

# **Socio-Economic Statistics And Public Policy: A New Role For Microsimulation Modeling**

by

**Michael C. Wolfson<sup>1</sup>**

**No. 81**

**11F0019MPE No. 81    ISSN: 1200-5223**

**ISBN: 0-662-21734-9**

24th floor, R.H. Coats Building, Ottawa, K1A 0T6

Institutions and Social Statistics

(613) 951-8216 wolfson@statcan.ca

Facsimile Number: (613) 951-5403

July 1995

A Paper Prepared for the  
50th Session of the International Statistical Institute, Beijing, 21-29 August 1995

Invited Paper Meeting on “Changes in User Expectations About the Quality of Survey  
Statistics in the Public and Private Sector

This paper represents the views of the author and does not necessarily reflect the opinions of  
Statistics Canada.

*Aussi disponible en français*

---

<sup>1</sup> Statistics Canada and the Canadian Institute for Advanced Research. The work reported here is very much a team effort principally by members of the Socio-Economic Modeling Group of the Analytical Studies Branch, Statistics Canada. Geoff Rowe and John Armstrong led the empirical analysis, while Steve Gribble created the ModGen software environment within which the LifePaths microsimulation model is written, and provided continuing leadership to the team. I remain responsible for any errors or infelicities in this paper.

## **Abstract**

Users of socio-economic statistics typically want more and better information. Often, these needs can be met simply by more extensive data collections, subject to usual concerns over financial costs and survey respondent burdens. Users, particularly for public policy purposes, have also expressed a continuing, and as yet unfilled, demand for an integrated and coherent system of socio-economic statistics. In this case, additional data will not be sufficient; the more important constraint is the absence of an agreed conceptual approach.

In this paper, we briefly review the state of frameworks for social and economic statistics, including the kinds of socio-economic indicators users may want. These indicators are motivated first in general terms from basic principles and intuitive concepts, leaving aside for the moment the practicalities of their construction. We then show how a coherent structure of such indicators might be assembled.

A key implication is that this structure requires a coordinated network of surveys and data collection processes, and higher data quality standards. This in turn implies a breaking down of the “stovepipe” systems that typify much of the survey work in national statistical agencies (i.e. parallel but generally unrelated data “production lines”). Moreover, the data flowing from the network of surveys must be integrated. Since the data of interest are dynamic, the proposed method goes beyond statistical matching to microsimulation modeling. Finally, these ideas are illustrated with preliminary results from the LifePaths model currently under development in Statistics Canada.

**Keywords:** microsimulation, social statistics, statistical frameworks

## 1. Introduction

It is an eminently reasonable expectation that a nation's statistical system provide reliable, coherent, and salient views of central socio-economic processes (e.g. Garonna, 1994; OECD, 1976). To an important extent, this is accomplished by the System of National Accounts (SNA). However, it has also long been appreciated that the SNA suffers from many serious limitations, particularly from the viewpoint of social concerns and policy. These limitations are implicit in the history of attempts to create sets or systems of internationally agreed social indicators.

To date, nothing has emerged from efforts to construct social indicators that compares to the SNA in breadth, coherence, and international acceptance. As a result, fundamentally new approaches appear necessary to meet these needs for socio-economic statistics, including the needs that have long motivated efforts in the area of social indicators.

Broadly speaking, three main strategies have been proposed for developing a statistical framework for the social sphere. One is extensions of the SNA, most prominently in either the form of Social Accounting Matrices (SAMs, e.g. Pyatt, 1990), or Satellite Accounts (e.g. Vanoli, 1994; Pommier, 1981). The second proposed strategy is construction of a framework designed specifically for social statistics -- the best known and most clearly articulated being Stone's System of Social and Demographic Statistics (SSDS; UN, 1975; Stone, 1973). The third approach foregoes the structure and coherence of an explicit framework, seeking only consensus on an ad hoc collection of statistical indicators. This is exemplified by the set of social indicators recommended by the OECD (Moser, 1973; OECD, 1976).

