2

Multistate life tables in public health

Abstract

The main objective of this chapter is to review the multistate life table (MSLT) method and to emphasize the potential utility of this method in public health research. During the past decade, various public health researchers have applied this method as a means to measure health expectancy, which is the best-known index for summarizing the health status of the population and addressing the issue of 'compression' and 'expansion' of morbidity. In this chapter, we describe the theoretical aspects of the multistate life table approach that relate to public health status measurement. We distinguish between prevalence- and incidence-based measurements, and conceptualize these from the public health point of view. A concise overview of the mathematical construction of multistate life tables is given in this chapter. We describe the mathematical equations, focusing on micro data instead of macro data. We discuss population-based and status-based measures of health expectancy. We generalize the multistate table, describing it in multiple covariate contexts. The input data construction and the possible output from multistate life table are also illustrated. We demonstrate possible methods to calculate the confidence intervals around multistate life table statistics. The fundamental difference between the multistate life table method and the Sullivan method is described as well. We found that the MSLT method provides a more accurate assessment of health expectancy in a population. Overall, this chapter gives an overview of theoretical and practical aspects of the multistate life table and indicates the methodological innovation and utility of this method in public health research.

2.1 Introduction

One of the most important and lively current debates in the study of public health and mortality revolves around the idea that as improvements in survivorship and life expectancy continue, the health of those individuals benefited by these improvements may deteriorate (Fries, 1980; Singer and Manton, 1994). This improvement of survivorship raises the question of whether or not an increase in the proportion of older people will result in an increase in the prevalence of chronic disease and disability i.e. expansion of morbidity. On the other hand, public policy debates in the areas of health and medical care have also emerged in recent years around the question of whether improvements in life style and medical technology will delay the onset of chronic illness and disability and result in a compression of morbidity at older ages (National Research Council, 1988). For modeling the life history of cardiovascular disease and addressing the issue of compression or expansion of cardiovascular morbidity, we chose to make use of the time-honored 'multistate life table' method. The objective of this chapter is to review the multistate life table, describe it from public health perspectives and investigate its utility and applicability to measure population health status.

Over the past three decades, public health researchers have focused on the development and application of health measures that combine mortality and morbidity data. Efforts to collapse mortality and morbidity into a single measure to provide summary measures of population health stretch back far into the past (Sander, 1964; Chiang, 1965; Moriyama, 1968; Sullivan, 1971a; Katz et al., 1983; Rogers et al., 1990; Preston, 1993; Crimmins et al., 1994; WHO, 2000). The volume of work from members of the Réseau de Espérance de Vie en Santé (REVES[1]) offers an indication of the activity in this field. The Sullivan method (see Hauet, 1997; Mamun, 2001) is one of the methods that is widely used in the public health domain to measure health status. The index is attractive for public health researchers, because this method is very simple and the data access is easy. There are, however, serious limitations to this method, such as the fact that it is unable to take into account any sudden change in health status (incidence) and re-entry into the life table population. A method that does not have these limitations is the multistate life table. The main advantage of this method is that it can accommodate both the increment and decrement according to various (health) statuses of the population. In this chapter, we explicitly distinguish between the concept of prevalence and incidence.

[1] REVES: International Network of Healthy Life Expectancy. REVES is an international organization of researchers and health planners from universities, governments, and international agencies dedicated to understanding the use of health expectancy as an indicator of population health and as a tool for health planning.

In public health, Katz and co-authors (1983) introduced multiple decrement life tables. They found that this method gives better estimates than the Sullivan method. This method allows several exits from the healthy population: through death, dependency in carrying out activities of daily living, or institutionalization. But this method does not include differential mortality by health status and transition from disability states to healthy states (Saito et al., 1999). That is, this method considers change in health states from healthy to disability and death, from which no return is allowed. In practice, however, people can recover from disabilities or from diseases (acute or chronic) that affect their activities of daily living, and can be released from institutions. "This clearly produces an estimate of active life expectancy that is biased downward" (Saito et al, 1999). The multistate life table, an extension of basic life tables, allows more complexity to enter the analysis: people can enter, as well as exit a population and can move back and forth across a variety of states within a population. In 1989, Rogers and colleagues constructed the multistate life table to calculate the active life expectancy. They adopted the multistate life table technique to incorporate both the decrements and increments of several interacting sub-populations (based on the activities of daily life) to measure the active life expectancy. This was the first time the MSLT method was applied to the field of public health, although many researchers (Rogers et al., 1990, Crimmins et al., 1994; Saito et al., 1999; Laditka and Hayward, 1999; Mathers and Robine, 1996; Barendregt et al., 1997; Nusselder, 1998; Peeters et al., 2002; Mamun, 2001) subsequently went on to use this method.

Most of the researchers compared the multistate method with the Sullivan method using the same data set. There is no conclusion as to which method is the best one. However, methodologically and theoretically, the multistate method is preferable (Mathers and Robine, 1996; WHO, 2000). The key issue of the debate centers on the use of prevalence proportions versus incidence rates in the estimation of health expectancies. The MSLT method is data-demanding in nature. The Sullivan method is widely used, as it is very easy to calculate and solely takes prevalence information into account. The Sullivan method is based on the status data, while the multistate method is based on flow data. For the theory of compression or expansion of morbidity and in the health dynamics of an individual life cycle, the emphasis should be on the multistate method (Barendregt et al., 1997).

A key component of the multistate method is its relaxation of the assumption of unidirectional health changes over the life course. The multistate method allows individuals, as they age, to experience the onset of health problems and the recovery from those problems. No assumptions are made about the hierarchy of health status changes. Thus, persons can experience multiple and recurrent health problems over their lives. This method deals with these transitions from one state to another: from alive to dead, healthy to unhealthy, unhealthy to recovery,

employed to unemployed, and so on. Some of the transitions may occur only once, whereas other transitions can be repeated. Some transitions may preclude other types of transitions or may make further transitions possible (Willekens, 1987). This method often offers answers to questions such as: what is the probability/rate that healthy men of 65 become unhealthy by age 75? Or what is the probability that healthy women of 50 suffer from cardiovascular disease by age 65? How does the probability of being in a health state differ in terms of duration? What is the average amount of time spent by an individual in good health or with disability in his or her life? What is the lifetime probability of developing cardiovascular disease? What are the differences between smokers and non-smokers in the number of years spent with disease? At the generic level, what is the impact of cardiovascular disease risk factors on life expectancy? A simple way to understand the events and relations involved and indeed to begin to answer these questions in public health research is the MSLT approach. This method is capable of providing estimates of the expected number of transitions over a person's lifetime or within a specified period of time. Thus, researchers can gain some sense of the health trajectories experienced by persons in the population.

The multistate method is very attractive because it can capture the natural course of a disease, and can encompass patients who are cured or have intermittent disease-free periods (Barendregt, 1994). The multistate model is used to chart the disease process or disease history and to detect the pattern of similarities or differences by basic explanatory factors such as sex, smoking habit etc. Recently, the multistate method has been used to model the population health and resource use (Niessen, 2002). A multistate life table can describe the transition of life and sojourn time in various stages of life. Using this dynamic approach the reconstruction of an individual biography is possible, even based on the elementary event (two states and one event) process. This method affords a unique opportunity to reconstruct the life history and to construct a 'synthetic biography' (Willekens, 1987).

The multistate method originated in the early 1970s in multiregional demography (Rogers 1973; 1975). Schoen (1975) and Schoen and Land (1979) applied it to the analysis of marital status. Since that time, the field has expanded and developed, as multistate methods have been found to be both powerful and flexible enough to capture the movement of the complex behavior of a cohort (Mills, 2000). Willekens (1987), following the work of Rogers (1973, 1975), Schoen (1975, 1988a) and others, argues that " the life table has made the transition from a method for estimating the length of life to a technique for describing the structure of life". The key to the development of this method was the use of the Rogers' mathematical demography, especially the application of matrix algebra in the life table.

The application of the multistate method includes diverse subject matters related to the family (Bongaarts, 1987; Yi, 1991), labor force participation (Willekens, 1980; Lalu, 1992), migration (Rogers and Willekens, 1986), spatial population distribution (Willekens and Rogers, 1978), marital status life table (Willekens, 1987), union formation and dissolution (Mills, 2000), contraceptive use (Islam, 1994), voting status (Land et al., 1985), active life expectancy (Rogers et al., 1989, 1990; Branch et al., 1991;Land et al., 1994; Crimmins et al., 1994; Hayward et al., 1998; Liu et al., 1995; Brouard and Robine, 1992; Izmirlian et al., 1997), mortality and morbidity projections (Rusnak et al., 1992) and health status and life expectancy (Crimmins et al., 1994; Nusselder, 1998). The majority of these studies used vital statistics or census data, for estimating the parameters of multistate life tables. Recently, researchers from variously different disciplines adopted this technique to measure health status (especially for aged population) by calculating the health expectancy or active life expectancy. However, from a public health point of view the explicit explanation of this method is inadequate. Here, we describe and explain the multistate method from the public health point of view. We present the theoretical issues that are essential to construct an MSLT. A concise overview of the mathematical construction of an MSLT is given in this chapter. The MSLT in multiple covariate contexts is discussed. From the technical point of view, we also describe the input data construction and the representation of output from the MSLT. For the statistical competence, we have described the available methods to calculate the confidence intervals of the outcomes from MSLT. The MSLT described in this study is a type of descriptive transition rates model with and without covariates. The multistate model is used to chart the cardiovascular disease process or disease history.

This chapter consists of the following sections. In Section 2.2, the distinction and interrelationship between prevalence and incidence has been made explicit. The Sullivan method and its limitations are discussed in Section 2.3. The multistate life tables are illustrated from a public health point of view in Section 2.4. The conceptual issues of MSLT are discussed in Subsection 2.4.1. In Subsection 2.4.2, the theoretical basis of the multistate life table, i.e. the Markovian property, is talked about. The MSLT equations are derived and discussed in Subsection 2.4.3. It includes the derivation of transition probabilities and equations for life table statistics. MSLT in covariate contexts is discussed in Subsection 2.4.4. The estimation of confidence intervals is described in Section 2.4.5. The input data to construct MSLT and the output from MSLT are described in Subsections 2.4.6 and 2.4.7. We conclude this chapter with a critical assessment of MSLT in public health.

2.2 Prevalence and incidence measures

Epidemiologists and demographers describe the magnitude of health problems either in terms of prevalence or incidence, or a combination of both. Incidence and prevalence are two basic measures of disease occurrences in populations (Selvin, 1991; Young, 1998). Prevalence (or proportion) reveals how many cases exist in a population at a given time. Incidence is usually expressed as the number of new cases occurring within a population at risk (for example, of contracting a disease) over a period of time. For epidemiological purposes, the occurrence of cases of disease must be related to the "population at risk". The concepts of incidence and prevalence are discussed here.

2.2.1 Prevalence

Prevalence refers to the proportion of a total population that has a defined health problem or disease. It is usually measured by surveying a particular population containing people with and without the condition of interest. The prevalence is usually reported as percentage. Usually, prevalence is of two types: point prevalence and period prevalence. We identify another type of prevalence, duration prevalence. Any measures of prevalence are proportions- as such they are dimensionless and should not be described as rates (Friis and Sellers, 1999).

Point prevalence

According to Selvin (1991), point prevalence (often called prevalence "rate") is the number of affected individuals in a population at a specific point in time divided by the size of the population under consideration. For example, the point prevalence proportion of coronary heart disease in the population age 65 and above in a specific country is the number of existing cases divided by the number of people above age 65 in the country on a specific date.

Period prevalence

Selvin (1991) defines a period prevalence (also usually called a "rate") as the number of affected individuals in a population plus a count of new cases over a defined period of time divided by the size of the population under consideration. The numerator is the combination of incidence and prevalence (point prevalence) and the denominator is the same as that in the point prevalence proportion measure. This measure is used less since it combines both incidence and point prevalence into a single number, which can be difficult to interpret.

