Capacity building for longitudinal research

Frans Willekens

1. Introduction

The raison d'être of longitudinal research is a better understanding of developmental processes and in particular the factors (internal and external) and the mechanisms that shape the processes. A process may result in a single outcome or several outcomes and the way it evolves may be influenced by several factors. In behavioural and social sciences, the processes of interest relate to human development and the life course. We want to know *what* is influencing the lives of individuals, their families, and their households. We also want to know *how* they are affected. These questions apply to different domains of life such as family life, work, health, education and income. A factor that influences the lives of people may have an immediate effect but, more often, the effect is felt with a time lag or it is spread out over several years. The effect is likely to be mediated by other factors in the same or different domains of life. The lifetime effect(s) of a given factor in the presence of other factors that change over time is (are) dependent on how factors interact and how the interactions evolve over time.

Observations on factors and their effects in the presence of other influencing factors are usually too short to evaluate lifetime consequences. In that case, observations on different cohorts are combined into synthetic lives or synthetic biographies that are fully consistent with the observations during the observation window. Observations may be augmented by data from other sources, including other studies, administrative records and domain experts. The challenge of longitudinal research is to understand and be able to predict the lives of individuals, families and households in specific settings, and to determine how lives are affected by events, experiences, conditions, regulations, policies and other interventions. In short, the challenge is to comprehend life as a whole, embedded in its context. Jessor et al. formulate it neatly: "Understanding the integrity of the life course, tracing its continuity over large segments of time, distinguishing what is ephemeral from what is lasting, grasping the role the past plays in shaping the future – al these, and more, are issues that yield only to

research that is longitudinal and developmental in design." (Jessor et al., 1991, p. 3; quoted by Cairns and Cairns, 2002, p. 289). Cairns and Cairns conclude their contribution to the book "*Looking at lives. American longitudinal studies in the twentieth century.*" with the observation that the interdisciplinary study of human beings over time and generations requires a new way of thinking and conducting research. Concepts and procedures extend beyond the disciplines from which studies originally emerged. A new discipline is emerging, which they call *developmental science*[1] (Cairns and Cairns, 2002, p. 292). In this paper, we approach longitudinal research from a developmental science perspective.

The life course is a developmental process and, as any other developmental process, it is embedded in a context with social (including institutional), cultural, economic, political, historical and technological aspects. The structure of the developmental process and the direction of change are outcomes of the combined effect of nature and nurture. Longitudinal studies track people's lives over relatively long periods of time to determine what factors influence the lives and how the effects are mediated by attributes of the individual and/or the environment. Most longitudinal studies observe people in a real setting (observational studies) and not in an experimental setting in which conditions are identical except for the factor(s) the effect of which needs to be determined[2]. Longitudinal studies are described by some authors as quasi experiments. Although no randomized allocation of individuals to treatments exists, the temporal sequencing of events does offer a means of control in comparison with a cross-sectional survey or sequences of cross-sectional observations of different persons of the same cohort, such as in a series of censuses. A dominant challenge in longitudinal research is to identify how and why given events, experiences or interventions affect different people differently and how and why the same individual in different situations responds differently to the same event, experience or intervention. Many longitudinal studies are designed to track the consequences of interventions such as government regulations and social programmes (see e.g. Buck, 2002). The understanding of human behaviour based on longitudinal research will further the development of public policies that are

[1] The journal *Developmental Science* has a much more narrow focus: human developmental cognitive neuroscience. The scope of *Applied Developmental Science* is considerably broader. It adopts the view that individual and family functioning is a combined and interactive product of biology and the physical and social environments that continuously evolve and change over time. In July 2004, a special issue was published on the impact on lives of the September 11th 2001 terrorist attack.

[2] In the hierarchy of research designs, the results of randomized, controlled trials are considered to be evidence of the highest grade, whereas observational studies are viewed as having less validity because they reportedly overestimate treatment effects. Concato et al. (2000) performed a meta-analysis of studies assessing the same intervention but with different research designs. They found that the average results of the observational studies were remarkably similar to those of the randomized, controlled trials. They concluded that well-designed observational studies (with either a cohort or a case–control design) do not systematically overestimate the magnitude of the effects of treatment as compared with those in randomized, controlled trials on the same topic.

more responsive to individuals' experiences across their lives (Bernard et al., 2005, p. 3)..

The objectives of longitudinal research and biographical analysis are: description, explanation, comprehension, prediction and effect analysis (or impact analysis). Description involves the *detection of a pattern* in the timing and sequencing of life events. It determines the particular 'life structure'. The second objective is to determine whether and how different life events are related, i.e. to *explain* the 'life structure'. The dominant method is statistical association. The third objective is to comprehend the life course by discovering causal links, i.e. to identify underlying elementary event processes and to describe how pathways and structure emerge from interactions among processes. The fourth objective is to *predict* or reconstruct life histories of individuals or cohorts (or other groups of individuals) from partial observations. In some fields, e.g. medicine and epidemiology, prognosis is an important dimension of life history analysis. The fifth objective is to determine, for a given individual or for an individual with a given set of attributes, the likely effect of an intervention compared with another. Many practical applications are limited to a single outcome and the factors affecting the outcome. More complex studies consider multiple outcomes or pathways, i.e. sequences of outcomes that may affect each other.