All three strategies have failed to achieve broad implementation within advanced countries, and have as a result failed to provide a basis for internationally comparable data. Three reasons can be identified for these failures. One is concern about feasibility, a second is lack of priority from national governments or statistical agencies; and a third is lack of salience. (Obviously, these reasons may be related.) The Social Indicators experience (OECD, 1982) did result in a completely specified and operational list of indicators -- based on agreement among both experts and senior government representatives of member countries. However, for many of the agreed indicators (e.g. healthfulness of life, use of time, income), the requisite data collection systems have still not been created in an internationally comparable manner. Since data systems for these purposes exist within some countries, it is clearly not a question of technical feasibility.

This leaves some mix of lack of salience and lack of sufficient priority -- including unwillingness to invest the necessary resources in statistical data collection -- as the main explanations. To the extent there is a lack of salience for social indicators, it is more likely in the area of international comparability than in domestic usefulness. Most countries do have a wide range of social statistics; the problem is that they are generally not comparable from one country to another. This revealed lack of interest in a comprehensive set of internationally comparable social statistics may simply be a result of the fact that economic concerns (as reflected in the successful efforts to create an internationally agreed SNA) are broadly seen as much more important than social concerns, at least from a comparative point of view. Another might be that a handful of basic social indicators (e.g. life expectancy, unemployment rates) are already available on an internationally comparable basis, and these are felt sufficient.

One source of relevance of internationally comparable data is that countries feel strong connections with one another in the given domain. In the economic sphere, it is obvious that countries are tied together substantively by trade and international financial flows, and by shared intellectual roots in macro-economic theory. These connections form a basis for the SNA. Corresponding substantive connections in the social sphere may simply be seen to be weaker (though cultural and intellectual flows via mass media and coincident international patterns of terrorism, unemployment, marriage breakdown, fertility decline, growing wage inequality, etc. make this view tenuous). In addition, if there is no shared theory, the basis for international comparability is weaker -- both at the level of individual statistical series (e.g. household income distribution), and how various social statistics series are put together (if at all). In short, countries' failures to invest in creating the data collection systems necessary for the OECD's Social Indicators may be the result of a lack of interest in comparability among an ad hoc collection of indicators.

However, the failures of implementation of the SSDS or Satellite Accounts in the social area must be due to more than a lack of interest in internationally comparable data; they must also derive from a lack of interest in their implicit underlying theoretical frameworks. The failures here are more profound because even within countries, it is difficult to find coherent and comprehensive structures for social statistics that come anywhere close to the SNA in scope and implementation.

The decades of failure of all of these strategies suggests that new approaches are required in order to provide a coherent structure for social statistics. Such coherence would help meet the needs of a growing segment of users. On one side, it would provide the basis for generalist needs, for example summary indications of broad trends. On the other side, bringing coherence to the complex of social statistics should help more specialized users who can currently become confused (justifiably) when they find several inconsistent estimates for the same item from different data sources.

## 2. Starting from First Principles

As a first step in thinking about new approaches to the construction of a framework for social statistics, it is useful to set out a series of basic measurement objectives. There are three which hopefully command broad support.

**a. overall outcome** -- A primary objective is to tell whether or not "things are getting better all the time" (in the words of the Beatles). Are people better off than they were last year, than a decade ago? Answering this question is difficult, primarily because there is no widely agreed summary approach to measuring how well off individuals are. Money income, health status, educational attainment, and social deprivation are all ingredients for such a measurement. But there is no agreement as to what other ingredients are essential, nor how the ingredients should be combined into an overall index. Moreover, there is lack of agreement on the appropriate outcome concept within various domains such as health and education.