Duration prevalence

We also identify a type of prevalence known as duration prevalence. This measure is used in some life table calculations. It is the proportion of time that a person is affected or a group of people are affected. It is a ratio of two durations that can be formulated as follows:

$$\text{Duration prevalence} = \frac{\text{The duration with a disease or impairment}}{\text{The total duration under observation}}$$

2.2.2 Incidence

Incidence measures the development of a disease or health problem in a population i.e. the frequency of NEW cases in the population at risk during a specified period. The term incidence refers to the absolute number of new cases and the term 'incidence rate' refers to the relative number of new cases (i.e. the new cases related to the population at risk). Incidence can be measured in two ways: (i) incidence proportion and cumulative incidence and (ii) incidence rate or incidence density.

Incidence proportion and cumulative incidence

For a given interval of time, we can express the increase in incidence number per unit increase in population size. If we measure size at the beginning of the interval and no one enters the population (immigrates) or leaves alive (emigrates) after the beginning of the interval, such a rate becomes the proportion of the people who experience events among those who entered the interval. This quantity is called the incidence proportion (Rothman and Greenland, 1998), which may also be defined as the proportion of a closed population at risk that becomes diseased within a given period of time. This quantity is often called the cumulative incidence (Miettinen, 1976 in Rothman 1998). The incidence proportion can be measured as:

$$\text{Incidence proportion} = \frac{\text{Number of new cases during stated period}}{\text{Number of persons at risk at beginning of period}}$$

The numerator relates to events. The denominator is the population at risk (risk set) of the event of interest. It consists of people who do not have the disease and who can have the disease. An example is the so-called attack rate[2] used in epidemic investigation. For example, if we start with 245 susceptible persons and 6 develop the disease over the study period, the risk during that study period = 6 / 245 = .0245, or 2.45 percent. The cumulative incidence is the proportion or

[2] Note that the attack rate is not a rate in the usual sense. It is a proportion instead.

probability of healthy individuals who contract the disease during a certain period. An often-used measure is lifetime incidence. Since incidence proportion requires the follow-up of individuals over time, it is a longitudinal measure of disease frequency. Incidence proportion is often expressed in terms of a population multiplier. For example, an incidence proportion of .0245 may be expressed as 24.5 per 1000.

Incidence rate or incidence density

Public health usually deals with large populations such as a city, state/province or country where it is not feasible to establish a disease-free group to be followed in time or calculate the person-time of observation for each person. Sometimes, measurement of incidence is complicated by changes in the population at risk during the period when cases are ascertained, for example, through births, deaths, or migrations. Therefore, instead of using the number of people at the start of the observation period as the denominator, we can define a period of observation and determine for each person the actual time period at risk, from the beginning of the period to the time the disease is detected. Or in the case of a person who does not become sick at all, to the end of the period of observation. The duration at risk is calculated by adding together the periods during which each individual member of a population is at risk during the measurement period. The duration is usually expressed in terms of *person-months* or *person-years* (i.e. is called person-time at risk). According to Rothman and Greenland (1998), the number of new cases of disease (i.e. incident number) divided by the person-time is the incidence rate of the population over the period:

$$\text{Incidence rate} = \frac{\text{Number of new cases during a period of time}}{\sum_{persons} \text{Time spent free of disease during the same period}}$$

The incidence rate is also defined as incidence density (Miettinen, 1976a in Rothman, 1998). This measure has also been called the person-time rate, force of mortality or morbidity, hazard rate and disease intensity, although the latter three terms are more commonly used to refer to the theoretical limit approached by an incidence rate as the time interval is narrowed down zero (Rothman and Grenland, 1998).

2.2.3 Prevalence versus incidence

Prevalence and incidence are interrelated. Prevalence is dependent on the incidence and the duration of disease. In a stable situation this relation may be expressed (Rothman and Greeenland, 1998) as:

$$P = I * D * (1 - P)$$

$$\Rightarrow \quad \frac{P}{1 - P} = I * D$$

where P= prevalence (i.e. not free from disease), I= incidence and D= duration. The left side of the equation is odds (ratio of a probability to its inverse). The denominator on the left side of the equation is the part of the population that is free from the disease. For rare diseases, where P is low such that (1-P)≅1, then the following approximation may be used:

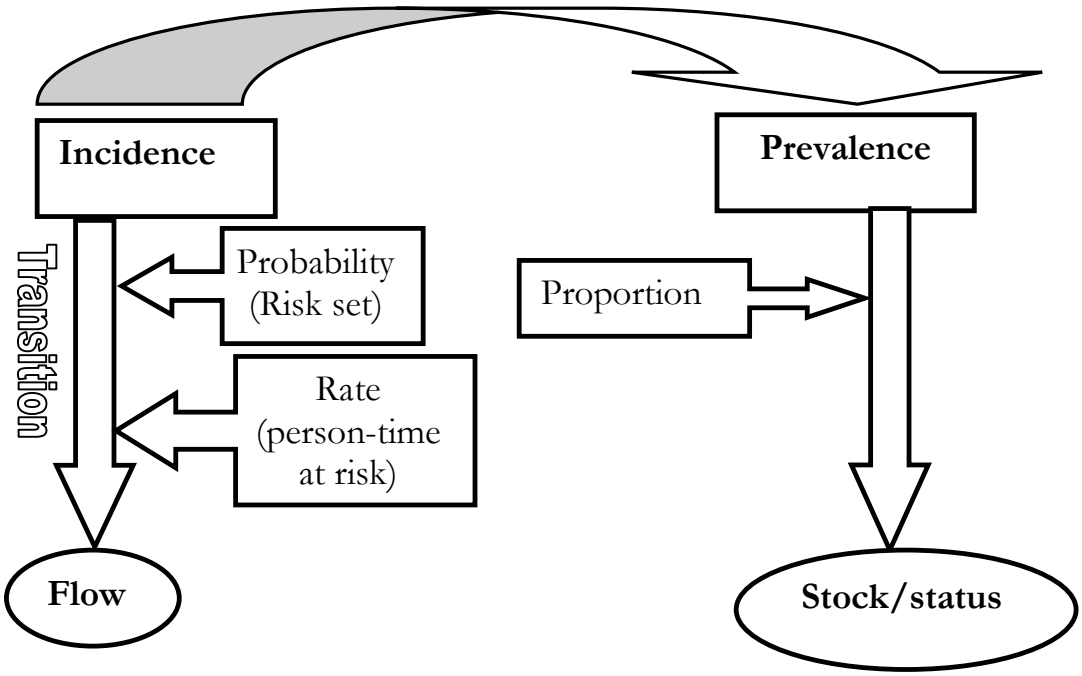
$$P=I*D$$

This relationship states that prevalence varies directly with both incidence and duration. For instance, if the incidence proportion is 10 new cases each month per 10,000 individuals and the duration of observation is 5 months, then the prevalence in this population is 50 cases per 10,000 individuals. This simple relationship, that prevalence is the multiplication of incidence proportion and duration, holds under rather strict steady-state conditions (Selvin, 1991). In realistic situations, these steady-state conditions rarely occur since diseases generally have complex incidence/prevalence dynamics influenced by background characteristics (race, age, sex, medical care etc). Despite only a few people in a group becoming ill each year, if the disease is chronic, the number of diseased people will mount and the prevalence will be relatively high in relation to the incidence. Whereas, if the illness is of short duration, because of either recovery or death or if there is migration of diseased persons from the area, the prevalence will be relatively low.

The differences between prevalence and incidence measurement are conceptualized in Figure 2.1 and summarized in Table 2.1. Incidence is measured over a period of time and prevalence is measured at a point in time. Prevalence differs from incidence in that it refers to status rather than change of status or event. The numerator is the count of existing cases rather than new cases (events). Incidence is about becoming, whereas prevalence is about being something or having something (Young, 1998). Prevalence is always a proportion and is dimensionless, whereas incidence could be either proportion or probability based on the risk set (cumulative incidence, dimensionless) or rate based on occurrence-

exposure (incidence density, dimension is per-unit of person-time). Prevalence data are stock data; incidence data are transition data (Flow).

Figure 2.1 Conceptualization of incidence versus prevalence measurement



The *risk set* is defined as the total number of individuals at risk of experiencing the event under study during a unit interval, accounting for censoring. In the multistate life table we use probability or rate to measure the risk level. We use risk set (as denominator) to measure the probability of an event. The *probability* is defined as the number of events to the risk set. Measuring risk level using *rate* we use exposure (as denominator) instead of risk set. The rate is defined as the number of events per unit of exposure. It is the occurrence-exposure rate i.e. occurrence/exposure. The basis is the duration of exposure or duration at risk. The concept of a rate differs from a probability. The rate used for dynamic analysis depends on the risk period or exposure time while the probability depends on risk set. A probability does not incorporate a direct reference to time whereas a rate is a measure of change per unit of time. A probability is a unitless value between 0 and 1. Rate has a unit and takes any value. Therefore, the estimated parameters of MSLT might vary depending on whether the estimation procedure is based on the risk set or exposure time.

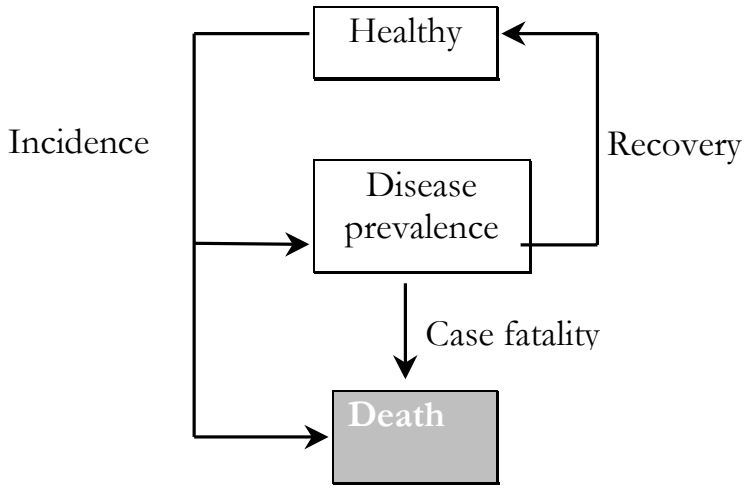
Table 2.1 Prevalence vs. incidence

	Prevalence	Incidence
Cases	Entities	Events
Source population	At risk to be a case	At risk to become a case
Time	Static (point)	Dynamic (interval)
Data	Stock	Flow

Source: www.sph.unc.edu/courses/EPID168, © Victor J. Schoenbach 2000, Measuring disease and exposure, cited: August 28, 2002

Mortality is the incidence of death. The interrelation of incidence, prevalence, and mortality is depicted in Figure 2.2. The mortality rate is a specific kind of incidence rate where the incidence or event is death. The key concept is a change in status, from healthy to sick, from alive to dead. From Figure 2.2, we can say that each new (incident) case enters a prevalence pool and remains there until either recovery or death.

Figure 2.2 Interrelationship of incidence, prevalence and mortality



The prevalence of disease in the population is an indicator of the current stock of health. It is the result of past rates of disease incidence, disease progression, and survival. Over time, the prevalence of disease may change because of increases or decreases in risk factors for various diseases and because of an increasing ability to treat diseases, thus delaying their progression to disability and death. The relative size of the change in incidence rates, recovery and death rates will determine change in disease prevalence (Crimmins and Saito, 2000). Prevalence is often used as an alternative to incidence in the study of rare chronic diseases such as multiple sclerosis, where it would be difficult to accumulate large numbers of incident cases. Again, however, care is needed in interpretation.

2.3 The Sullivan method

The Sullivan method originated in 1964, when a researcher, Sander, proposed a health indicator that combined information on mortality and morbidity. The first example of such an indicator was published in a report of the US Department of Health Education and Welfare in 1969. It contained preliminary estimates of disability free life expectancy calculated using a method devised by Daniel Sullivan (Sullivan, 1971a). He developed a life table technique that collapses both mortality and morbidity into a single composite index of disability free life expectancy. The method uses the observed prevalence of disability at each age in the current population at a given point in time, to divide the years of life lived by a period life table cohort at different ages into years with and without disability (Mathers and Robine, 1997).