Capacity building for longitudinal research involves description, explanation, comprehension, prediction and effect analysis. The lives of people are shaped by genetic factors, antecedents or historical factors, and contemporary factors. They are predispositions, endowments, childhood experiences, intervening factors, cumulative causation, living conditions, social and cultural embedding, institutional and technological context, accumulation of wealth and social capital, and values. Some of these factors are person-centred while others are experienced by many; they are collective experiences. Lives are also shaped by needs and motives; aspirations, expectations, anticipations and prospects; opportunities and constraints; and many contingencies. How these factors and processes operate and interact in shaping the course of life requires a theoretical model that encompasses the factors and processes. The theoretical model constitutes the main object of life course theory. A comprehensive life course theory does not yet exist, however.

The paper is organized as follows. Section 2 presents a brief account of the life course as a developmental process. It serves as a theoretical basis for empirical longitudinal studies of the human life course. Capacity building should prevent the isolation of data analysis from established and new theories. Data collection and theory development are two parallel processes that call for coordination and synchronization. Section 3 presents a few ideas on what can be done to obtain a

cadre of researchers that is comfortable with longitudinal research. Section 4 concludes the paper.

2. Developmental processes: a brief note

The human life course may be viewed as a developmental process. Developmental processes are characterised by continuity and change, or periods of stability and periods of transition. Processes can be identified and studied at different levels of aggregation, from the population level in economics, sociology and demography, to the individual level in psychology, education, and evolutionary economics, and the cell level in developmental biology. All developmental processes have important features in common which points to a meta-theory of developmental processes. The processes consist of relatively stable periods that are connected by periods of transition. During transitional stages, pre-conditions are created for take-off to the next stage and the *developmental readiness* that results is a major factor in progression to the next stage. Some periods are critical to the future of the process and events during these periods may have long-term implications and may even determine the development path. Another feature that all developmental processes have in common is the effect of history or path dependence. Usually several developmental processes are present simultaneously and they interact. Synchronization and other coordination mechanisms result in an harmonious development of the organic whole, provided they work properly (for details and an application in demography, see Willekens, 1991).

Life is a composite developmental process. Erikson (1982, p. 25) asserts that a human being's existence at every moment depends on three processes of organization that must complement each other. They are biological processes that govern the development of the organism (soma), psychic processes that govern the identity development and ego synthesis (psyche), and social processes that govern the interdependence of persons. In order to understand life, we must understand how these processes interact and how they respond to the environment and help create a suitable environment. By separating the processes to address disciplinary research questions, the interdependence is lost and at most partial insights can be obtained. We will never understand the lives of people as long as we separate the different dimensions of life into homo economicus, homo sociologicus, etc. What is needed is a comprehensive life course theory. No such theory exists but building blocks do exist, such as notions of (1) causal pathways, resulting from risk accumulation over the life course, (2) critical periods and sensitive periods (Kuh and Ben-Shlomo, 1997; Arthur, 1994), (3) positive feedbacks, path dependence and lock-ins (Arthur, 1994). A building block is also the empirical evidence that most episodes of poverty, unemployment or disease are short-term. At the abstract level, a comprehensive theory of the life course is a generic theory of a developmental process. It

represents an integrative perspective on human development in a changing world, connecting ontogenetic, historical, and institutional perspectives. Such a comprehensive approach is practiced in the LIFE Research School organized by the Max Planck Institute for Human Development in Berlin (http://www.mpib-berlin.mpg.de/en/life/index.htm) and the Developmental Science programme at the University of Iowa (http://www.psychology.uiowa.edu/devscience/index.html). A comprehensive theory is inspired strongly by biology.

Few people will dispute that the development of an individual is in part directed by his genetic code (*ontogenesis*) and in part by the interaction with the environment. Three classes of factors need to be distinguished to understand the lives of people: the innate characteristics (genetics, heredity), the contextual factors (environment) and the acquired characteristics that result from the interaction between the genetic code and the environment (*epigenesis*). Today, epigenetics is changing our evolutionary thinking after the discovery that context is crucial for gene expression.[3] In the 1950s Erikson introduced the epigenetic principle in the study of psychosocial development. The principle says that we develop through a predetermined unfolding of our personalities in stages. Our progress through each stage is in part determined by our success, or lack of success, in all the previous stages and that success is partly determined by our environment. The phenomenon is known as path dependence and cumulative causation. Epigentic rules are "the algorithms of growth and differentiation that create a fully functioning organism". (Wilson, 1998, p. 150; quoted by Machalek, 2004)).