The character of this partial consensus has major implications for statistical framework development. One is that flexibility is required. To the extent that there is agreement on some of the ingredients in measuring overall well-being, these ingredients must be included in the underlying statistical program. But given that there is no single "right answer" as to how they should be combined, users should have flexibility in combining them. It should be possible to assemble alternative summary indices from the basic ingredients -- both at the level of domains like health and education, and overall.

A second implication is that social statistics should be neither subservient to nor separate from economic statistics, as has characterized the three main strategic approaches so far. Economic status is clearly a central ingredient in any broad measure of whether or not people are becoming better off. Thus, it is better to think of a statistical framework for *socio-economic* statistics than for social statistics alone. In turn, and in contrast to the views of some national accountants (e.g. Vanoli, 1994), the prospects for frameworks for social statistics that build out from SNA concepts are not good. Completely fresh approaches are required. They should build on new premises, taking the whole of the "social economy" as the domain. The SNA would then become an important component of a broader "System of Social and Economic Statistics" (Wolfson, 1994; Ruggles and Ruggles, 1973).

Yet a third implication is that summary methods for comparing two or more social economies -- either over time or between countries -- must be found that go beyond simple linear aggregation. The concepts of interest are not always amenable to aggregation based on a single numeraire as is the case with money values in the SNA. Fortunately, work in areas like methods for comparing household income distributions, or distributions more generally using graphical methods (e.g. Easton and McCulloch, 1990), and flexible data base architectures allowing set-theoretic and pointer-based arithmetic over complex multivariate, longitudinal micro data sets -- which are now readily feasible with modern informatics -- indicate that SNA-style aggregation is not essential. Indeed, such approaches complement and support the second following objective.

**b. variety** -- Another basic measurement objective is to enable users to see the variety in the social economy. Variety covers all kinds of heterogeneity -- for example going beyond aggregates and

averages to meet the long-standing criticism of the SNA that it conveys nothing of the rich, the poor, and the degree of inequality in the size distribution of income. Variety is also reflected in the dispersion of educational attainments and household structures in a country's population.

Capturing variety has fundamental implications for a statistical system. Essentially, it requires explicit micro data foundations. The SNA, as the pre-eminent statistical framework, predates the revolution in computing. This has evidently constrained creative thinking for social and socio-economic statistical frameworks. In the pre-computer era in which the basic structure of the SNA was developed, aggregation was not only part of the theoretical foundation, it was also a practical necessity. Today with modern database technology, aggregation is not only inimical to accurate reflection of variety, it is also practically unnecessary.

(The idea of explicit micro data foundations in such statistical endeavors is certainly not new, e.g. United Nations, 1979; Ruggles, 1981. On the other hand, the pervasiveness of the "aggregation culture" is evident in the OECD Social Indicators efforts wherein it was felt necessary to define a set of *a priori* "basic disaggregations of main social indicators"; OECD, 1977. Such agreement, while helpful, is certainly not essential when appropriate and internationally comparable micro data sets are readily available to analysts.)

**c. what if** -- The third fundamental measurement objective is to provide the basis for the careful posing of and answering "what if" questions. There are two basic motivations. The obvious one is that government policy departments and private sector decision-makers, as major users of socio-economic statistics, want answers to these kinds of questions. For example, what would the distribution of disposable incomes be if such and such a tax/transfer policy change were implemented; or what would spending on such and such commodities be in five years if current trends prevailed?

Less obvious but equally important is that key statistical indicators are in fact answers to such "what if" questions. The best example is life expectancy. Life expectancy in 1990, say, is the answer to the following hypothetical question, "how long could a birth cohort expect to live if all its members were always exposed to the mortality rates that were observed in 1990?" Thus, life expectancy is not a datum that is directly observed like counts of deaths by age and sex. Rather, it is the outcome of a numerical simulation, one that is tightly coupled to disaggregated numerator and denominator count data on deaths and populations at risk respectively.