A large number of studies presented in the REVES network employed the Sullivan method to calculate estimates of healthy life expectancy (Colvez and Blanchet, 1983; Katz et al., 1983; Wilkins and Adams, 1983; Bebbington, 1988; Crimmins et al., 1989; McKinlay et al., 1989; Rogers et al., 1989; Mathers, 1991). Robine et al., (1995) calculated disability-free life expectancy for over 30 countries using the Sullivan method. Robine et al., (1998) examined trends in disability-free life expectancy in France in the period 1981 to 1991, and concluded that the proportion of life lived free from significant disability increased during the period studied. A similar conclusion was reached by Crimmins et al., (1997) who used the Sullivan life table model to examine trends in life expectancy in the United States from 1970 to 1990. They found evidence of morbidity compression in the 1980s. Hayward and Heron (1999) recently used the Sullivan approach to examine racial heterogeneity in active life expectancy in the United States and they found substantial differences in total and active life expectancy across racial groups.

Up to the second half of the 1980s, most researchers used the Sullivan method to calculate the healthy expectancy. The method uses information on the prevalence of health states in the population. The estimates of active life expectancy for the USA had also been made using double decrement life tables (Katz et al., 1983) with two competing events: disease and death. The Sullivan method reflects the current health status of a real population that is adjusted for current mortality levels. It relies on the input of prevalence measures of the age-specific proportions of the population health conditions, generally gathered in cross-sectional surveys, and on information taken from a period life table (Hauet, 1997). The method is an attractive one, with its simple calculation, low data demand, relative accuracy and easy to interpret outcome.

The Sullivan method resembles the method used in the construction of life tables of working life developed by Wolfbein in 1949 and later on used by others for the construction of tables of working life in studying labor force participation.

The method relies on restrictive assumptions such as the unimodality of the labor force participation curve, entrance into labor force only at ages before the peak and retirement only at ages after the peak, and independence of mortality from labor-force status (Willekens, 1980). These assumptions were implicit in working-life tables that had been published in a large number of countries. After 30 years Hoem and Fong (1976) and Willekens (1980) calculated a multistate table of working life that had no restrictive assumptions. The basic difference in their method is its focus on flows instead of stocks. The Sullivan method uses prevalence, or stock data instead of incidence, or flow data. In the Sullivan method any number of unhealthy states can be adapted- for example, short term and long term disability, with and without disease, severe and less severe disability. Details of the Sullivan method and its consequences are described elsewhere (Hauet, 1997; Mamun, 2001).

In general, eliminating a disease decreases the age-specific probabilities of dying, as well as the age-specific prevalence of disability. Therefore, the application of the Sullivan method would result in a considerable overestimation of the disability effect (WHO, 2000). Even though problems are recognized in this method, it is widely used around the world.

On the other hand, the MSLT method uses data on transitions between health states e.g. from healthy to ill health and vice versa, or from either state to death. The MSLT could be health status-based (Rogers et al., 1989) and population-based (Crimmins et al., 1994). The difference between status-based and population-based measures is discussed in subsection 2.4.3. The limitation of the MSLT is that it requires information for at least two points in time that may not be available. Nevertheless, the use of this method is increasing because this method is both more accurate and more informative than the Sullivan method (WHO, 2000). According to a WHO (2000) comparison of the MSLT approach and the Sullivan method, the

> *'multistate method is more accurate, specifically because it is based on transition rates which reflect ways in which the health state of the population changes and develops at the same time. In contrast, prevalence rates are the result of the previous history of the population'.*

The Sullivan method and the MSLT will give the same estimate of health expectancy if actual transition rates have been stable for a very long time (Barendregt et al., 1994; WHO, 2000). Otherwise, health expectancies calculated by the Sullivan method lag some way behind what is actually happening and consequently, they may indicate a compression or expansion of morbidity when in fact the reverse is taking place (Barendregt et al., 1994). Laditka and Hayward (1999) point out that the development of the multistate method provides demographers with a means to model the dynamics of health in the population in a

more realistic fashion than the Sullivan method. In our study, we make extensive use of multistate life tables to describe the life course of cardiovascular disease.

2.4 The multistate life table

Of the models used by mathematical demographers, the life table model is the most representative. It follows a group of people, born during the same period, experiencing transitions between two or more states over time and age. In the simplest situation, i.e., in the case of a conventional life table, there are two states, namely alive and dead. The transition is irreversible from the former to the latter; the state of death is an absorbing state. The extension of this ordinary life table is the multiple decrement life tables, which distinguishes more than one final absorbing state. A well-known example is cause of death. This model cannot deal with transient states, i.e., states that can be entered and left again. Such states can easily be handled using the *multistate life table* (MSLT) or *increment-decrement life table* (IDLT*)*. The MSLT models allow individual members of the life-table population to move into states (increment) and out of states (decrement). These movements or transitions are the fundamental concepts of the MSLT framework.

The MSLT approach can be viewed from two perspectives. The first is the population (macro) perspective: the life table is seen as a description of a stationary population. The second is the life history (micro) perspective: the life table can be viewed as describing the life course or life history of members of a synthetic cohort i.e. the cohort biography (Willekens, 2002). The life history perspective is a micro perspective since it focuses on transitions at the individual level. In the micro perspective, multistate models describe transitions in life and sojourn times in the various stages of life. It visualizes the structure of life. For instance, age-specific transition rates and transition probabilities produce a life structure, and can answer the question of, how t life structure changes when transition rates or probabilities change. The life history perspective is also a longitudinal perspective that focuses on the life course development of cohort members (intra-cohort development) (Willekens, 1987). In the macro perspective, the MSLT describes the dynamics of interdependent sub-populations, where the sub-populations are defined on the basis of personal attributes (health status, place of residence, marital status, employment status etc). The MSLT described in this study is a transition rate model with and without covariates. The multistate model is used to chart the cardiovascular disease process or disease history.

2.4.1 Conceptual issues

A life table is a systematic way to keep track of the experiences (e.g., disease) of a group of people. The experiences of this group can be gathered from birth until

death or over a short period of time. Several conceptual issues that are important in constructing life tables are described here. The issues are: types of MSLT, state and state space, events and transitions, risk and exposure, observation period, and exact age and age in completed years.

Types of MSLT

There are different types of MSLT. The life table can be nonhierarchical or hierarchical, population-based or status-based, uni-radix or multiradix. A hierarchical model is when all states, except the initial state, have only one possible transition into the state. Hierarchical models deal with cases moving sequentially through successive states without reentry. In nonhierarchical models reentry into a given health state is possible. The preference for a nonhierarchical model is determined by the fact that a regaining of health functioning, for example, reentering the non-disabled or non-diseased state, has shown to be a significant force of increment (Rogers, et al., 1989; Mathers and Robine, 1997).

MSLT measurement can be population-based or status-based. The population-based multistate model deals with the entire population irrespective of initial (health) states. It is an unconditional measure, which describes the life history of an average person of a given age, irrespective of the person's health status at that age. The health status-based measures depend on the health status of the person for which the measure is calculated (Willekens, 1987). Both the population-based and status-based measures are possible for the same data set. For example, the Longitudinal Study of Aging has been used to estimate life expectancy with and without disability of persons who were disabled at age 70, which is status-based, and of the total 70 year- old population, which is population-based (Crimmins et al., 1994). The choice between population-based and status-based depends on the research objectives and research questions. If the primary focus is on change of the health status of the population and on addressing the compression of morbidity discussion, a population-based model is used (Crimmins et al., 1994).

Based on the observation period[3], an MSLT can be of two types: cohort-life table or period-life table. Cohort refers to a group of persons who experience a particular event (e.g. heart stroke) during the same period of time. If the event experienced is birth, then the corresponding cohort is called a birth cohort. The period life tables deal with the study of different cohorts at the same point in time i.e. the cross-section of the population. For example, the life table constructed based on the period from 1968 to 1978 of the Framingham Heart Study (FHS) is a period life table. Depending on the research questions, research objectives and the availability of data we use cohort life table or period life table. As observed in the foregoing, the MSLT could be uniradix or multiradix. In a uniradix life table the initial cohort is concentrated in one state (e.g. healthy or never married) while in the

[3] The period during which the events are recorded

multiradix life tables, the initial cohort is allocated to several states. Life tables are also classified according to the length of age interval in which the data are presented. A complete life table contains data for every single year of age. In an abridged life table, information is given only for broader age intervals, e.g. 5-year age groups.

State and state space

State is defined as a specific attribute of an individual at a given age and time. Attributes are individual characteristics or traits. Persons with a given set of attributes are said to occupy a given state. In general, the attributes are objective and we can measure them. They may also be subjective and refer to values, attitudes, opinions, perceptions, or evaluations (Scott and Alwin, 1998). In modeling, a category of an attribute or a combination of categories of a set of attributes is referred to as a state (Willekens, 1999). For instance, if a person is healthy at exact age 50, he or she occupies the healthy state at that age. At the same age, he or she might be married, may have two children, highly educated and have a high-ranking job. All together constitute his or her attributes at age 50 and healthy is a specific attribute that is a specific state as well.

The number of states is finite and the state variable is a discrete variable. Therefore, the possible number of states can be represented by a discrete or categorical variable. States are classified as recurrent and non-recurrent. A non-recurrent state is entered only once and a recurrent state allows multiple entries. Those states from which exits are possible are called transient states, and the states from which no exit can be made are called absorbing states. In a multistate framework, recurrent and non-recurrent states can be combined to examine transitions between several states, which can be either transient or absorbing (Rajulton, 1999).

The collection of all possible states constitutes the *state space*. For example, to analyze changes in health status, a state space may consist of two states: healthy and unhealthy. To analyze changes in marital status, a state space may consist of the following five states: never married, cohabitation, first marriage, divorced/widowed and second marriage. In this example, the states never married, first marriage, and second marriage are non-recurrent (because these states can be entered only once), while cohabitation and divorced or widowed are recurrent (they can enter more than once).

Events and transitions

A change in attribute is called an *event* and an event is an outcome of some process. For instance, infection, recovery, entry/exit labor market, graduation, migration, marriage, birth of a child, are life events. These events are associated with underlying causal mechanisms. An event can also be thought of as a change that

places an individual in a new status, which is different from the previous status, he or she was in before the change took place. From a demographer's point of view, "A vital event is a major change in an individual's status which leads to a change in composition of the population" (Pressat, 1985). From an epidemiological point of view, recovery from a disease or incidence of a disease is an event. This general definition of an event enables us to visualize events as transitions between statuses (Rajulton, 1999). MSLT is the method that reconstructs the life history based on the given information.

An event is a *transition* from an origin state to a destination state. For instance, transitions may be from unhealthy to health (recovery), from never migrant to migrant, from single to cohabiting, from second to third births, or unemployed to employed. Events can be non-renewable or renewable i.e. some of the transitions can be experienced only once, whereas other transitions can be repeated. For example, migration and marriage are renewable events. First marriage, first birth, death are all non-renewable events because they only happen at the most once in a lifetime. Some transitions may prevent particular types of transitions or may make further transitions possible. Any transition may be defined by the state before and the state after the transition (Willekens, 1987). A sequence of events can be simply considered as shifts between successive states in the state space (Rajulton, 1992).

An individual experiences an event at a specific age, at a specific length of time since an earlier event, or at a specific calendar time (Rajulton, 1992). Sequence and timing of events are important for the multistate method. Events can be measured by different time scales (see Willekens, 1999; Rajulton, 1999). A sequence of events and stages of life constitute a career, for instance, marital career, fertility career. When two or more careers run simultaneously they are parallel careers. A set of interdependent or parallel careers constitutes the life course. In this study, we construct the life course of cardiovascular disease.

Risk and exposure

Risk and *exposure* are two other concepts important for understanding the multistate method. An individual is exposed or at risk of an event if he or she has a chance of experiencing the event. For example, only healthy individuals are at risk of disease, only married persons are at risk of marital dissolution and only fecund couples are at risk of conception. Risk is conceptualized in a variety of ways. Being at risk means that there is a chance of experiencing an event, i.e. the probability of an event is not zero. If the probability of occurrence is zero it means the individual is not at risk. Life table analysis is a risk analysis.