One acquired characteristic of fundamental importance in understanding behaviour and behavioural patterns of individuals, and hence the life course, is the mental model or models individuals develop to interpret the world, to attach meaning to events and experiences, and to direct their attitude, opinions and behaviour. As people interact with the environment and exchange information, they develop a mental model that represents a coherent and internally consistent view of the world and that enable them to interpret the world as it is experienced. At the same time, however, it constrains the ability to comprehend new and unfamiliar experiences and events. Experiences are always filtered through the current ways of understanding. The *organization* of knowledge in cognitive schemes and the *adaptation* of these schemes to the demands by the

[3] The term epigenetics was coined by Waddington in 1940s for the causal study of embryological development. For him epigenetics was the link between the genotype, which belonged to the geneticist, and the phenotype, which in the evolutionary context belonged to the taxonomist. Modern epigenetics is primarily concerned with the mechanisms through which cells become committed to a particular form or function, and that functional or structural state is then transmitted in cell lineages.

environment are two complementary processes underlying cognitive development (Piaget, 1952; quoted by Miller, 1983, p. 71). The first is an internal process and the second relates to the interaction with the environment. Adaptation involves assimilation (fitting information into one's current cognitive scheme) and accommodation (adjustment of the scheme in order to accommodate discrepant information). If new information does not fit into the scheme, the information is disregarded. As a result, cognitive schemes act as a filter and they make individuals selectively responsive to information that is congruent with their self-view and their world-view (D'Andrade, 1986; Roberts and Caspi, 2003).

This theoretical perspective on the life course has consequences for longitudinal research. Data on ***life histories***, i.e. timing and sequences of life events ('the string of facts in chronological order'), must be augmented by life stories. ***Life stories*** are oral histories that enable the understanding of the individual's interpretation and the extent at which reality is real or constructed (perceived). Life stories are individual and social constructs that may involve considerable biographical illusion (Bourdieu, 1986). The life history and the life story represent two perspectives on the life course. The life history is an etic construct, i.e. a description of life using the cognitive scheme (concepts and categories) of the scientific observer. The life story is an emic construct, i.e. a description of the life course in terms of the concepts and categories that are considered as meaningful and appropriate by the individual being observed. Life stories are particularly useful to explore the subjective meanings individuals attach to events, experiences and situations (for details, see Willekens 2006). Abbott adds that plausible narratives, i.e. life stories, about particular individuals, are required to justify the entailed relation between variables in statistical analysis (Abbott, 2001, pp. 132).

During the past decades the life course paradigm, which approaches events, experiences and behaviour as embedded in the life course and which approaches life as a developmental process, emerged across disciplines including sociology (Elder, 1999), political sciences (Heinz and Krueger, 2001), psychology (Elder and Kirkpatrick Johnson, 2000), psychiatry (Kaplan and Sadock, 1981, chapter 1), criminology (Macmillan, 2001; Sampson and Laub, 2005) and anthropology (Linde, 1993). In recent years, the life course perspective emerged significantly in epidemiology and public health (Kuh and Ben-Shlomo, 1997; Barker, 1998; Ben-Shlomo and Kuh, 2002; Kuh and Hardy, 2002; WHO, 2002 and Halfon and Hochstein, 2002; Osler, 2006). In health sciences, there is considerable interest in understanding the effects of early-life experiences and lifestyle on diseases at older ages and the factors and interventions that prevent, postpone or slow down these diseases. In demography the interest is on the effect of early life

conditions on adult mortality. In sociology, the lifetime consequences of early life transitions constitute a major subject of study.

The increased individual freedom of choice in parts of the world raises new life course issues never witnessed before. In 1990, Légaré and Marcil-Gratton expressed the vision that, as the controls over reproduction and death are being transferred to the exclusive prerogative of the individual, 'Individual programming of life events such as childbirth and death could truly be a privilege of the individual human being in the next century.' (Légaré and Marcil-Gratton, 1990, p. 104). An equally important view was expressed by Held: 'marriage and procreation may come to be seen as experiences of an exploring, growing self and not longer as role and status positions.' (Held, 1986, p. 162). Longitudinal research is challenged to integrate these social changes in studies of the lifepaths of individuals in varying contexts.

3. Capacity building

"If important and relevant insights are to be generated through longitudinal data, it will be by skilled researchers. If insufficient research capacity is brought to bear on longitudinal surveys, a shortfall of relevant findings will result." (Picot and Webber, 2005).

In this section, five suggestions are made on what may need to be done to obtain a cadre of researchers that is comfortable with longitudinal research.

1. *Train researching in asking the right research questions*

Researchers who have access to longitudinal data may not make optimal use of the data because they do not ask the research questions these surveys are designed to answer or are able to answer. Better life course theory is expected to enhance the ability to ask the right questions. It is felt that a right question is one that approaches the life course as a developmental process and that aims at uncovering causal mechanisms rather than statistical associations.