While life expectancy is a hypothetical construct that is not directly measured, it is also an intuitive and broadly accessible concept that can provide the framework for related families of indicators. This is clearest in the health area, where an ad hoc group of researchers has come together in the REVES (reseau pour esperance de vie en sante; Mathers and Robine, 1993) to develop and seek consensus on just such a related family of health indicators.

Drawing this discussion together, three basic measurement objectives have been proposed:

- broad indicators of the extent to which individuals are becoming better off;
- capacity to display variety and heterogeneity in the population; and
- tools for posing and answering "what if" questions.

In turn, in order to meet these objectives, the underlying statistical framework must:

- be flexible;
- encompass both social and economic aspects;
- have explicit micro data foundations;
- utilize modern informatics and database technology; and
- incorporate simulation models that are tightly coupled to data.

Given these premises, what might serve as the key elements of a framework for socio-economic statistics? The following points are apt:

- at any point in time, the population is best represented by a sample of individuals, each of whom is characterized by a set of attributes and relationships;
- attributes include income, educational attainment, consumption, various aspects of health status, and time use patterns of activity;
- relationships include conventional kinship ties as well as cohabitation (i.e. in database or graph-theoretic terms, such relationships can be represented by various kinds of pointers to other individuals -- each of whom is also in the database);
- relationships also include interactions with the major institutions of society -- school, work and government programs. These contacts, relationships or transactions between individuals and major institutions can also be considered part of the set of individual attributes. They can take the form of pointers to descriptions of the institutions -- schools, workplaces, and government programs -- with which the individuals were interacting;
- this individual database can then easily be viewed as comprised of a hierarchy of various types of units -- e.g. individuals, nuclear families, extended families, and households;
- each unit (individual, family or household) can be described by any one of a number of summary attributes such as disposable income, leisure time, or self-reported satisfaction;
- measures of variety can then be defined by summary statistics over this multivariate joint distribution of units (e.g. Gini coefficients, quantile shares);
- over time, the population is best represented by a series of individual biographies, the equivalent of a broad and deep longitudinal panel survey;
- given this longitudinal representation, a coherent family of summary indicators can be constructed from generalizations of the notion of life expectancy -- including partitions of life expectancy into cumulative sojourn times in various life states;

In essence, this socio-economic framework would contain a complete longitudinal micro data sample, a microcosm of the actual population and its relationships to major social and economic institutions. From this microcosm, a wide variety of statistical indicators could be readily constructed -- effectively with no more effort than pressing the ubiquitous <Enter> on a computer keyboard to launch the appropriate software algorithm and have it pass through the microcosm data.

By construction, all such summary indicators would be coherent because they would be derived from the identical underlying micro data base. The summary indicators would not obscure the population's variety and heterogeneity, because the underlying micro data base would always be open (at the click of a mouse button, say, in terms of contemporary informatics functionality) for detailed inspection.

The main question is from where would this microcosm come? For the very practical reasons of cost, respondent burden, and concerns for individuals' privacy, it could not come from an omnibus longitudinal household survey. Moreover, there is not time to wait half a century or more for such a longitudinal survey to be substantially completed, by which time many things will have likely changed dramatically. The unavoidable conclusion is that the microcosm will have to be synthesized.

Such synthesis would be an extension of the synthesis of a population cohort already implicit in indicators such as life expectancy. It would differ methodologically, because the semi-aggregate or cell-based approach inherent in the underlying life table is incompatible with explicit micro data foundations. Instead, microsimulation is required.