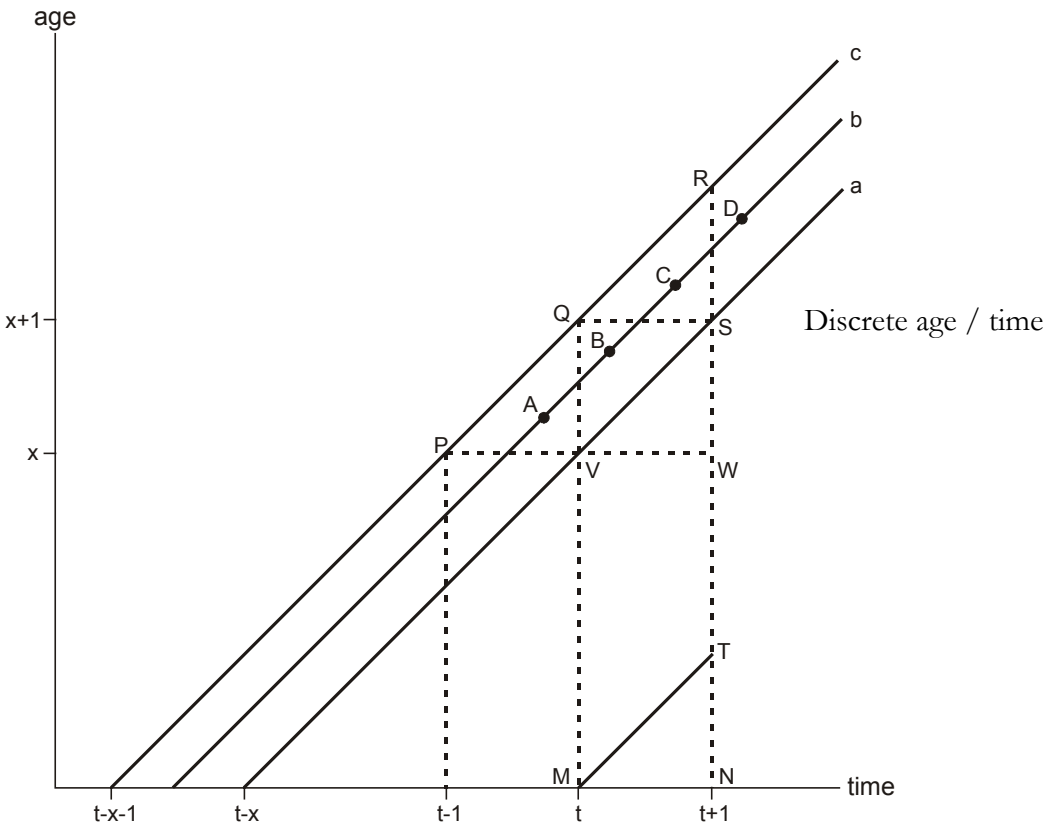
Exposure is defined as being exposed to the risk of an event; and exposure time is the duration at risk. For example, in case of CVD, the exposure time is from entry into the survey (provided the person is without CVD) to the moment of CVD or end of observation. People generally differ in risk levels and duration at

risk (Willekens, 1999). Risk indicators measure the risk level. The risk indicators are counts, probabilities, rates, odds, and relative risks. Exposure analysis engages measurement and estimation of the length of exposure and the level of exposure for those exposed (Willekens, 2002).

Observation period

Multistate life tables reconstruct the individual biography based on partial observation. The partially observed information is gathered during an *observation period*. The observation period is the period during which the events are recorded. The MSLT can be viewed as describing the life histories from birth or from a given age *y* (reference age) to the death of the members of a synthetic cohort (Willekens, 1987). The individual biography may be displayed graphically in the Lexis diagram (Figure 2.3). The explanation and interpretation of this diagram is similar to the Lexis diagram presented for the marital status life table by Willekens (1987).

Figure 2.3 Lexis diagram



The Lexis diagram is a popular choice for demographic analysis because of its focus on both personal time (age) and historical time. The horizontal axis of the Lexis diagram represents historical or calendar time and the vertical axis gives the person's age. Consider the lifeline of individual *b*. The events are located on the

lifeline. For example, A (Angina Pectoris: AP), B (Myocardial Infraction: MI), C (Other cardiovascular disease: OD) and D (Death: D). On the lifeline b, four events are shown (A, B, C, and D). Figure 2.3 shows an age interval (x, x+1), a time interval (t, t+1) and a cohort interval (t-x-1, t-x). The cohort consists of the group of people who experienced the initial event (birth) in the time interval (t-x-1, t-x).

The timing of events and the age at the time of events may be illustrated in a Lexis diagram. In this regard, two central questions are of importance:

i). How is time or age measured?

ii). When is time or age measured?

When measuring age, it is important to note whether the time intervals are discrete. It may be measured at any point of time in the interval. Suppose an interval has the length of one year. Age is measured either at the exact time of the event, at the end of the year or at the beginning of the year in which the event occurs. In a retrospective survey, the measurement of *age in completed years,* at the end of the interval, is common. It is equivalent to measuring seniority[4] in period difference. This difference is obtained by subtracting the year of occurrence of the initial event from the year of occurrence of the event under study (Willekens, 1987). The measurement of age at the beginning of the year is equivalent to the measurement of the year of birth. A person born between *t-x-1* and *t-x* is at time *t* between *x* and *x+1* years of age, i.e. of age *x* in completed years.

Implicit in the measurement of the timing of events in completed time units is the definition of the observation period. In the Lexis diagram, the observation period refers to the segment of the lifeline for which events are recorded. Events on parts of the lifeline not included in the segment remain invisible. The different ways of fixing the observation period leads to different observation plans (Willekens, 1987). We can present a few of these with reference to the Lexis diagram (Figure 2.3). We refer to read Willekens (1984) for the elaborate typology of observational plans.

Four observational plans may be distinguished, if we assume that the observation period is fixed (say, one year).

a. Cohort-age (cohort) observation

Cohort-age observational plan records for a person experiencing an event the cohort to which the person belongs and the *age in completed* years at the time of the event. The parallelogram PQSV in Figure 2.3 shows this observation plan. The observation period extends over two calendar years.

[4] The seniority is the time elapsed since age *y*, the age at which the event occurred that started the process (event-origin) (Willekens, 1987).

b. Period-cohort observation

This observational plan records the calendar year in which the event occurs as well as the cohort to which the person belongs. The parallelogram VQRS in Figure 2.3 shows this observational plan. The observation interval covers two age classes.

c. Period-age (period) observation

This observational plan records the calendar year in which an event occurs, as well as the age of the person in completed years at the time of the event. The square VQSW in Figure 2.3 represents this observation plan.

d. Age-period-cohort (APC) observation

For a person experiencing an event, an APC observational plan records the calendar year in which the event occurs, the year of birth of the person and his/her age in completed years at the time of the event. The triangle VQS in Figure 2.3 shows this observational plan. The observation period covers only one age, one cohort and one calendar year. The data that are recorded by an APC observational plan are commonly referred to as doubly classified data.

The ways in which the observation period is fixed affects the estimation of transition probability in the MSLT. The transition probability, from which all measures that are relevant in MSLT are derived, must be estimated from the data.

Exact age and age in completed years

Each of the life table functions refers to a specific age or age interval. In life table measurement, the word "age" is used very precisely, and the precision is emphasized by the addition of the modifier "exact". For instance, when we say that a person is exact age 0, we mean that person was just born. When that person is exact age 5, that individual has lived exactly five full years (i.e. 5th birthday). In contrast we can say that someone "is 5 years old," meaning that the person is between exact age 5 and exact age 6, i.e. turned 5 on the last birthday or aged 5 in completed years.

2.4.2 The Markov model

The application of probability theory, statistical inference and matrix algebra has shifted life table analysis from length of life (single decrement) to the structure of life (MSLT), from non-renewable event to renewable or repeatable events and from event of death to various types of events. In life table methodology, this change can be defined as a paradigm shift. Behind this shift, the underlying assumption is that a stochastic process generates the events (Namboodiri and Suchindran, 1987; Rajultan, 1992; Willekens, 1987; 1991) within the defined state space in the individual biography. Willekens (1991), described this shift

'as s*ituating demographic analysis as stochastic processes represents a distinct shift in life table theory from accounting or data driven approaches to one that addresses complexity and process'.*

The probabilistic structure of the MSLT estimated in this study is based on the Markov process with discrete state spaces. We assume that the occurrence of an event is an outcome of a random process. A random variable is defined by a set of possible values; which is associated with a probability. The MSLTs of this study can be described as time-inhomogeneous, finite-space, continuous-time Markov models. In technical terms, the time-inhomogeneous property means that rates of transition can vary between age intervals (Schoen, 1988a). The finite state space of the model is assumed to contain J (j=1,2,….., J) states, where J is greater than 1 and a positive integer. The J^{th} state is an "absorbing" state (e.g. "dead" state, from which there are no decrements).

We define a stochastic process {S(x): x≥0} in the state space, where the continuous time parameter x indicates the exact age attained (Schoen, 1988a). For a population, S(x) denotes the individual's position in the state space (i.e. state occupancy) at age x. This information can be summarized in terms of a set of state probabilities, P{S(x)=j}, where j is a specific state. The transition probability between two states is defined as:

$$p_{ij}[x, x+1) = \Pr\{S(x+1) = j \mid S(x) = i\} \tag{2.1}$$

where $p_{ij}[x, x+1)$ represents the probability that a person is in j at x+1 provided he or she is in i at x. The transition intensities from state i to state j are defined as

$$\mu_{ij}[x, x+\Delta x) = \lim_{\nabla x \downarrow 0} p_{ij}[x, x+\Delta x) / \Delta x \quad \text{for } i \neq j \tag{2.2}$$

where, Δx is a small time interval.

Equations (2.1) to (2.2) denote a Markov process, since the occurrence of an event of interest depends directly on only the occurrence of the state occupied and the person's age. In state space terminology, it means that a transition from one state i, the origin state, to another state, j the destination state, depends only on the origin state. This characteristic is known as the Markovian property and constitutes the basis of the construction of MSLT.

A Markov process, therefore, ignores the pathway through which the preceding state was reached. The Markov model has been applied to a variety of social phenomena such as labor mobility, attitudes change, collective violence and in the last two decades to the development of the multistate demography (Rogers, 1975; Schoen and Land, 1979; Willekens and Rogers, 1978; Hoem and Funck-Jensen, 1982; Schoen, 1988a; Crimmins et al., 1994).

2.4.3 The equations

To construct an MSLT we usually have information for a segment of life course. The event history data contains the exact time of transitions, to the month or week or day. In this study, we construct an MSLT using individual level information. First we estimate the transition probabilities. After estimating the transition probabilities other life table statistics are rather easy to calculate. The presentation of equations in MSLT is split up into two main parts. In the first part (A), we describe the estimation procedure of transition probabilities. The second part (B), describes the equations to construct an MSLT.

A. Estimation of transition probabilities

In constructing the MSLT, the first step is to estimate the age-specific transition probabilities from the data. The transition probabilities determine the changes in state occupancies. They are the ultimate parameters of the multistate models (Willekens, 2002). The estimation of these parameters depends on the observational plan. Since the life table provides information on the status of cohort members at consecutive exact ages, the ideal observational plan for life table analysis is the cohort (i.e. cohort-age) observational plan (Willekens, 1987). The formulas to estimate the transition probabilities based on the cohort data are shown in this study. We refer to Willekens (1984) for the estimation of life table transition probabilities from other types of observational plans.

In a cohort observational plan, transitions or events are recorded by age in completed years and by year of birth of the persons experiencing the transition or event (Willekens, 1987). Using micro data, there are two alternative approaches to estimate the transition probability (I) risk set approach and (II) occurrence-exposure rates.

(I) Risk set approach

In the risk set approach, the transition probability is the ratio of the number of events to the population at risk during an age interval. The population at risk is defined in Section 2.4.1 of this chapter. The number of people in state i at exact age x+1 is defined as:

$$k_i(x+1) = k_i(x) + entry_i[x, x+1) - exit_i[x, x+1) - death_i[x, x+1) \qquad (2.3)$$

where $k_i(x)$ is the number of people who reach exact age x in state i, $entry_i[x,x+1)$ is the number of people who enter in state i in age interval x to x+1; $exit_i[x,x+1)$ is the number of people who leave the state i (either make the transition to other states or (right) censored) in the age interval x to x+1 and death$[x,x+1)$ is the number of deaths in state i within that age interval. Assuming that the censoring

and new entries are uniformly distributed over the age interval x to x+1. Therefore, the risk set at age (x+1) is:

$$R_i(x+1) = k_i(x+1) - \frac{1}{2}c_i[x,x+1) + \frac{1}{2}f_i[x,x+1) \tag{2.4}$$

where, $c_i[x,x+1)$ is the number of people who are right censored within the age interval x to x+1, $f_i[x,x+1)$ is the number of people who enter into observation between x and x+1, and $R_i(x+1)$ is risk set. It is assumed that they entered in the middle of the age interval and remained in the risk set for half of the age interval.