2. *Train researchers in using the full potential of longitudinal data by studying multiple events and transitions between multiple transient states*

Many studies that use longitudinal data focus on a single event. They are concerned with the effect of personal attributes (current and during earlier stages of life), institutional factors and other contextual factors on the time to event using models of duration data (duration models, survival models. Event history analysis, the label used in sociology, may use data on event histories (sequences of events recorded retrospectively), but generally address questions about a single event (treated as nonrepeatable event) and disregard the potential of longitudinal data to study repeatable events and sequences of events (pathways). Influential textbooks in the field, Yamaguchi (1991) and Blossfeld and Rohwer

(2002), focus on the timing of single transitions or nonrepeatable events, although both have some discussion on repeatable events (Two-way transitions in Chapter 3 of the Yamaguchi book and models with multiple episodes in Chapter 4 of the Blossfeld and Rohwer book) and include models that incorporate the effect of events in the same or other domains of life (included in models with time-varying covariates). The same applies to textbooks in econometrics such as Lancaster (1990).

Duration models and other models of a single event may be considered as transition models or models for transition data analysis with two states only (origin 0 and destination 1). The occurrence of the event implies the transition from origin to destination, e.g. from unemployment to employment. This two-state perspective is implicit in most duration models and is made explicit by Blossfeld and Rohwer (2002). Models with multiple states were introduced in the social sciences long ago (e.g. Bartholomew, 1982; Tuma and Hannan, 1984), but they are not much used. The reasons are not clear. Software availability may be an explanation (the package RATE by Tuma has been replaced by other packages such as TDA by Rohwer [http://steinhaus.stat.ruhr-uni-bochum.de/tda.html]).

Whereas event history models do not consider the whole sequence of multiple transitions between states simultaneously, sequence analysis does. Sequence analysis studies whether sequences of events typically happen in a particular order (Abbott, 1985, 2000; Abbott and Tsay, 2000; Billari and Piccarreta, 2005). Such questions are central in studies of careers, such as occupational careers, criminal careers, marital careers, etc.. A representative application is by Malo and Muñoz–Bullón (2003). The package TDA includes a procedure for describing sequence data. Sequence analysis is very different from regression analysis and other variable-based methods of analysis. In essence, it is a pattern-based method of analysis (Abbott, 2001, p. 260). It aims at detecting patterns in data.

Multistate transition models consider simultaneously transitions among several states of existence. Although these models have been around for decades and are documented extensively by Bartholomew (1982), Tuma and Hannan (1984), Rogers (1975) and others, they are not yet used much in longitudinal research in the social sciences. Hougaard (2000, p. vi) asserts that models with multiple states (multistate models) are 'the most classical way of analysing life history data'. Multistate models are much more common in epidemiology and public health (for reviews, see Hougaard, 1999, and Commenges, 1999). Journals such as *Statistics in Medicine* publish regularly methodological innovations and new applications. An illustration of the potential of these models in the study of short-term and long-term effects of different interventions is the paper by Pérez-Ocón et al., (2001). The authors study the impact of different treatments and combinations of treatments on the chances for survival from breast cancer after

breast surgery. The effect of treatment is incorporated in transition models via the transition rates. Alternative ways of incorporating interventions are presented by Brookmeyer and Gray (2000).

3. *Train researchers in life-table and simulation techniques to generate synthetic biographies*

The dependent variable in statistical models of time to event data, duration data or event history data is the transition probability (logit model, logistic regression model, probit model) or the transition rate (transition rate models such as the exponential model, the Weibull model, the Gompertz model and the Cox model). These models quantify the effects of personal attributes (covariates) and other factors on the probability of a transition during a unit time interval, the rate of transition and ultimately on the time to transition. The effects are likely to vary with age and may vary with other duration variables (e.g. duration of marriage, duration of unemployment). To determine the long-term and even lifetime consequences of events, experiences and conditions, the data and the regression models are not enough. To be able to determine what to expect if the probabilities and rates prevail and what to expect when personal attributes or other factors change, simulation models are needed. Wolfson (2005) gives a fine illustration of the questions models can answer that data alone cannot. In what follows, some additional illustrations will be given. Models summarize the information in the raw data but enable the researcher to go beyond the data. For instance, the dependence of a disease (e.g. lung cancer) on a risk factor (e.g. smoking) in the presence or absence of other factors may be determined by a transition rate model or another regression model. The model does not provide information on a most important health measure, the lifetime probability of the disease. The lifetime probability may be derived from the incidence rate(s), which are transition rates, but the derivation implies the specification of a model of the life course and consequently at least one assumption. The common model specification, which is usually not made explicit however, is whether an individual who did not yet get the disease may some time in the future die from another disease. In the absence of competing causes of death, the lifetime probability is obtained as the cumulative or lifetime incidence. In the presence of competing causes of death, it is obtained using the life table. Even today, a relatively simple concept as the lifetime probability of an event is not adequately understood (see e.g. De Backer and Bacquer, 1999). The problem is often discussed in the medical literature. For an interesting discussion of the discrepancy between the number of AIDS cases predicted and the number reported in France and the UK, see Rude et al. (1993).