In effect, what is being proposed is a weaving together of the ideas in Stone's SSDS (UN, 1975), with the idea of explicit integrated micro data bases proposed by a subsequent international expert group (UN, 1979). The first step is the recognition that the SSDS implicitly rests on longitudinal micro data. Indeed, Stone (1973) notes that,

"Of course, if statistics are collected by means of a linked system of compatible records or, better still, by a continuously updated, comprehensive system of individual data (*i.e. longitudinal micro data*), a discussion of sequence (*i.e. representations in terms of discrete time, first order Markov chains*) becomes largely irrelevant since the information in a vast, computerized data bank can be combined in any desired manner. But while these may be the methods of statistical collection in the future, they are not, with very limited exceptions, in operation at present, and so it makes sense to discuss the systematization of social statistics in terms of more familiar methods of collection." (p152, italics added)

In this sense, the future has arrived, so Stone's matrix algebra, restrictive first order Markov assumptions, and "familiar methods of data collection" need no longer be constraining.

The second step is extension of the ideas of creating integrated data bases (IDBs) synthetically, by means of statistical matching methods, so clearly articulated almost two decades ago by the UN's IDB working group (UN, 1979). They recognized the great utility of more highly multivariate micro data, as well as the practical limitations of collecting such data directly. As a result, they recommended that the desired micro data be synthesized, even if it meant that the underlying micro data records were artificial. The IDBs in that earlier work referred generally to *cross-sectional* micro data.

The bridge between these two broad ideas -- a Stone/SSDS-style framework based on longitudinal dynamics, and synthetic statistical matching of micro data -- is synthetic *longitudinal* micro data. The difference is that creation of synthetic longitudinal micro data requires more than techniques of statistical matching, since this matching idea does not carry over well to combining disjoint longitudinal micro data sets. The synthetic longitudinal micro data must instead be generated by dynamic microsimulation modeling (again not a new idea; see Ruggles, 1981). In essence, what is being "matched" across longitudinal micro data sets by microsimulation is not the character of individual observations, but rather observed patterns of dynamic behaviour for groups of observations in each micro data set (e.g. as sketched in a later section).

Furthermore, the synthesis of the microcosm using microsimulation means that the marginal cost of developing a "what if" capacity is negligible. For example, once a life table has been constructed, relatively little extra work is required to compute cause-deleted life expectancy. The analogous situation applies to a microsimulation basis for constructing the population microcosm. Once the investment has been made in the capacity to synthesize a "baseline" microcosm, synthesis of "variant" microcosms is relatively straightforward.

Finally, as will become evident in the description to follow, this lifecycle microanalytic approach means that one need no longer be faced with a choice between time-based and demographic styles of social accounting as discussed in Juster and Land (1981). The approach being developed here encompasses both.

### **3. Implications for Data Collection Systems**

Consideration of a socio-economic statistical framework along the lines sketched above has major implications for both conceptual and operational aspects of data collection systems in a national statistical agency. These implications may not be that costly (relative to primary data collection costs), and most are relatively straightforward:

- data collection processes cannot exist as "stovepipe" systems, in isolation from one another;
- one kind of coordination across data collection processes is use of common concepts and definitions (e.g. identical definitions and methods for eliciting educational attainment);

- the other kind of coordination is assuring appropriate overlap in content -- basically to anticipate the need for synthetic statistical matching (or equivalent methodologies); and
- microanalytic uses of raw data are far more demanding of data quality than aggregative uses.

In effect, this means that data collection systems must be jointly planned, and that micro level data quality standards must be more stringent.

The joint planning requirement is not new. Construction of the SNA also requires some coordination of data feeder systems, not least to assure that there is some method to cover all sectors of the economy. However, this coordination is much less onerous than that entailed by microsimulation. The reason is that inconsistencies across data collection systems uncovered by the SNA can be resolved at the high level of "macro editing". Adjustments are made to broad aggregates, notwithstanding the fact that this introduces inconsistencies between various SNA aggregates and their source micro data. However, for the microsimulation purposes here, a key objective is internal consistency across source data sets at the micro level.

The micro level data quality requirement is also not new. It has been faced most acutely whenever a public use micro data set is produced. Knowing that users will subject the data to intensive inspection and analysis, for example as part of assessing regression "outliers", extensive editing and imputation is applied to these data. Similar but weaker micro level data quality concerns are faced with population census micro data files that, while not publicly available, are open to generalised ad hoc cross-tabulation requests.