The transition probability from origin state i to destination state j is defined as:

$$p_{ij}[x,x+1) = \frac{D_{ij}[x,x+1)}{R_i[x,x+1)} \tag{2.5}$$

where, $D_{ij}[x,x+1)$ is the number of transfers (e.g. CVD) from the origin state i to destination state j during the age interval x to x+1.

In matrix notation, the J by J matrix of transition probability can be written as:

$$\boldsymbol{P}[x,x+1) = \begin{bmatrix} p_{11}[x,x+1) & p_{21}[x,x+1) & .. & p_{J1}[x,x+1) \\ p_{12}[x,x+1) & p_{22}[x,x+1) & .. & p_{J2}[x,x+1) \\ .. & .. & .. & .. \\ p_{1J}[x,x+1) & p_{2J}[x,x+1) & .. & p_{JJ}[x,x+1) \end{bmatrix} \tag{2.6}$$

Each column in the $\boldsymbol{P}[x,x+1)$ matrix sums to 1.

(II) Occurrence-exposure rate approach

The observed transition rates are estimated by dividing the number of occurrences to the duration of exposures within the age interval x to x+1. Following equation 2.2, the observed occurrence-exposure rate from state i to state j can be defined as

$$M_{ij}[x,x+1) = \frac{D_{ij}[x,x+1)}{PY_i[x,x+1)} \tag{2.7}$$

where, $D_{ij}[x,x+1)$ represents the observed number of transfers from state i to state j between the ages of x and x+1, and $PY_i[x,x+1)$ is the risk period or person years in state i within the age interval x to x+1. Because of the Markovian assumption the observed rate applies to all persons aged x to x+1 in state i, regardless of the state they were in at exact age x (Schoen, 1988a). Now the age-specific occurrence-exposure rates can be written as a J by J matrix

$$\boldsymbol{M}[x,x+1)=\begin{bmatrix} M_{11}[x,x+1) & -M_{21}[x,x+1) & .. & -M_{J1}[x,x+1) \\ -M_{12}[x,x+1) & M_{22}[x,x+1) & .. & -M_{J2}[x,x+1) \\ .. & .. & .. & .. \\ -M_{1J}[x,x+1) & -M_{2J}[x,x+1) & .. & M_{JJ}[x,x+1) \end{bmatrix} \quad (2.8)$$

with $M_{ij}[x,x+1)$ representing the observed rate of transition from state *i* to state *j* during the interval from *x* to *x*+1. The diagonal element is expressed as:

$$M_{ii}[x,x+1) = -\sum_{j\neq i}^{J} M_{ij}[x,x+1)$$

where, J^{th} state is absorbing state. Each column in the $\boldsymbol{M}[x,x+1)$ matrix sums to 0. Defining the matrix of life table transition rates, $\boldsymbol{m}[x, x+1)$, in the same fashion, we arrive at the orientation equation

$$\boldsymbol{m}[x,x+1)=\boldsymbol{M}[x,x+1) \quad (2.9)$$

where, $m_{ij}[x,x+1)=\dfrac{d_{ij}[x,x+1)}{L_i[x,x+1)}$.

$L_i[x,x+1)$ is the number of person years spent by the members of the cohort (in the life table) in state *i* between exact age *x* and *x*+1, $d_{ij}[x,x+1)$ represents the number of transfers from state *i* to state *j* between the ages *x* and *x*+1 and $m_{ij}[x,x+1)$ is the life-table transition rate. We assume that the observed occurrence-exposure rate, $M_{ij}(x)$ is equal to life table rate $m_{ij}(x)$. The configuration of observed $\boldsymbol{M}(x)$ is similar to that of life-table $\boldsymbol{m}(x)$.

There are different alternatives to transfer the transition rates to transition probabilities, of which the linear and exponential methods are most common. The methods differ in the distribution of transitions over the age interval. The linear and exponential methods are briefly discussed here.

Linear assumption

The assumption of uniform distribution of the events over the observation interval and of births of cohort members over the interval *t-x*-1 to *t-x* leads to a piecewise linear survival function. That is, transitions occur on average in the middle of the interval. Assuming the distribution of transition is uniform within the interval, the formula yields an estimate of $\boldsymbol{P}[x,x+1)$ suggested by Rogers and Ledent (1976) and Willekens et al., (1982):

$$\boldsymbol{P}[x,x+1)=[\boldsymbol{I}+\frac{1}{2}\boldsymbol{M}[x,x+1)]^{-1}[\boldsymbol{I}-\frac{1}{2}\boldsymbol{M}[x,x+1)] \quad (2.10)$$

where $\boldsymbol{P}[x,x+1)$ is the transition matrix defined in equation 2.6, $\boldsymbol{M}[x,x+1)$ is the occurrence-exposure rates same as in equation 2.8 and $\boldsymbol{I}$ is an identity matrix. This

linear approach is applied to construct the multistate smoking status life tables (Chapter 5).

Exponential assumption

The exponential method assumes that the forces of transition or intensities are constant within the age intervals. The occurrence-exposure rates are converted to probabilities for use in the life table by the following equation:

$$\boldsymbol{P}[x, x+1) = \exp[-\boldsymbol{M}[x, x+1)] \tag{2.11}$$

Where $\boldsymbol{P}[x, x+1)$ and $\boldsymbol{M}[x,x+1)$ are the same matrices as defined before. The relation is derived from the Kolmogorov equation, which is a system of differential equations. This constant rate approach is applied in Chapter 3, Chapter 6 and Chapter 9 of this study. However, when the age interval is small the results from both the methods are equivalent.

Under very general conditions, the linear method is to be preferred on the grounds of simplicity and ease of calculation (Palloni, 2001). However, it is acknowledged that when the underlying risks are decreasing rapidly the assumption of linearity leads to a fair amount of inaccuracy (Schoen, 1988a). That may even lead to impossible negative values when some of the transition rates are very large (Hoem and Funck-Jensen, 1982; Nour and Suchindran, 1984). Therefore, the exponential method, or alternatively the so-called "mean duration of transfer method" (Schoen, 1988a), are to be preferred on the grounds of consistency. In this study, the exponential method is used for most of the applications to get more consistent estimates.

B. Construction of life table

Once we have the transition probabilities from the data, we can construct the life tables. Here we briefly illustrate the construction of a life table following the description of life table number of survivors, number of person years (i.e. sojourn time in a state), and life expectancy.

Life table number of survivors or survival probability (state occupancies)

The life table number of survivors is the number of people surviving at the beginning of the age interval x to x+1 or at exact age x. As with a standard life table, the calculation begins by the specification of a radix for the first age interval. The value of the radix is arbitrary. In demographic literature, it is usually considered to be 100,000 or 10,000 or 1000. If the beginning value is 1, it indicates the survival proportion or survival probability. Throughout this study, we use beginning value 1, i.e. the survival proportion or survival probability. Sometimes the radix population also starts from the observed prevalence in each of the health states from the

starting age of that life table (Crimmins et al, 1994). Tuma and Hannan (1984), who refer to the survival probability as the survivor function make a similar distinction. Following Rogers (1975), the relationship between survivorship values and the transition probabilities can be written as follows:

$$\boldsymbol{l}(x+1)=\boldsymbol{P}[x,x+1)*\boldsymbol{l}(x) \tag{2.12}$$

The survivorship vector $\boldsymbol{l}(x+1)$, denotes the proportion of survival in a state at exact age $x+1$.

Number of person years lived

The time spent in each state between two exact ages by a cohort member may be estimated from the life table number of survivors or survival proportion. Assuming that the events that occur in a unit interval are uniformly distributed over the interval, the average sojourn time in years spent in i^{th} state between ages x and $x+1$ can be approximated by the following expression:

$$L_i[x,x+1)=\frac{1}{2}[l_i(x)+l_i(x+1)] \tag{2.13}$$

Assuming exponential gross flow functions, the sojourn time spent in i^{th} state between ages x and $x+1$ can be approximated by the following expression (Chiang, 1984):

$$L_i[x,x+1)=\frac{l_i(x)-l_i(x+1)}{-\ln[l_i(x+1)/l_i(x)]} \tag{2.14}$$

The number of times a member of the cohort may expect to move from state i to state j between exact ages x and $x+1$ is $m_{ij}(x)L_i(x)$, where, $m_{ij}(x)$ is the occurrence-exposure rate and $L_i(x)$ is the duration of exposure in state i between ages x and $x+1$, irrespective of the state occupied at exact age x or at any prior age (i.e. population-based measure). For details of this calculation, both for population-based and health status-based measures, we refer to Willekens (1987).

To close the life table we need to assume that the population was stationary above some high age, ω. This enables us to set up the following equation (Palloni, 2001):

$$\boldsymbol{L}[\omega,\infty)=\boldsymbol{l}(\omega,\infty)[\boldsymbol{M}(\omega,\infty)]^{-1} \tag{2.15}$$

In our application, we estimate $\boldsymbol{M}(\omega,\infty)$ from the observed life expectancy at that age using the equation $\boldsymbol{M}(\omega)=\boldsymbol{e}(\omega)^{-1}$. For example, the life expectancy of the

Massachusetts population between 1989-91 is 4.55 years at age 90, so $M(90+)=1/4.55=0.22$.

Total number of years lived

The total number of years lived in state i beyond age x is equal to

$$T_i(x) = \sum_{t=x}^{\omega} L_i(t) \tag{2.16}$$

where ω is the highest age or age group. $T_i(x)$ depends on the $(0,x)$ survival probability and the time spent in each state beyond age x by a person who reaches age x.

Life expectancy

The last key measure of the synthetic cohort's life cycle disease experiences is the expectation of life in each disease or health status. The life expectancy in state i at exact age x is:

$$e_i(x) = \frac{T_i(x)}{l_+(x)} \tag{2.17}$$

where $l_+(x) = \sum_i l_i(x)$ is the number of survivors at exact age x.

This life expectancy is a population-based measure. It does not depend on the state occupied at age x.

The total life expectancy is:

$$e_+ = \sum_i e_i(x) \tag{2.18}$$

The proportion of the remaining lifetime spent in state i is $\frac{e_i(x)}{e_+(x)}$

Population-based and status-based measures

As we discussed earlier (in subsection *Types of MSLT*), the MSLT can be population-based or status-based. Population-based MSLTs show the life expectancy and health status of the entire life table population by age. Status-based MSLTs indicate the life expectancy by status at x, the reference age. The health status-based life table measures depend on the state occupied at the reference age. Their estimation is only meaningful for reference ages at which at least two states are occupied (Willekens, 1987). In status-based measures, a reference age y is selected at which at least two health states are occupied. The minimum reference age, which satisfies this condition, is generally only a few years more than the minimum age at which the transition (such as cardiovascular disease, incidence) starts. In our application, we construct population-based life tables in Chapter 2,

Chapter 6 and Chapter 7. For details of the mathematical construction and interpretation of population-based and status-based life tables we refer to Willekens (1987).

2.4.4 The introduction of covariates

Most studies that use the MSLT model and that consider covariates, stratify the population by covariate status and construct life tables for each stratum separately. For example, we construct separate life tables for males and females in order to identify their different cardiovascular life histories (e.g. Chapter 3). To discover the cardiovascular life history of smokers and non-smokers, we construct MSLTs for smokers and non-smokers separately (e.g. Chapter 6). However, to obtain estimates of the net or partial effect of a specific covariate on life table measures, for instance-life expectancy without cardiovascular disease, we also often need to control for other covariates. Therefore, the introduction of covariates in MSLT methodology has potential demographic, epidemiological and sociological applications beyond the computation of traditional "health expectancy" for a group of people.