The lifetime probability of an event, such as a disease, divorce or unemployment, is only one of the measures used in studies that cover the lifespan. The measure cannot be obtained from the data directly. It requires a model of the life course

that considers competing events. The (multistate) life table is commonly used. Another measure that requires the life table is the life expectancy for individuals with specific risk-factor profiles. Bonneux (2000) uses the model to study the impact of cholesterol-lowering therapies on the risk of cardiovascular disease and the life expectancy for smokers and non-smokers. The lifetime derived from the life table is the *expected* lifetime if the age-specific incidence rates of a disease and age-specific death rates that have been observed during a period of observation apply to a synthetic cohort of individuals with given characteristics. Life tables translate sets of age-specific rates into life-course measures such as lifetime probability of an event, probability that an event triggers another event, life expectancy, expected duration of episodes between two events, expected durations of stages of life, etc.. The life table, in particular the multistate life table, is increasingly being used to predict life trajectories for people with given risk factors and risk profiles. Peeters et al. (2002, 2003) studied the impact of obesity, smoking and other risk factors on the total life expectancy and the expected numbers of years with cardiovascular disease (CVD). The years of life lost due to obesity are also estimated using a different method by Fontaine et al. (2003). Mamun (2003) and Mamun et al., (2004) studies the impact on years with and without CVD. For a quick overview of multistate models from a more demographic perspective, see Willekens (2003) and for a technical introduction to multistate models in life course research, see Willekens (2005)[4].

The life table produces expected values of sojourn times. Individuals with the same attributes (and histories) have the same transition rates and transition probabilities, unless random effect models have been specified where the regression coefficients vary across individuals. Except for random effect models, individuals with the same attributes have the same *expected* sojourn times and individual variation in sojourn times is the effect of chance. Microsimulation techniques may be used to determine individual life trajectories that are consistent with observed probabilities and rates of transition. Most microsimulation models start from transition probabilities, defined for discrete time periods. A few models, including SOCSIM of the University of California at Berkeley (http://www.demog.berkeley.edu/~wachter/socstory.html) and LifePaths of Statistics Canada (http://www.statcan.ca/english/spsd/LifePaths.htm) start from transition rates

[4] The report is part of the MicMac project. The aim of the project is to develop a multistate model that bridges the gap between micro- and macro-level analysis in demographic forecasting. MicMac is being developed by a team of researchers from the Vienna Institute of Demography (VID), the Institut National d'Etudes Demographiques (INED), Paris, Bocconi University, Milan, Erasmus Medical Centre, Rotterdam, Max Planck Institute for Demographic Research, Rostock, International Institute for Applied System Aanalysis (IIASA), Laxenburg, and University of Rostock, under the leadership of the Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague. The project is funded by the European Commission under the 6th Framework Programme. For details, see the MicMac website www.micmac-projections.org

and the associated waiting times to the event. The second approach has a number of advantages. First, simultaneous events can be handled without additional assumptions (using the theory of competing risks). Second, the length of intervals between events can be estimated precisely instead of approximately.

4. *Organize general courses on longitudinal data analysis and specialized courses on specific data sets, statistical models and/or simulation models*

Many social scientists show an interest in longitudinal data analysis but do not have the adequate training. Remedial courses and general-purpose courses may be necessary. To make the courses most effective, they may be short courses organized at different locations. In the case of Canada, one might consider having the courses at the Research Data Centres (RDCs) established throughout Canada by Statistics Canada, the Social Sciences and Humanities Research Council and other organizations such as universities. To make the courses most effective, public-use longitudinal data sets should be made available. They may be subsets of full data sets and they may include a few variables only. They should be realistic files corresponding to the master files and one should be able to use the subset outside the RDCs. That allows researchers greater flexibility in terms of where and when they do their work (see also Spencer, 2005). Researchers are required to make a relatively large initial investment to become familiar with complex longitudinal surveys (Picot and Webber, 2005, p. 16). Training provided by or in cooperation with Statistics Canada may enhance interest.

To enhance the application of microsimulation in longitudinal research, the availability of a simplified and generic (general-purpose) version of LifePaths is likely to stimulate interest. Microsimulation is generally presented as complex and inaccessible. Although most microsimulation models may be inaccessible, the technique of microsimulation is increasingly accessible to everyone thanks to the range of random number generators that are included in statistical software including SPSS. It is interesting to note that some authors view microsimulation not as modelling but as the generation of data. Wolf sees microsimulation fundamentally as an exercise in sampling: "Microsimulation consists of drawing a sample of realizations of a prespecified stochastic process." (Wolf, 2000, p. 2). The same view is adopted in the MicMac project mentioned earlier.

Life tables and simulation models to generate synthetic biographies do not always need specialized software. Abridged life table, even multistate life tables, can be developed relatively easily in Excel, provided the transition probabilities or the transition rates have been estimated using statistical procedures in standard statistical packages. Stochastic life tables can be estimated using the @RISK software for Excel (http://www.palisade-europe.com/) (see e.g. Niessen, 2002, p. 126).

