

LIFEHIST : Analysis of Life Histories, a State Space Approach

Rajulton Fernando
Population Studies Centre
University of Western Ontario

1999

Paper presented at the Workshop on Longitudinal Research in Social Science: A Canadian Focus, Windermere Manor, London, Ontario, Canada, October 25-27, 1999.

1. 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 visualize events as **transitions** between statuses.

As events are classified as *recurrent* or *non-recurrent*, so also states are classified as recurrent and non-recurrent. A non-recurrent 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 analysing life histories in a multistate framework, one can combine recurrent and non-recurrent states to examine transitions between several states which can be either transient or absorbing.

2. Sequence of Events or Transitions

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 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 analysing 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 that 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, $P(B/A) \text{ at lag one} = P(AB_{+1})/P(A) = 3/4$ 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.

3. 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 highly contingent on other events preceding them and is also highly 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 which provide such information are called **event-history** or **life-history** data that are mainly collected either through a **prospective** or through a **retrospective** observation plan.

The collection of longitudinal data has recently become fashionable among researchers, and longitudinal research is touted as a panacea for establishing temporal order, measuring change, and making causal interpretations. Although there are many advantages associated with longitudinal data, there are offsetting costs and difficulties. First, it requires that information be collected over a time span long enough to encompass a detectable change in developmental status. Second, over long periods of time, attrition of the sample occurs, and it is often difficult to compare the initial and final groups to get a true picture of cause and effect relationships. Third, maintaining contact with and sustaining the motivation of the subjects is difficult and costly, even with a small group of 300 to 500. Fourth, the passage of time can lead to changes in the hypotheses and in the questions raised, and to unforeseen changes in both instrumentation and theory. In fact, as time goes on, most of the collected material is discarded without any systematic re-evaluation of the study in the light of its interim findings.

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** that traces the course of events and the different states an individual visits, along with durations in each state. A set of life history data

would therefore consist of all the 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.

4. 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.

Confusion may arise here because a study may analyse 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.

5. State Space

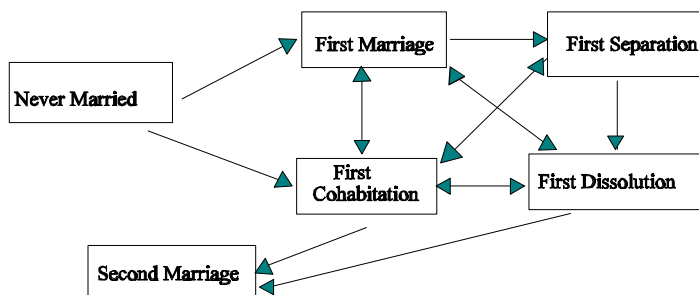
This paper is based on the premise that to examine both the sequence and timing of events, there is no better approach than the state space approach, which follows directly from the definition of an event as a transition between two states. A **state space** is a collection of all states considered for 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 analyse 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 non-recurrent 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 non-recurrent 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.

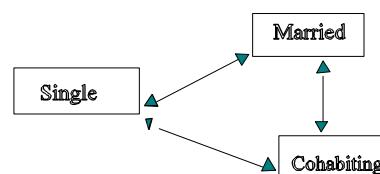
Figure 1.1 illustrates the direct transitions involved in a system of (a) non-recurrent states only, (b) recurrent states only, and (c) a mix of recurrent and non-recurrent states.

A state space will normally consist of some **transient** states (that is, states from which exits are possible) and some **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 shifts between marital states until the time of second marriage, then the state Second Marriage can be considered an absorbing state.

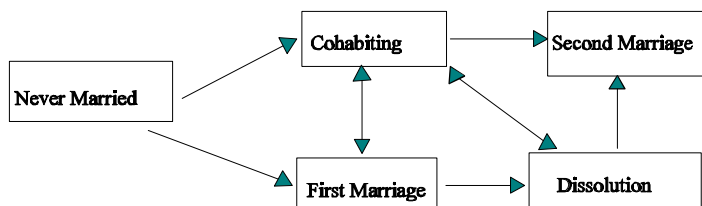
The above discussion makes it clear that **the first step in a life history analysis is to correctly specify the types of states in a state space**. This mainly depends on how complete or incomplete is the available information and on the scope and extent of a research problem. The way a state space 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 1(B)) would obviously lead to conclusions that are different from an elaborate analysis done with six (non-recurrent) states of Never Married, First Cohabitation, First Marriage, First Separation, First Dissolution and Second Marriage) as in Figure 1(A). 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.



A) Transitions in a system of non-recurrent states only



B) Transitions in a system of recurrent states only



C) Transitions in a mix of recurrent and non-recurrent states

6. Multistate Analysis of Event Dependencies

A fundamental assumption in life history analysis through a state space approach is that a specific stochastic process generates the observed events, and that the available data can be appropriately

analysed 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 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 a state space has been clearly defined according to a selected research problem, the next step in analysing 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.

a) Markov Model (⇒ 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 a variety of 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. It is true that the initial versions of multistate analyses predominantly 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 analyse 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 transition. Thus, a Markov-based multistate analysis of life history data can provide 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.** These two aspects of timing capture the influence of social norms. In life history analyses, 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), one can prefer to deal with a Markov model mainly as a tool for examining the age-dependence of transition probabilities.

b) Semi-Markov model (⇒ Duration Makes a Difference)

Of particular interest in analysing 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 study of migrations, 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 analysing life histories, besides duration in a state, the age at entry into the state can also influence the probability of transition. A minor modification 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, 1985). For example, in analysing 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 $p(t)$.