Still, micro level data quality concerns will be much more acute in the context of an integrative microanalytic framework such as that about to be described. It is one thing for a process of edit and imputation to assure that each record in a given micro data set is plausible and internally consistent. It is quite another to assure that multiple micro data sets are mutually consistent -- for example that a health survey and a disability survey yield the same age- and sex-specific distributions of disability by severity, or that a longitudinal survey on labour dynamics produces cross-sectional estimates of labour force participation that agree with those generated by the mainline labour force survey, or that a time series of administrative data on school enrollments is consistent with census data on educational attainment by age and sex.

This requirement for mutual consistency highlights a concern raised by Wilk (1987), namely the relative weakness of statistical methods for addressing non-sampling error. For example, item non-response or bias in household surveys typically causes serious under-reporting of selected income sources. However, conventional edit and imputation processes usually address this in only a limited manner (e.g. income components falsely reported as zero are not changed). It is only the community of microsimulation modelers for tax/transfer policy who have, of necessity, had to grapple with this problem (Citro and Hanushek, 1991; Bordt et. al., 1990; Wolfson et al., 1989). Moreover, household survey editing does virtually nothing about response rounding (e.g. giving income to the nearest \$100 or \$1000) -- even though there is evidence that such respondent behaviour can cause errors in some statistics (e.g. quantiles) of the same magnitude as conventional sampling error (Rowe and Gribble, 1994).

Finally, there is a growing recognition of the importance of longitudinal surveys, which are clearly fundamental to developing descriptions of dynamics, and to disentangling causal pathways. Using longitudinal micro data for these purposes will entail the use of more sophisticated inferential methods than statistical agencies typically encounter, for example hazard regression as compared to cross tabulation. This in turn should expose the data to far more critical scrutiny.

#### **4. The LifePaths Project**

We turn now to an illustration of these general points. The LifePaths project is an effort to construct a prototype socio-economic statistical framework. The project is being undertaken by Statistics Canada on behalf of the Canadian Ministry of Human Resources Development, the recently created "super-ministry" responsible for welfare, pensions, unemployment insurance, and labour market policies, among others.



The basic objective of the LifePaths statistical framework is to provide a coherent and multi-faceted series of "views" of the socio-economic status of the Canadian population. This framework is designed to have the general characteristics indicated earlier, namely a capacity to indicate overall outcomes and variety, and to provide answers to "what if" questions. The substantive domain includes how Canadians are spending their time in various activities such as working, learning, family roles, participating in government programs, and leisure.

Generalizations of working life tables form one of the central views or facets to be provided. Table 1, for example, shows for Canadian male birth cohorts, not only conventional life expectancies, but also the average ages at which men could expect their first entry and last exit from the paid labour force. By examining a series of these (period) birth cohorts, each representing a successive decade, the analysis vividly displays the long run trends of more time spent in schooling, ever earlier ages of retirement, and a general reduction in working years for men. The final column also gives a clear indication of the impacts of these trends on public pension costs. (Note that while old, these are apparently the most recent working life table estimates available.)

**Table 1 -- Historical Stationary Male Life and Working Life Expectancies at Age 15**

Year	average age at			number of		working years per year of retirement
	entry to labour force	retirement	death	working years	retirement years	
1921	16.5	63.7	67.6	47.2	3.9	12.1
1931	17.0	64.0	68.4	47.0	4.4	10.7
1941	17.2	64.1	69.1	46.9	5.0	9.4
1951	17.5	63.9	70.4	46.4	6.5	7.1
1961	18.2	64.0	71.2	45.8	7.2	6.4
1971	19.8	63.3	71.3	43.5	8.0	5.4

Source: Gnanasekaran and Montigny (1975) and Wolfson (1979)

The LifePaths framework extends these basic working life table results in several directions. Annual work patterns are considered in more detail, going beyond a two-way breakdown between working and non-working years. For example, part-time work, increased duration of paid holidays and vacations, changes in typical hours worked per week, sub-annual spells of unemployment or withdrawal from the labour force, periods where work and schooling are simultaneously pursued, and more participation in self-employment are all taken into account. In addition, the time aspects of work are combined with the economic aspects, particularly income.