Few people may be interested in specialized courses that address complex data sets or advanced methods of analysis. International cooperation may represent a cost-effective way to build the necessary skills. Some examples exist. In Europe, the European Science Foundation is funding a programme to strengthen the human capacity in Europe to analyse large and complex datasets. This Quantitative Methods in the Social Sciences (QMSS) programme is coordinated by professor Chris Skinner of the Southampton Statistical Sciences Research Institute (S^3RI). (http://www.s3ri.soton.ac.uk/qmss/). The programme tries to improve the quantitative skills by a series of workshops and seminars (2003-2007). Longitudinal data analysis is one of the themes of the QMSS programme. In the UK, the ESRC National Centre for Research Methods is a national programme to promote the methodological skills used by the social science research community. The strategy is a wide range of workshops and short courses. The Centre, which has affiliated centres in many places in the UK, is being coordinated by professor Chris Skinner of S^3RI (http://www.ncrm.ac.uk/index.php).

International cooperation at the PhD level exists too. Examples include the European Doctoral School of Demography (EDSD) (http://www.eds-demography.org/overview/) and the International Max Planck Research School "The Life Course: Evolutionary and Ontogenetic Dynamics (LIFE)" (http://www.imprs-life.mpg.de/).

Lack of easy access to longitudinal data is a major constraint in capacity building. Some data sets that can be accessed relatively easily have contributed considerably to longitudinal research and the development of life course theory[5].

5. *Organize a longitudinal research competition*

Longitudinal research is highly compartmentalized. Longitudinal research in social sciences, econometrics and epidemiology differ considerably in language and method. Although some cross-fertilization exists, it remains limited. A longitudinal research competition may change that and may enhance

[5] Among the longest running observational studies are the Framingham Heart Study (FHS), which started in 1948, and the MRC National Survey of Health and Development (NSHD), better known as the British 1946 birth cohort study. Both studies aim at mapping pathways to health and disease. The Framingham Heart Study resulted in over 1,000 articles in refereed journals and extensive insights in the risk factors of cardiovascular disease (outcome). The concept of cardiovascular life course was introduced and tested using the FHS to complement conventional epidemiological measures with time-based policy measures such as life years lost and years lived with disease (Peeters et al., 2002). The NSHD also triggered the development of a life course perspective of health and disease. Kuh's concern with early life influences on adult health resulted in the definition of the new field of *life course epidemiology* (Kuh and Ben-Shlomo, 1997; Kuh et al., 2003).

interdisciplinary cooperation. It should involve all disciplines and should call on participants to select the best method(s) to answer a given research or policy question using a given data set. It should be a real question and a real data set. The outcome should consist of a research paper ready for submission to a refereed journal. A multidisciplinary panel should assess the results and the best papers should be published in one of the respected journals, e.g. the Journal of the Royal Statistical Society, C (Applied Statistics).

4. Conclusion

As Michael Wolfson observes in his background paper for this conference, the genome project appeals to the public's deep interest in health. A well-defined project on how individual lifepaths evolve as a result of the genetic predispositions, context-specific gene expressions, interpretative schemes that enable us to comprehend the world and our own lives, and environmental influences may generate an even greater public's interest. Computer modelling will be essential to capture the complexities. Individual life courses may be combined and the interactions and transactions that result may generate social constructs such as institutions and networks that both facilitate and inhibit actions by individuals. The outcome is a synthetic, virtual or artificial society that provides a laboratory to study the effects of events, experiences and conditions. The project would not have to start from scratch. Important building blocks exist such as LifePaths of Statistics Canada that creates a virtual nation of about 4 million synthetic Canadians whose characteristics mirror those of real Canadians (http://www.statcan.ca/english/spsd/LifePaths.htm), the artificial society developed by Epstein and Axtell (1996), and the advances in such areas as computational demography (Billari and Prskawetz, 2003), computational economics, computational psychology, and computational biology.

Synthetic biographies generated by combining the best data available on real people represent a powerful instrument for research and policy analysis. The power can be enhanced by adding a developmental perspective briefly introduced in this paper and by using techniques developed in artificial life and complex adaptive systems in general (Langton, 1997). The technology that results would be helpful in training researchers in life course research. In the meantime, other ways exist to build the research capacity for life course research and longitudinal data analysis. Some cost-effective training schemes are presented in this paper.

References

Abbot, A. (1985) Sequence Analysis: New Methods for Old Ideas, *Annual Review of Sociology*, 21: 93-113.

Abbott, A. (2001) Time matters. On theory and method. The University of Chicago Press, Chicago.

Abbott, A. and A. Tsay (2000) Sequence analysis and optimal matching methods in sociology: review and prospect. *Sociological Methods & Research*, 29(1): 3-33

Arthur, W.B. (1994) Increasing returns and path dependence in the economy. The University of Michigan Press, Ann Arbor.

Barker, D.J.P. (1998) Mothers, babies and health in later life. Churchill Livingstone, Edinburgh.

Bartholomew, D.J. (1982) Stochastic models for social processes. Third edition. Wiley, Chichester.