As we have mentioned earlier (Chapter 1), one of the major aims of this study is to construct multistate life tables in a multiple-covariates context i.e. bridging the multivariate regression model and the multistate life table. In multiple covariate contexts, the age-specific transition rates or transition probabilities are predicted using multiple regression models. Later on these adjusted rates or probabilities for a group of people with specific characteristics are transferred into life table estimates. The definition of specific characteristics depends on the research objectives and interest. For example, multistate life tables are constructed for the high and low risk profiles of cardiovascular disease (Chapter 9). Once we estimate the transition rates or probabilities, the construction of life tables follows the same rules as those formulated in part B.

Recently, multistate life tables with more than one covariate have been applied in the field of work and retirement (Hayward et al., 1990), active life expectancy (Land et al., 1994) and life cycle model of labor force inequality (Hayward and Lichter, 1998). They estimated transition rates for the MSLT model using a log-linear modeling approach:

$$\begin{aligned} \ln \mu_{ij}(x) &= \exp(\beta_{0ij} + \beta_{1ij} z_1 + \beta_{2ij} z_2 + \ldots\ldots + \beta_{pij} z_p) \\ &= \exp(\beta Z') \end{aligned} \qquad (2.19)$$

where β is a row vector of p+1 regression coefficients. Z is a px1 column vector of p covariates. The covariates are all discrete variables. The effects of the covariates on the transition rates, as specified in equation 2.19 can be given an interpretation

like that of typical conventional hazard regression models. The effect of z_k, $k=1,2,3,\ldots..p$, when expressed as $\exp(\beta_k)$ is greater than 1, the transition rate is amplified by the covariate k, whereas if $\exp(\beta_k)$ is less than 1, the transition rate is depressed by variable k (Land et al 1994).

The model specified in equation 2.19 was used for the panel data where transition is assumed to occur in the middle of the two panels or intervals (i.e. interval censored) or at a discrete time point i.e. in the absence of the exact timing of an event or censoring. This type of model is called piecewise exponential or discrete time hazard model (Hannan, 1984; Land et al 1994; Hayward et al, 1998; Crimmins et al, 1994). This model was suggested and discussed conceptually by Hannan (1984). Following Hannan's suggestions and adapting the nonparametric discrete-time approach to hazard regression to the multistate, multiple-covariate context, as described by Trussell and Hammmerslough (1983) and by Guilkey and Rindfuss (1987), Hayward and Grady (1990) constructed an MSLT. Gill (1992) provided a survey of this approach and studies of associated statistical issues.

When the exact dates of transitions into and out of the states in continuous event histories are known, the estimation of hazard regression models at individual level with the computation of group-specific multistate life tables is directly applicable (Land et al., 1994). In our study, we use the FHS to construct the MSLT with multiple regression contexts. The FHS recorded disease incidence and mortality at the exact time, i.e. we have continuous event histories. Therefore, we construct hazard regression models at individual level to estimate the adjusted age-specific transition rates to construct the MSLT in covariate contexts. We use parametric and semi-parametric regression models. We describe these hazard regression models and estimations in Chapter 7.

2.4.5 Confidence interval estimation

In statistical analysis, we are not only interested in obtaining a point estimate of a statistic but also an estimate of the variation around this point estimate, and a confidence interval for the true value of the parameter. Traditionally, we depend on the central limit theorem and normal approximations to obtain standard errors and confidence intervals. These techniques are valid only if the statistic, or some known transformation of this, is asymptotically normally distributed (Efron and Tibshirani, 1993; DiCiccio and Efron, 1996). Therefore, we cannot use traditional methods if the normality assumption is not valid. For the same reason, the traditional method is not valid with MSLT. In this study, we describe a non-parametric bootstrapping approach to estimate the 95 percent confidence intervals of the multistate life table statistics.

Prior work on estimating confidence intervals for the MSLT model is relatively limited (Hayward et al., 1999). There are two basic approaches to measuring the CIs

of the outcomes of the MSLT model: analytical and numerical. Analytically, *delta method* is used to estimate the CIs (Ferrucci, 1999). For the numerical approach, *bootstrapping* is used (Calhoun, 1997; Laditka and Wolf, 1998; Hayward et al., 1999**;** Barendregt and Peeters, 2003). In this study we describe the bootstrapping method, as this procedure is easier to estimates CIs compared to analytical method.

The bootstrap method allows us to obtain an approximation of the distribution of an estimator in the absence of any priori information about the true distribution of the estimator or the original data (Efron and Tibshirnani, 1993). It is a resampling technique, which can be parametric or non-parametric. The parametric bootstrapping method employs epidemiologic knowledge about the distribution of the transition variables in the MSLT and number of cases and exposure time or person years at risk (Barendregt and Peeters, 2003). For example, the occurrence of cardiovascular disease is assumed to follow a Poisson distribution. In the non-parametric bootstrap method, *N* new samples, each of the same size as the observed data are drawn from the observed data *with replacement*. The statistic is calculated for each new set of data, yielding a bootstrap distribution for the statistic. The fundamental assumption of this method is that the observed data are representative of the underlying population (Efron and Tibshirnani, 1993). For a more detailed description of the bootstrapping method and its applicability, see Efron and Tibshirani (1993); Chernick, (1999); DiCiccio and Efron (1996). By resampling observations from the observed data, the process of sampling observations from the population is mimicked (Efron and Tibshirani, 1993). To estimate the CI of the MSLT outcome, a random sample (with same population size) with replacement from the distributions around all variables is drawn, and the model outcomes calculated. If this procedure is repeated at sufficient length (e.g. 2000 times), a distribution of the outcomes will result, from which a CI can be calculated (Hayward et al., 1999).

Non-parametric bootstrapping can be regarded as the gold standard (Hayward et al., 1999, Barendregt, 2002)**,** leading us to choose to focus on this method. . Also, when using regression to smooth the transitions by age, non-parametric bootstrapping would be the method preferred. The basic steps of this method are described here, and are similar to the methods described by Hayward et al., (1999).

Step 1: Draw a sample from the original data and calculate observed occurrence-exposure rates or statistically estimate transition rates for each possible transition in the state space.

Step 2: Use the observed occurrence-exposure rates or the parameter estimated from the statistical models to calculate a set of predicted rates (e.g. smoothed rates) for each possible transition and from these calculate the life table using the equations described in part B for each sample.

Step 3: Repeat step 1 and step 2 a large number of times (e.g. 2000).

Step 4: Calculate summary statistics (e.g. bootstrap standard deviation of 2000 replications) based on the sampling distribution of life table estimates.

Step 5: CIs are estimated from the summary statistics (e.g. the bootstrap standard deviation is used to estimate the standard deviation of the life table estimates).

Repeating steps 1 to 5, we can estimate the confidence intervals of any outcomes we estimate from the MSLT. For an application of this method we refer to Chapter 6 of this study.

2.4.6 Input data for the multistate life table

MSLT calculation depends on what types of data the researcher is going to use. The input data could be from macro to micro levels. Macro-level data refers to grouped or aggregated or tabulated data, such as the number of events during a year by the mid-year population. Micro level data refers to the information obtained from an individual, such as the smoking status of an individual at age 45. When calculating health expectancy in public health research we usually use the micro-level data from a prospective panel study to construct the life table.

Since MSLT calculations stem largely from Roger's (1975) extension of matrix algebra and its formula as a stochastic process, Schoen (1988a) states that with the exception of a few (e.g., Espenshade, 1983), most multistate projects have used census or vital statistics data. This could be attributed to the fact that at the time of MSLT inception and also during its development in the late seventies and early eighties, the event history data were not available. As a result the majority of the MSLT models have been applied to period data from census or vital registration system to estimate the occurrence-exposure rates, $\boldsymbol{M}(x)$. Rogers (1975; 1995), Willekens (1980; 1987), Schoen (1988a), Namboodiri and Suchindran (1987) documented details of these issues.

For this study, we focus on micro-level information from a prospective panel study the Framingham Heart Study. Using this longitudinal data, three opportunities arise. First, unlike most census or vital statistics registration data, FHS contains the exact time of transitions (i.e. not only the aggregated number of events over the year). If we know the exact timing of events, we can easily calculate the exact age at the time of events and the exact exposure time. Second, due to the continuous recording of the events (e.g. CVD) we are provided with continuous information for a substantial segment of the life courses. This enables us to

calculate the occurrence-exposure rates or transition probabilities directly from the different sequences or pathways in the prospective data set. The transition probabilities may be estimated directly, due to the fact that the risk set of those making a transition from state *i* to state *j* is known (Mills, 2000). After the estimation of the transition probabilities, it is easy to extend these calculations to produce age-specific transition probability matrices (Rajultan, 1992) which is the basis for the construction of MSLT. Third, due to time to time recording of the risk factor status over 48 years of follow-up, the FHS study gives us the opportunity to investigate the risk factor career using the multistate life table technique.

The first step in constructing the MSLT is to prepare the input data. The input data and the construction of MSLTs using macro level information are described in LIFEINDEC (Willekens, 1979), SPA (Willekens and Rogers, 1978), SPACE (Rogers, 1995) and LIPRO (Van Imhoff, 1994) and a few other packages. They are usually applied in multiregional demography. One point that is rarely discussed in the published literature is how to obtain the input data from the micro-level survey data set to produce the life table. Rajulton (1999) developed the package LIFHEHIST for constructing MSLTs with the help of micro data. This package constructs MSLT based on the risk set approach. The major limitation of this package is its inability to calculate the occurrence-exposure rate needed to construct an MSLT. Hence, we cannot use this package to construct an MSLT if left censoring or delay entries are present in the data. Recently, Willekens (2002) extended his LIFEINDEC to SURVEYLIFE, which allows the construction of MSLTs using micro data. However, most of the available software is neither sufficiently general nor flexible to handle a very broad class of applications or to implement alternative solutions.

For this study, we have constructed MSLT based on the long time follow-up data. In most of the cases, we combine general software packages such as S-PLUS, SPSS, STATA, EXCEL that are conducive to mixing preprogrammed routines (such as matrix inversion) with user defined subroutines. The construction of input data using SPSS syntax command is described in Chapter 3 of this study. The construction of input data and subsequently the construction of the MSLT, along with the confidence intervals of estimated life table parameters, are described in Chapter 6. The basic input data for all applications are the exact occurrence-exposure rates.

2.4.7 Output from the multistate life table

Output from the MSLT depends on the research question and objective of the study. In general, survival probabilities, sojourn times, expected years of life spent in each health or disease state and number of transitions are the major valuable output data obtained from the MSLT. The MSLT provides insight into the

processes of health change and health state structure of the estimated years in specified states. This output makes the multistate method so valuable.

The multistate life table enables an estimation of the expected magnitude of flows between health states for exact age *x* and above to be made. For example, we can calculate the expected number of moves into and out of the "cardiovascular disease" state above a certain age or between ages. These numbers are useful descriptors of individuals' experiences in health status over the life cycle.

Another important output from the MSLT connects the prevalence proportion and the incidence rate. The prevalence proportion is a product of the incidence rates and the stock of persons in each status, which in turn is a product of temporal prior behavior that has left its stamp on the composition of the population (Schoen, 1988b). This approach allows the depiction of the underlying processes that determine the prevalence proportions of health status at given age.

We illustrate the connections between prevalence proportions and incidence rates by adopting a simple life table model consisting of two "alive" states- healthy (1), diseased (2), and an absorbing state of death (3). The implied duration prevalence (IDP) (i.e. synthetic duration prevalence) of disease among persons aged *x* to *x+1*, $IDP_2[x,x+1)$, can be expressed in terms of life table notation, as follows (Schoen and Woodrow, 1980):

$$IDP_2 = \frac{L_2[x, x+1)}{L_2[x, x+1) + L_1[x, x+1)} \tag{2.20}$$

Where IDP_2 is duration prevalence and $L_i[x,x+1)$ refers to the person-years lived in the health state *i* during the interval *x* to *x+1*. Similarly, the implied prevalence (IP) (i.e. state probability of the synthetic cohort) at exact age *x* in terms of life table notation can be defined as:

$$IP_2 = \frac{l_2(x)}{l_2(x) + l_1(x)} \tag{2.21}$$

Where IP_2 is point prevalence.