Other major forms of activity are also included. One is formal schooling; another is familial context (e.g. living alone or with other family members). Thus, involvements with the major institutions of society -- work, school, and family -- are covered. The LifePaths framework therefore combines both the "active" (learning and earning) and "passive" sequences ("succession of family groupings to which individuals are attached in the course of their life", p145) in Stone's (1973) demographic accounting SSDS proposal. This is a capacity that in practice becomes combinatorially intractable with the matrix methods he used.

Additionally, participation in major social programs is planned -- for example, Social Assistance (SA), Unemployment Insurance (UI), and Workers Compensation (WC) disability pensions. Generally, a more fine-grained account of time use is included, based on data from time use surveys -- the time-based accounting proposed, for example, by Juster et. al. (1981). Thus, major categories of activity include not only work and school, but also unpaid housework, personal care, care for others, sleep, commuting, TV, other passive leisure, active leisure, interaction with family members, and other social interactions.

The LifePaths framework encompasses all these human activities, from a complete life cycle perspective, in a coherent and integrated manner -- thereby combining and nesting both time-based and demographic accounting approaches as debated in Juster and Land (1981). Constructing the LifePaths framework is challenging, and the results to be presented here are a prototype.

Methodologically, the LifePaths framework is premised on several major statistical innovations.

First, no single data set contains all the required information, for example detailed data on human activities from both economic and social perspectives. Current and planned data sets in this domain are partial and fragmentary. Moreover, as already noted, considerations of cost, respondent burden, and privacy suggest that fully integrated household survey data will never be practical. Thus, processes of synthetic integration, utilizing multiple data sets, are inevitably required.

Second, the framework is intended to cover individuals' full life cycle histories. Doing so with actual longitudinal data would require decades of survey follow-up, by which time many things will have changed. Thus, the basic idea is to build on and generalize the concept of *period* life expectancy and its underlying life table. In turn, this means that the analysis will focus on realistic, but hypothetical, population cohorts.

A third part of the basic objective is the detailed reflection of individual variety or heterogeneity, and in turn, a capacity to view distributional phenomena such as income inequality. This capacity requires explicit micro data foundations. Since data on the actual life paths of a representative sample of individuals is infeasible, the underlying micro data must be synthetic. Yet at the same time, these data must be sufficiently realistic to be essentially indistinguishable from the partial sets of characteristics observed in real data from actual population samples, including longitudinal surveys.

These requirements imply that the heart of the LifePaths statistical framework must be a microsimulation model. In other words, the core of the statistical framework is a sample of realistic, but synthetic, individual life paths.

## 5. On Synthetic Data

Before presenting initial results, it is important to explain the sense in which the LifePaths framework is based on synthetic data, and the extent to which these synthetic results are a reasonable reflection of current realities.

Human lifetimes typically span about three-quarters of a century. However, given the relatively rapid pace of change over a wide range of human activities, it is almost impossible to have consistent and stable socio-economic observations for this length of time. Statistics that have been well accepted for decades simply did not exist 75 years ago -- for example the unemployment rate, GDP per capita, and measures of leisure time. Correspondingly, it is quite possible that 75 years from now, in 2070, these basic statistics, whose importance is taken for granted today, will have been superseded by new kinds of statistics we can barely imagine.