Ben-Shlomo, Y. And D.L. Kuh (2002) A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology*, 31:285-293 (Editorial)

Bernard, P., P. Bélanger et al. (2005) Learning, debating and deciding : the contribution of longitudinal and life course research to public policy in Canada. Available at: http://www.socio.umontreal.ca/personnel/documents/LLSRCFinalReportMay132005_finalversion_.pdf#search=%22Learning%2C%20debating%20and%20deciding%20bernard%22

Billari, F. and R. Piccarreta (2005) Analyzing demographic life courses through sequence analysis. *Mathematical Population Studies*, 12(2): 81-106

Billari, F. and A. Prskawetz (2003) Agent-based computational demography. Springer Verlag, New York.

Blossfeld, H.P. and G. Rohwer (2002) Techniques of event history modeling. New approaches to causal analysis. Lawrence Erlbaum, Mawah, New Jersey. Second Edition.

Bonneux, L. (2000) Cholesterol-lowering therapy for smokers and non-smokers: a life-table analysis. *The Lancet,* 356(9246): 2004-2006

Bourdieu, P. (1986) L'illusion biographique. *Annales de la Recherche en Science Sociales*, 62/63, pp. 69-72.

Brookmeyer, R. and S. Gray (2000) Methods for projecting the incidence and prevalence of chronic diseases in ageing populations: application to Alzheimer's disease. *Statistics in Medicine*, 19:1481-1493

Buck, N. (2002) National strategy for longitudinal studies. Report to ESRC Research Resources Board. UK Longitudinal Studies Centre, University of Essex.

Cairns, R.B. and B.D. Cairns (2002) Plotting developmental pathways: methods, measures, models, and madness. In: E. Phelps, F.F. Furstenberg Jr., and A. Colby eds. Looking at lives. American longitudinal surveys in the twentieth century. Russell Sage Foundation, New York, pp. 267-296.

Commenges D. (1999) Multi-state models in epidemiology. *Lifetime Data Analysis*, 5:315-327

Concato, J., M.P.H., Nirav Shah and R. I. Horwitz (2000) Randomized, controlled trials, observational studies, and the hierarchy of research designs. *The New England Journal of Medicine*, 342(25):1887-1892

D'Andrade, R. (1995) The development of cognitive anthropology. Cambridge University Press, Cambridge.

De Backer, G. and D. de Bacquer (1999) Lifetime-risk prediction: a complicated business. *The Lancet*, 353 (January 9): 82

Elder, G.H. Jr. (1999), The life course and aging; some reflections. Distinguished Scholar Lecture, American Sociological Association, August 1999.

Elder, G.H. Jr. and M. Kirkpatrick Johnson (2000), Perspectives on human development in context. Paper presented at the XXVIIth International Congress of Psychology, Stockholm. Available at http://www.unc.edu/~elder/stockholm.htm

Epstein J.M. Axtell R. (1996); Growing artificial societies, social sciences from the bottom up, Brooking Institution Press, MIT Press Washington D.C., Cambridge Mass.

Erikson, E.H. (1982) The life cycle completed. A review. Norton, New York.

Fontaine, K.R., D.T. Redden, C. Wang, A.O. Westfall and D.B. Allison (2003) Years of life lost due to obesity. *Journal of the American Medical Association*. 289:187-193.

Halfon, N. and M. Hochstein (2002) Life course health development: an integrated freamework for developing health, policy, and research. *The Milbank Quarterly*, 80(3). Available at http://www/milbank.org/quarterly/8003feat.html

Heinz, W.R. and H. Krueger (2001) Life course: innovations and challenges in social research. *Current Sociology*, 49(2). Pp. 29-45.

Hougaard, P. (1999) Multi-state models: a review. *Lifetime Data Analysis*, 5(3):239-264.

Hougaard, P. (2000) Analysis of multivariate survival data. Springer Verlag, New York

Kaplan, H.I. and B.J. Saddock (1981) Modern synopsis of comprehensive textbook of psychiatry. Fourth Edition. Willams and Wilkins, Baltimore.

Kuh, D.L. and Y. Ben-Shlomo (1997) A life course approach to chronic disease epidemiology. Tracing origins of ill-health from early to adult life. Oxford University Press, Oxford.

Kuh. D.L. and R. Hardy (2002) A life course approach to women's health. Oxford University Press, Oxford.

Kuh, D., Y. Ben-Shlomo, J. Lynch, J. Hallqvist and C. Power (2003) Life course epidemiology. *Journal of Epidemiology and Community Health*, 57:778-783.

Lancaster, T. (1990) The econometric analysis of transition data. Cambridge University Press, Cambridge.

Langton, C. (1997) Artificial life: an overview. MIT Press, Cambridge.

Légaré, J. and N. Marcil-Gratton (1990) Individual programming of life events: a challenge for demographers in the twenty-first century. In: C. Maltoni and I.J. Selikoff (eds.). Scientific issues of the next century: convocation of world academies. *Annals of the New York Academy of Sciences*. 610: 99-105.

Linde, C. (1993) Life stories: the creation of coherence. Oxford University Press, New York.

Machalek, R.S. (2004) Evolutionary biology and human nature: the archaeology of epigenetic rules. *Reviews in Anthropology*, 33:193-207.