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.

c) 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 very explicit, and a model that considers the history of events becomes non-Markovian.

Nobody doubts that past history greatly 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 analysing 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 very 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. Only a non-Markovian framework can handle such dependencies.

However, it is generally very 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.

7. 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.

¹ I say "may leave untouched", because in my opinion, a non-Markovian 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.

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 very 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 modelling procedures, in which there is a function of timing which is adjusted in many ways by the effects of observed heterogeneous covariates. 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 make an attempt 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.

8. Multistate Analysis: Data File and Commands in LIFEHIST

The first practical problem encountered in analyzing life histories is arrangement of relevant data. This problem is accentuated when one considers several transitions among states. Any 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.

A data file for analyzing life histories generally 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 described in this book. 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 input file for running most of the multistate programs in LIFEHIST will contain

- i) a set of **ages at occurrence of events** considered in the state space; and,
- ii) a set of **instructions** to a program in terms of commands and arguments.

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 way to arrange the data on ages at occurrence of events (in short, timings) is explained below with illustrations. A set of instructions to the program are then added while running the program on Windows; this procedure is called Parameter Specification. These specifications instruct the program how to analyse and present the results in the output 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.

Step 3: Prepare the data file consisting of individual sample weight (if any), age at the time of survey, the state in which life history starts and ages at which all the events considered for analysis occur. If an individual does not experience an event, assign a missing value for the age at occurrence of this event.

The data file can be prepared by extracting the necessary information from a source file through any available mechanism. The number of variables 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 quite easy to prepare the file under when the state space contains only non-recurrent states. Extra effort will be needed to prepare the data file with one or more recurrent states.

The essential information for analyzing life histories is contained in the following list of variables:

- 1) Surveys often carry sampling weights, neglect of which would seriously bias the inferences made from an analysis. This point is often ignored in many studies, particularly those that deal with multivariate analyses. If sampling weights are used, LIFEHIST is instructed to read the weights from the file. Individual sampling weight should be the first variable in the data file. If no weights are used, unit weight will be assumed for all cases and the weight variable is ignored in the data file and the first variable will be the age at the time of survey.
- 2) The second variable in the data file is an individual's *age at the time of survey*. LIFEHIST uses this variable to find at what time (age or duration) an individual is (right) censored and in which state.
- 3) Then follow the **ages** at which specific events occur in an individual's life. 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.

- ▶ The number of event-specific ages depends on whether the state space consists only of non-recurrent states or whether there is a mix of recurrent and non-recurrent states. If only non-recurrent states are used, the ages when events occur are given in *the same order in which the states were assigned integer values* in Step 2. If recurrent events are included, the ages are to be presented in a more detailed way (not shown here).
- ▶ The missing value can be uniform for all events and for all individuals, say for example, 99.99. Or, they can vary from individual to individual and from event to event. Many survey data are coded with multiple missing values such as 99.4, 99.5, 99.9, etc. The instruction to the program asks, therefore, for the *lower limit* of missing values. Users should be careful in choosing this missing value, preferably beyond the range of ages at which different events occur. LIFEHIST uses this information along with the age at the time of survey to find at what time a life history ends (technically speaking, right censored) and in which state. All the programs in LIFEHIST assume Type I or right censoring only.
- ▶ LIFEHIST finds the sequence of events by using the information on ages at occurrence of events. It is important therefore that these ages are given with as much precision as possible. Most statistical procedures developed for life history analysis assume that the times of events (or transitions) are measured precisely. In practice, however, timings of events are never measured precisely. Instead, they are always rounded or grouped. Some surveys record the year and the month when an event occurred, which though not very precise is good enough for many analyses. Many 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 very unsatisfactory and sometimes may be erroneous unless extra care is taken in giving proper instructions to LIFEHIST about admissible transitions (see below). When data are grouped or rounded, only discrete procedures can be used for analysis, provided no events are systematically omitted. Anyhow, users should carefully check the sequencing output produced by LIFEHIST.

Box 1 presents as an example the data file prepared for a marital history analysis with six states: (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 (among the events considered here, all but first cohabitation). The analysis starts either in the Never Married State or in the First Cohabitation state at age 15 and ends with second marriage. Therefore, the state Second Marriage is considered absorbing.

Box 2.1. Example of data file containing only non-recurrent states
Life history starts in the first state

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
.....							

Box 1 gives the marital histories of seven women. The first woman is aged 38 at the time of survey and has been assigned a weight of 0.84. This is the woman's age *in completed years* at the time of survey. First Cohabitation, reported by the sixth woman is also recorded in complete years. In fact, a more precise information on 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 very 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. For the first woman, the age at first marriage has been recorded as 13.42 which is earlier than the assumed age at the start of marital history. If any such discrepancies 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.

LIFEHIST finds the sequence of events by their recorded ages, and finds in which state a woman is censored.

In addition to the data on timings of transitions, the following commands and arguments are used for running various multistate programs:

- 1) **WEIGHT**
- 2) **MISSING VALUE**
- 3) **N OF STATES**
- 4) **STATE NAMES**
- 5) **Number of TRANSIENT States**
- 6) **Any RECURRENT states**; if yes, additional commands
- 7) **TIME TYPE**
- 12) **Admissible transitions**
- 13) **Pairs of origins and destinations**