Another important outcome of the MSLT can be produced using the number of person- years function (i.e. $L_i[x,x+1)$). We can estimate the life years lost to disease or number of years lived with disease of the life table population. For example, if we are interested in knowing at which age or until which age smokers spend more time with cardiovascular disease compared to non smokers, we can estimate the differences in the number of life years spent with disease by smoking status using the following equations-

$$LYSD = L_{CVD}[(x,x+1)) \text{ of smokers} - L_{CVD}[(x,x+1)) \text{ of non-smokers} \tag{2.22}$$

Where, LYSD indicates the differences of the life years spent with disease, $L_{CVD}[x,x+1)$ is the number of years lived with cardiovascular disease.

2.5 A critical assessment of the multistate life table

In this chapter, we described the multistate life table and its utility in public health. We illustrated several conceptual issues that are basic to the construction of the MSLT. The MSLT has been explained from a public health point of view, especially as a means to measure the chronic disease life history. To measure health status, we illustrate how the mathematical theory of multistate regional demography can be applied to calculate health expectancy. The fundamental difference between the MSLT approach and the Sullivan method is described. The MSLT method focuses on flows; the Sullivan method focuses on stocks. In the Sullivan method, only changes in the stock (net flows) of the population by health status and age are considered. The MSLT provides a more accurate assessment of the health expectancy in a population. We have also illustrated some of the research questions, questions in public health that can be best answered from a life course perspective. There are several methodological features of this study that may serve as an added contribution to the field of public health research, more especially to summarize the population health status and address to debate of expansion versus compression of morbidity. Some of the points can be summarized as nine basic points.

First, the debate between the Sullivan method and multistate method centers on the use of prevalence versus incidence measures of public health. Therefore, the concept of prevalence and incidence is explicitly discussed in this chapter. The precise meaning and exact application of prevalence or incidence measurement is important to get the standard estimation of a population health status. The prevalence is the basis for the construction of the Sullivan method and the incidence is the root for the multistate method.

Second, the major strength of the multistate method is its ability to capture the implications of age-related declines and improvements in health (i.e. inflow and outflow). This provides a more accurate assessment of the expected life cycle health experiences of the average person in the population. Moreover, this method allows the explicit assessment of how disability and mortality processes contribute to the structure of population health, or the changing prevalence of health problems associated with age (Crimmins, et al., 1994). These strengths have prompted some scholars to urge for the adoption of the multistate method as the methodological "standard" means of calculating active life expectancy in public health. The Sullivan method yields an inaccurate portrayal of the timing and volume of a cohort's disability and hence could mislead the theory on compression of morbidity. This is because Sullivan method does not take into account the recovery from a particular disability that results from some intervention measures. This method only deals

with the prevalence in population. In addition, the Sullivan method also can produce what appears to be a counter-intuitive relationship between changes in disability prevalence proportions and active and inactive life expectancy (Laditka and Hayward, 1999). During periods of rapidly declining mortality, Sullivan based calculations of active life remain stable relative to the growth in overall life expectancy (Bebbington, 1992; Crimmins et al., 1997).

Third, the multistate life table equations are described in a simplified way. Since we use micro data to construct an MSLT- the derivation of transition rates or probabilities is an issue of discussion. Two concepts, the occurrence-exposure rates and the observed transition probabilities are discussed. The concept of a rate differs from a concept of a proportion due to the fact that the former is a dynamic concept that depends on the risk period or exposure time. We distinguished between the risk set and the occurrence-exposure rates approach. In this study, the occurrence-exposure rates are used as the basic input data to construct the MSLT. The occurrence-exposure rates can deal with any censoring (left, right or double) that occurs in the data. Therefore, a multistate life table constructed based on the occurrence-exposure rates is more accurate than the risk set approach.

Fourth, we discussed both population-based and health status-based measures. Population–based measure describes the potential life cycle events for the whole population; status-based tables can be used to compare the perspective life cycles of those who reach specified ages in different health states. It is only in the multistate perspective that a status-based life-table analysis can be carried out, because such an analysis requires simultaneous consideration of all status-specific cohorts.

Fifth, using the multistate life table method, population heterogeneity can be dealt with in a better way than by adopting other methods. In a mortality context, Vaupel (2002) points that heterogeneity should be taken into consideration because not everyone benefits the same way from mortality improvements. That is also true for the chronic disease process. In the MSLT context, the heterogeneity problem is resolved by stratifying the population into sub-populations or groups on the basis of important attributes such as sex, socioeconomic status and risk factors. During the life course, memberships of the groups shift and alter, for instance, by changing the health status. Neither are members of the same group homogeneous; they might differ in many ways and the differences are likely to affect their chances for survival, disease process and other aspects of life.

Sixth, the introduction of multiple covariates in the context of the MSLT is a recent innovation i.e. the combination of multistate life table models with event history models has a more general appeal beyond the ordinary MSLT. For instance, we can link the MSLT of cardiovascular disease and risk factors of CVD and measure the adjusted life expectancy with disease and without disease. That results in the multivariate-multistate life tables. This type of model may be useful for health intervention (e.g. treatment exposure) and policy making. From a life course

perspective, this method will facilitate the integration of the risk factor exposure at different stages (see Figure 1.3) of life and their impact at later ages of life.

Seventh, to evaluate the significance group differences in life table estimates we discussed the non-parametric bootstrapping method. Generally, this approach allows us to estimate variances around the MSLT functions, thereby permitting formal statistical tests for group differences. Using this procedure the statistical competence of the MLST estimates could be addressed.

Eight, a Markov model ignores past experiences and examines the effect of the most recent experiences. For the chronic disease modelling, the duration of a disease is important. In that situation we need an extended Markov model, called the Semi-Markov model. The Semi-Markov model considers duration dependence i.e. the occurrence of an event depends on the length of duration between two successive events. Most-real world problems are not fully Markov in nature- they are often non-stationary, history-dependent and/or not fully observable. In order to solve such problems, we need to extend the Markovian properties that are history dependent.

Ninth, it would appear from the literature that t most of the public health researchers have shown that the multistate life table approach is data demanding. To estimate the transition probability either we need to conduct a longitudinal survey (at least two rounds or waves) or collect information from vital statistics or census in at least two time points. In the multiregional demography, MSLTs were also constructed based on the information collected in vital statistics or census (e.g. Rogers, 1975). Day by day, the number of longitudinal studies is increasing in developed countries. Therefore, the utility of the MSLT approach to measure the public health status would not be problematic in terms of data accessibility. For the developing country there could be some alternatives. We may also obtain individual health status on the basis of calendar records, which is very simple and less costly. This calendar recording can be applied to construct the period MSLT. If we measure the health status-based on the activities of daily living (ADL) or self-perceived health or self-reported health, then the calendar history recording will be an alternative to longitudinal study. In that case sampling design will not affect the MSLT results.

In this chapter, we drew an explicit distinction between prevalence and incidence based measures of population health status and shed light on the utility of the multistate life table to public health questions regarding population health status. To address this debate, researchers agreed that when the required data is available, the multistate life table approach provides a better solution than does the prevalence based method. The conceptual issues, mathematical equation, generalization, statistical competence, input and output of the multistate life table as discussed in this chapter are consistently applied throughout this book.

References

Barendregt J, Bonneux L, van Der Maas PJ, (1994). Health expectancy an indicator for change? *Journal of Epidemiology and Community Health*, 48:482-487.

Barendregt J, Bonneux L, van Der Maas PJ, (1997). How good is Sullivan's method for monitoring changes in population health expectancies? Letters to the editor. *Journal of Epidemiology and Community Health*, 51:578-582.

Barendregt J, Peeters A, (2003). Confidence intervals for Health Expectancy: a comparative analysis. Manuscript, Erasmus University Rotterdam, The Netherlands.

Bebbington AC, (1988). The expectation of life without disability in England and Wales. *Social Science and Medicine*, 27:321-326.

Bebbington AC, (1992). Expectation of life without disability measured from the OPCS disability surveys. In: Robine JM, Blanchet M, Dowd EJ, (eds.), *Health Expectancy*. HMSO, London.

Bongaarts J, (1987). The projection of family composition over the life course with family status variables. In: Bongaarts J, Burch TK, Wachter KW, (eds.), *Family Demography: Methods and Their Application*. Oxford University Press, Oxford.

Branch LG, Guralnik JM, Foley DF et al., (1991). Active life expectancy for 10,000 Caucasian men and women in three communities. *Journal of Gerontology*, 46:M145-50.

Brouard N, Robine JM, (1992). A method for calculation of health expectancy applied to longitudinal surveys of the elderly in France. In: Robine JM, Blanchet M, Dowd EJ, (eds.), *Health Expectancy*. HMSO, London.

Calhoun C, (1997). Bootstrapping the multi-state life table: preliminary results. Presented in annual meeting of the Population Association of America (PAA). Washington, D.C.

Chernick, MR (1999). Bootstrap methods a practitioner's guides, Wiley, New York.

Chiang CL, (1965). An index of health: mathematical models. *PHS Publication* No. 1000, Series 2, No. 5., U.S. Government printing office, Washington, D.C.

Chiang CL, (1979). Survival and stages of diseases. *Mathematical Biosciences,* 43:159-171.

Chiang CL, (1984). *The Life Table and Its Applications*. Krieger, Malabar, Florida.

Crimmins EM, Saito Y, Ingegneri D, (1989). Changes in life expectancy and disability-free life expectancy in the United States. *Population and Development Review,* 5(2):235-267.

Crimmins EM, Hayward MD, Saito Y, (1994). Changing mortality and morbidity rates and the health status and life expectancy of the older population. *Demography,* 31(1): 159-175.

Crimmins EM, Saito Y, (1993). Getting better and getting worse: transitions in functional status among older Americans. *Journal of Aging and Health,* 5:3-36.

Crimmins EM, Hayward MD, Saito Y, (1996). Differentials in active life expectancy in the older population of the United States. *Journal of Gerontology: Social Sciences,* 51B(3):S111-S120.

Crimmins EM, Saito Y, Ingegneri D, (1997). Trends in disability-free life expectancy in the United States, 1970-1990. *Population and Development Review.* 23(3):555-572.

Crimmins EM, Saito Y, (2000). Change in the prevalence of diseases among older Americans: 1984-1994 © 2000 Max-Planck-Gesellschaft ISSN 1435-9871. http://www.demographic-research.org/Volumes/Vol3/9.

DiCiccio TJ, Efron B, (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3):189-228.

Efron B, Tibshirani R, (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Espenshade TJ, (1983). Marriage, divorce, and remarriage from retrospective data: a multiregional approach. *Environmental and Planning A.* 15:1633-52.

Ferrucci L, G Izmirlian, Leveille et al., (1999). The effects of smoking and exercise on active life expectancy. *American Journal of Epidemiology*, 149:545-653.

Fries JF, (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine*; 303(3):130-35.

Friis RH, Sellers TA, (1999). Epidemiology for public health practice, 2nd edition, Aspen Publishers.

Gills RD, (1992). Multistate life-tables and regression models. *Mathematical Population Studies*, 3(4):259-276.

Hannan MT, (1984). Multistate demography and event history analysis. In: Diekmann A, Mitter P, (eds.). *Stochastic Modeling of Social Processes*. Academic Press, New York.

Hauet E, (1997). Practical Guide on health expectancy calculation: Sullivan method. European concerted action on harmonization of health expectancy calculation in Europe. Sub-committee on research design and calculation methods. June 1996 (revised 1997). (http:/sauvy.ined.fr/ euroreves/ methods/cookdef/cookdef.html), accessed on January, 2001.

Hayward MD, Grady WR, (1990). Work and retirement among a cohort of older men in the United States, 1963-1983. *Demography*. 27:337-56.

Hayward MD, Lichter DT, (1998). A life cycle model of labor force inequality- extending Clogg's life table approach. *Sociological Methods and Research*, 26(4):487-510.

Hayward MD, Crimmins EM, Saito Y, (1998). Cause of death and active life expectancy in the older population of the United States. *Journal of Aging and Health,* 10:1922-213.

Hayward MD, Heron M, (1999). Racial inequality in active life among older Americans. *Demography*, 36:77-91.