Macmillan, R. (2001) Violence and the life course: the consequences of victimization for personal and social development. *Annual Review of Sociology*, 27, pp. 1-22. Available at http://soc.annualreview.org/cgi/content/abstract/27/

Malo, M.A. and F. Muñoz–Bullón (2003) Employment status mobility from a life-cycle perspective: A sequence analysis of work-histories in the BHPS. *Demographic Research*, 9(7): 119-161.

Mamun, A.A. (2003) Life history of cardiovascular disease and its risk factors. Rozenberg Publishers, Amsterdam (PhD dissertation, Population Research Centre, University of Groningen, The Netherlands). Also available from http://dissertations.ub.rug.nl/faculties/rw/2003/a.al.mamun/?FullItemRecord=ON

Mamun, A.A., A. Peeters, J. Barendregt, F. Willekens, W. Nusselder, L. Bonneux (2004) Smoking decreases the duration of life lived with and without cardiovascular disease: a life course analysis of the Framingham Heart Study. *European Heart Journal*, 25, 409–415

Miller, P.H. (1983) Theories of developmental psychology. W.H. Freeman and Co., San
Francisco.

Niessen, L. (2002) Road to health. Multi-state modelling of population health and resourse use. Rozenberg Publishers, Amsterdam (PhD dissertation, Population Research Centre, University of Groningen, the Netherlands). Also available from http://dissertations.ub.rug.nl/FILES/faculties/rw/2002/l.w.niessen/c6.pdf

Osler, M. (2006) The life course perspective: a challenge for public health research and intervention. *The European Journal of Public Health*, 16(3):230-.

Peeters, A., A.A. Mamun, F.J. Willekens and L. Bonneux (2002) A cardiovascular life history. A life course analysis of the original Framingham Heart Study cohort. *European Heart Journal*, 23 (2002), pp. 458- 466

Peeters, A., J. Barendregt, F.J. Willekens, J.P. Mackenbach, A.A. Mamun, and L. Bonneux (2003) Obesity in adulthood and its consequences for life expectancy: a lifetable analysis. *Annals of Internal Medicine*, 138:24-32

Pérez-Ocón, R., J.E. Ruiz-Castro and M.L. Gámiz-Pérez (2001) A piecewise Markov process for analysing survival from breast cancer in different risk groups. *Statistics in Medicine*, 20:109-122

Picot, G. and M. Webber (2005) Taking stock. The future of longitudinal surveys. *Horizons* (Statistics Canada), 8(1): 16-23.

Roberts, B.W. and A. Caspi (2003) The cumulative continuity model of personality development: Striking a balance between continuity and change in personality traits across the life course. In R.M. Staudinger and U. Lindenberger (Eds.), Understanding human development: Lifespan psychology in exchange with other disciplines (pp. 183–214). Kluwer, Dordrecht, The Netherlands.

Rogers, A. (1975) Introduction to multiregional mathematical demography. Wiley, New York.

Rude, N., D. Costagliola and A.-J. Valleron (1993) Cumulative incidence of HIV infection and AIDS case prediction in France. *The Lancet,* 342(8868): 436-437.

Sampson. R.J. and J.H.Laub (2005) A life-course view of the development of crime. *Annals of the American Academy of Political and Social Sciences*, 602:12-45.

Spencer, B.G. (2005) A user's perspective. *Horizons* (Statistics Canada), 8(1): 40-41.

Tuma, N. B. and M.T. Hannan (1984) Social dynamics. Models and methods. Academic Press, New York.

Willekens, F.J. (1991) Understanding parallel processes. In: J.J. Siegers, J. de Jong Gierveld and E. van Imhoff eds. Female labour market behaviour and fertility: preferences, restrictions and behaviour. Springer Verlag, Berlin, pp. 11-31.

Willekens, F.J. (2001) Theoretical and technical orientations toward longitudinal research in the social sciences, *Canadian Journal of Population*, 28(2): 189-217

Willekens, F.J. (2003) Multistate demography. In P. Demeny and G. McNicoll (eds.). Encyclopedia of population. Revised edition. MacMillan Reference, New York.

Willekens, F.J. (2005) Multistate models for biographic projections. Report prepared in the context of the MicMac project (Combining micro- and macro-approaches in demographic forecasting), 6th Framework Programme,

Willekens, F.J. (2006) Biographies. Real and synthetic. Manuscript.

Wilson, E.O. (1998) Consilience: The Unity of Knowledge. Alfred A. Knopf, New York.

WHO (World Health Organization) (2002), Life course perspectives on coronary heart disease, stroke and diabetes. The evidence and implications for policy and research. WHO/NMH/NPH/02.1, Department of Noncommunicable Disease Prevention and Health Promotion, Workd Health Organization, Geneva.

Wolf, D. (2000) The role of microsimulation in longitudinal data analysis. http://www.ssc.uwo.ca/sociology/longitudinal/wolf.pdf

Wolfson, M. (2005) Data are not enough. *Horizons* (Statistics Canada), 8(1): 27-30.

Yamaguchi, K. (1991) Event history analysis. Sage Publications, Newbury Park, Ca.