incorporated in a study of heterogeneity are nothing but vestiges of early life course transitions.

References:

- Blumen, I., M. Kogan and P. J. McCarthy, 1955. *The Industrial Mobility of Labour as Probability Process*. Cornell Studies of Industry and Labour Relations, no.6. Ithaca, New York: Cornell University Press.
- Coleman, J. S. 1964. *Introduction to Mathematical Sociology*. New York: Free Press.
- Hoem, J. M. and U. Funck-Jensen, 1982. Multistate life table methodology: A probabilist critique. In K.C.Land and A.Rogers (eds.). *Multidimensional Mathematical Demography*. New York:Academic Press.
- Mode, C., 1982. "Increment-decrement life tables and semi-Markovian processes from a sample path perspective," in K.C. Land and A. Rogers (eds.). *Multidimensional Mathematical Demography*. New York: Academic Press.
- Mode, C., 1985. *Stochastic Processes in Demography and their Computer Implementation*. New York: Springer-Verlag.
- Rajulton, F. and T.R.Balakrishnan, 1990. "Interdependence of transitions among marital and parity states in Canada," *Canadian Studies in Population*, 17: 107-132.
- Ravanera, Z. R., F. Rajulton and T. K. Burch, 1994. "Tracing the life courses of Canadians," *Canadian Studies in Population*, 21: 21-34.
- Ravanera, Z. R., F. Rajulton and T. K. Burch, 1998. "Early life transitions of Canadian women: A cohort analysis of timing, sequences, and variations," *European Journal of Population*, 14:179-204.
- Rogers, A., 1975. *Introduction to Multiregional Mathematical Demography*. New York: Wiley.
- Schoen, R. and K. Land, 1979. "A general algorithm for estimating a Markov-generated increment-decrement life tables with applications to marital status patterns," *Journal of the American Statistical Association*, 74:761-776.
- Schoen, R., 1988. *Modeling Multigroup Populations*. NY:Plenum Press.
- Spilerman, S., 1970. "The causes of racial disturbances: A comparison of alternative explanations," *American Sociological Review*, 35: 627-649.

Analysis of Life Histories: A State Space Approach

Willekens, F. and A. Rogers, 1978. *Spatial Population Analysis, Methods and Computer Programs*. RR-78-18. Laxenburg, Austria: International Institute of Applied Systems Analysis.

Zhao, J., F. Rajulton and Z. R. Ravanera, 1995. "Leaving parental homes in Canada: Effects of family structure, gender and culture," *Canadian Journal of Sociology*, 20: 31-50.