Hayward MD, Hsinmu C, Crimmins EM, (1999). Evaluation group differences in healthy life expectancy: the estimation of confidence intervals for multistate life table expectancies. Unpublished manuscripts. Population research institutes, the Pennsylvania state university.

Hoem I, Fong M, (1976). A Markov chain model of working life tables: A new method for the construction of tables of working life. Copenhagen, University Laboratory of Actuarial mathematics, Working paper no. 2.

Hoem J, Funck-Jensen U,(1982). Multistate life table methodology- a probabilistic critique. In: Land KC, Rogers A, (eds.), *Multidimensional Mathematical Demography*. Academic Press, New York.

Islam MA, (1994). Multistate survival models for transitions and reverse transitions: an application to contraceptive use data. *Journal of the Royal Statistical Society,* 157(3): 441-456.

Izmirlian G, Brock DB, Ferrucci L, Phillips C, (1997). Active life expectancy from annual follow-up data with missing responses. Proceedings of the American Statistical Association, Social Statistics Section:159-164.

Katz S, Branch LG, Branson MH et al., (1983). Active life expectancy. *New England Journal of Medicine*; 309:1218-1224.

Laditka SB, Wolf DA, (1998). New methods for analyzing active life expectancy. *Journal of Aging and Health*,10:214-241.

Laditka SB, Hayward MD, (1999). The evolution of demographic methods to calculate health expectancies. This manuscript was prepared for the REVES Collective Book. In: Robine JM, Jagger C, Crimmins EM, Suzman R, (eds.), *Health Expectancies at the Dawn of the Third Millennium*.

Land KC, Guralnik JM, Blazer DG, (1994). Estimating increment-decrement life tables with multiple covariates from panel data: the case of active life expectancy, *Demography,* 31(2):297-319.

Lalu NM, (1992). Multistate life tables and its application to the analysis of labour force participation. In: Krishnan P, Tuan CH, Mahadevan K, (eds.), *Readings in Population Research: Policy, Methods, Perspectives*. BR Publishing, Delhi, India.

Liu X, Liang J, Muramastsu N, Sugisawa H, (1995). Transition in functional status and active life expectancy among older people in Japan. *Journal of Gerontology/Social Sciences,* 50B:S383-S394.

Mamun AA, (2001). *Multistate Models in Public Health- Review and Application to the Framingham Heart Study*. Master thesis series, No. 01-3. Population Research Centre, University of Groningen, The Netherlands.

Mathers C, (1991). Health expectancies in Australia 1981 and 1988. Technical report, Australian Institute of Health: AGPS, Canberra.

Mathers DC, Robine JM, (1994). Health expectancy indicators: recommendations for terminology. Seventh International Meeting of REVES (International Network on Health Expectancy), Canberra, 23-25 February 1994.

Mathers CD, Robine JM, (1997). How good is Sullivan's method for monitoring changes in population health expectancies? *Journal of Epidemiology and Community Health*, 51:80-86.

Mills M, (2000). *The Transformation of Partnerships-Canada, the Netherlands, and the Russian Federation in the Age of Modernity.* PhD dissertation, University of Groningen.

Moriyama IM, (1968). An Indicators of social change. In: Sheldon EB, Moore WE (eds.), *Problems In The Measurement Of Health Status*. Russell Sage Foundation, New York.

National Research Council (1988). *The Aging Population in the Twenty-first Century: Statistics for Health Policy*. National Academy Press, USA.

Namboodiri K, Suchindran CM, (1987). *Life Table Techniques and Their Applications.* Academic Press, Orlando.

Niessen L, (2002). Roads to Health Multi-state Modelling of Population Health and Resource Use. PhD Dissertation, University of Groningen. Rozenberg publishers, Amsterdam.

Nour ES, Suchindran CM, (1984). The construction of multistate life tables: comments on the article by Willekens et al., *Population Studies*, 48(2):193-222.

Nusselder W, (1998). *Compression of Morbidity? A Life-table Approach*. PhD Dissertation, Erasmus University, Rotterdam, The Netherlands.

Palloni A, (2001). Increment-decrement life tables. In: SH Preston, P Heuveline, M Guillot (eds.) *Demography- Measuring and Modeling Population Process.* Blackwell Publishers, Malden, Massachusetts.

Pressat R, (1985). The Dictionary of Demography. In: Wilson C, (eds.), Blackwell Reference, Bell and Bain Ltd. Glasgow.

Preston SH, (1993). Health indices as a guide to health sector planning: a demographic critique. In: Gribble JN, Preston SH, (eds.), *The Epidemiological Transition. Policy and Planning Implications for Developing Countries.* National Academic Press, Washington D.C.

Rajulton F, (1992). Life history analysis. Guidelines for using the program LIFEHIST. Population Studies Center, University of Western Ontario, Canada.

Rajulton F, (1999). LIFEHIST: Analysis of life histories, A state space approach. http://www.sscl. uwo.ca/sociology/longitudinal/Approaches.htm, accessed on December 2000.

Rogers A, (1973). The multiregional life table. *The Journal of Mathematical Sociology,* 3: 127-137.

Rogers A, (1975). *Introduction to Multiregional Mathematical Demography*. Wiley, New York.

Rogers A, Ledent, (1976). Increment-decrement life tables: A comment. *Demography*, 13:287-290.

Rogers A, Willekens FJ, (1986). *Migration and Settlement: A Multiregional Study*. Reidel: Dordrecht, The Netherlands.

Rogers RG, Richard A, Belanger A, (1989). Active life among the elderly in the United States: Multistate life-table estimates and population projections. *Milbank Quarterly*. 67:370-411.

Rogers A, Rogers RG, Branch LG, (1989). A multistate analysis of active life expectancy. *Public Health Reports*, 104(3):222-226.

Rogers A, Rogers RG, Branch LG, (1990). Longer life but wrote health? Measurement and dynamics. *The Gerontologist,* 30:640-649.

Rogers A, (1995). *Multiregional Demography. Principles, Methods and Extensions*. Wiley, Chichester.

Rothman K, Greenland S, (1998). Measures of disease frequency. In: Rothman K, Greenland S, (eds.), *Modern Epidemiology*. Lippincott Williams and Wilkins, Philadelphia.

Rusnak M, Scherbov S, Cider B, (1992). Mortality and morbidity projections: Lung Cancer. In: Morgensten W, Chigan E, Prokhorskas R, Rusnak M, Schettler G, (eds.), *Models Of Noncommunicable Diseases- Health Status and Health Service Requirements*. Published in collaboration with the World Health Organization. Springer-Verlag, Berlin Heidelberg.

Sander B, (1964). Measuring community health level. *American Journal of Public Health*, 54: 1063-1970.

Saito Y, Crimmins EM, Hayward MD, (1999). Health expectancy: an overview. *NUPRI* research paper series no. 67. Nihon University population research institute, Tokyo, Japan.

Schoen R, (1975). Constructing increment-decrement life tables. *Demography*, 12:313-24.

Schoen R, Land K, (1979). A general algorithm for estimating a Markov-generated. increment-decrement life table with applications to marital status patterns. *Journal of the American Statistical Association*, 74:761-776.

Schoen R, Woodrow K, (1980). Labor force status life tables for the United States. *Demography*, 17:297-322.

Schoen R, Woodrow K, (1984). Marriage and divorce in twentieth-century Belgian cohorts, *Journal of Family History*, 9:88-103.

Schoen R, (1988a). *Modelling Multigroup Populations*. Plenum Press, New York.

Schoen, R, (1988b). Practical uses of multistate population models. *Annual Review of Sociology,* 14:341-61.

Scott J, Alwin D, (1998). Retrospective versus prospective measurement of life histories in longitudinal research. In: Giele JZ, Elder Jr JH, (eds.), *Methods of Life Course Research*. Sage, Thousand Oaks.

Scott GR, Filerman GL, Lesar JW, (2000). Attaining global health challenges and opportunities. *Population Bulletin*, 55(1). A publication of the population reference bureau. http://www.prb.org/Content/NavigationMenu/PRB/AboutPRB/Population_Bulletin2, accessed December 2002.

Selvin S, (1991). *Statistical Analysis of Epidemiologic Data*. Monographs in epidemiology and biostatistics. Vol. 17. Oxford University Press, Oxford.

Singer BS, Manton K, (1994). What is the fuss about compression of morbidity? *Chance*, 7(40):21-30.

Sullivan DF, (1966). Conceptual problems in developing an index of health. Vital and Health Statistics, Series 2 (17), National Center for Health Statistics, Washington, D.C.

Sullivan DF, (1971a). A single index of mortality and morbidity. *HSMHA Health Reports*, 86:347-354.

Sullivan DF, (1971b). Disability components for an index of health. Office of health statistics. Data Evaluation and methods research, Series 2, No. 42. Washington, D.C.

Tuma N, Hannan M, (1984). *Social Dynamics: Models and Methods*. Academic Press, San Diego.
Vaupel JW, (2002). Life expectancy at current rates vs. current conditions. A reflection stimulated by Bongaarts and Feeney's "How long do we live?" *Demographic Research*, 7, article 8.
Van Imhoff E, (1994). LIPRO user's guide. Version 3.0. Working Paper no. 1994/A1, NIDI, Voorburg, The Netherlands.
Willekens FJ, Rogers A, (1978). Spatial population analysis. Methods and computer programs. IIASA, Research Report RR-78-18.
Willekens FJ, (1979). Computer program for increment-decrement (multistate) life table analysis: a user's manual LIFEINDEC, working paper WP-79-102, IIASA, Laxenburg, Austria.
Willekens FJ, (1980). Multistate analysis of tables of working life. *Environment and Planning*, 12:563-588.
Willekens FJ, Shah I, Shah JM, Ramachandran P, (1982). Multi-state analysis of marital status life tables: theory and application. *Population Studies*, 36(3):129-44.
Willekens FJ, (1987). The marital status life table. In: Bongaarts J, Burch T, Wachter KW, (eds.), *Family Demography: Models and Applications*. Oxford University Press, Oxford.
Willekens FJ, (1991). Life table analysis of staging process. In: Becker HA, (eds.), *Life Histories and Generations*. ISOR, Volume II. University of Utrecht, The Netherlands.
Willekens FJ, (1991). Understanding the interdependence between parallel careers. In: Singers JJ, de Jong-Gierveld J, Van Imhoff E, (eds.), *Female Labor Market Behavior and Fertility: A Rational-choice Approach*. Spinger-Verlag, New York.
Willekens FJ, (1999). The life course: models and analysis. In: van Wissen LJG, Dykstra PA (eds.), *Population Issues: an Interdisciplinary Focus*. Kluwer Academic/Plenum Publishers, Dordrecht.
Willekens FJ, (2002). "Life Histories: SURVEYLIFE", manual written for the use of multi-state life table. Population Research Center, University of Groningen.
Willekens FJ, (2002). Forecasting the life course. Paper presented at the Population Association of America, May 9-11, 2002, Atlanta.
Willekens FJ, (2002). Multistate demography. *Encyclopaedia of Population*. Revised edition, Paul Demeny and Geoffrey McNicoll editors Macmillan Reference USA, New York.
Willkins R, Adams OB, (1983). Health expectancy in Canada, late 1970s: Demographic, regional, and social dimensions. *American Journal of Public health,* 73:1073-80.
World Health Organization, (2000). WHO issues new healthy life expectancy rankings. Press Releases WHO, Released in Washington, D.C. and Geneva, Switzerland 4 June 2000. (http://www.who.int/inf-pr-2000/en/pr2000-life.html).
World Health Organization, (2000). International Classification of Functioning, Disability and Health. Prefinal draft. Full version, December 2000. Internet: http://www.who.int/ icidh/ICIDH-2PFDec-2000.pdf.
Wolf DA, Laditka SB,(1997). Stochastic modeling of active life and its expectancy. Papers in Microsimulation series paper no. 4, ISSN 1084-1695, The Maxwell Center for Demography and Economics of Aging, Syracuse University.
Wolfbein SL, (1949). The length of working life. *Population Studies*, 3:286-294.
Young TK, (1998). *Population Health: Concepts and Methods*. Oxford University Press, New York.
Zeng Yi, (1991). *Family Dynamics in China. A life table analysis*. Madison: University of Wisconsin Press. National Academy Press, USA.