Yet there is a very broad interest in statistical indicators that do reflect processes spanning a human lifetime. The most obvious is life expectancy. Other such statistics are the proportions of marriages that can be expected to end in divorce, the number of different jobs an individual can expect to have over his or her working career, the expected adequacy of public pensions relative to pre-retirement earnings, and the portion of life expectancy that will be spent in good or ill health. Clearly, such lifetime statistical indicators exist and are more or less widely accepted. The LifePaths framework generalizes such indicators.

While it may not be widely appreciated, life expectancy is a "made up" statistic. It is analogous to a statement about where a car is heading based on its position and velocity while ignoring any acceleration. Life expectancy is based on current age- (and sex-) specific death rates so, like vehicle speed, it is based on real data. But (period) life expectancy applies to a hypothetical individual who has been taken out of calendar time, and spends his or her entire lifetime exposed to the mortality rates of the early 1990s. In essence, any acceleration or deceleration of mortality rates is ignored.

It is, of course, well known that mortality rates have generally fallen over past decades, and it is widely anticipated that these declines will continue. Thus, while life expectancy itself ignores these trends in mortality rates, trends in life expectancy provide very convenient summary indicators of these changes in underlying mortality rates, since they track a form of weighted average of age- (and sex-) specific mortality rates, all of which are changing over time. The underlying age-specific mortality rates are always available for inspection, but trying to make sense of the evolution of even one hundred numbers is a complex task. (More numbers are involved if mortality is broken down by sex and marital status as well as by single year of age.) Life expectancy is a helpful indicator precisely because it collapses these hundred numbers into a single intuitively accessible indicator -- one whose changes over time correspond reasonably to the changes over time in the underlying age-specific mortality rates.

The LifePaths framework is designed to be completely analogous. However, it builds on a much richer variety of processes and statistical descriptions of individuals' transitions among various life states. For example, in addition to mortality, explicit account is taken of demographic states such as marital status, and the associated transitions of entering a common law or legal union, and leaving a union to a separated or divorced state. Similarly, other socio-economic status classifications like working or engaging in learning have been included, based on real data on recent distributions and transition rates amongst these states.

In order to achieve this kind of generalization of life expectancy, the underlying concept of a life table has had to be generalized as well. In a life table, the finest level of detail is a group of individuals -- for example as defined by sex and single year of age. Within such a group, all individuals are assumed to be homogeneous. In LifePaths, this level of detail is insufficient. Explicit consideration of heterogeneous individuals characterized by multiple attributes is essential in order to make the best use of, and to reflect most accurately, results emerging from analyses of dynamic behaviour patterns in rich longitudinal micro data sets.

In an important sense, this implies that LifePaths produces results that are much more realistic than life expectancy produced from a conventional life table. For example, in LifePaths, mortality rates are broken down by marital status as well as age and sex; and in turn marital status depends in a complex way on factors like educational attainment, fertility history, and labour force activity durations.

On the other hand, the synthetic character of the "data" underlying LifePath results is inevitably more explicit than in the case of a life table. While a population of individuals underlies any conventional life table, the individuals themselves are only implicit -- all that is calculated is the numbers of individuals in each cell or category, e.g. by age and sex. In contrast, in the LifePaths framework, all individual life paths must be explicit.

So what meaning should be attached to a LifePath result such as a breakdown of life expectancy into the number of years an individual can expect to spend in the paid labour force and in learning? The interpretation should be analogous to conventional life expectancy -- a kind of summary of recent population flow rates. LifePaths results show how things would be if recent rates of transitions among socio-economic states (conditional on attributes for heterogeneous individuals) were constant.

## **6. Initial Results**

The LifePaths statistical framework consists, fundamentally, of a sample of complete (synthetic) individual life cycle histories. However, this longitudinal micro data base of sampled life histories is far too complex to be examined directly, so we offer here only a few summary "views" of the underlying microcosm. The specific views start from conventional demographic analysis.

Note that these "views" stop short of summary scalar indicators like GDP; they show simultaneously a number of basic population attributes. This need not be seen as a weakness, as the inability to make the last step to a single overall measure as in the SNA. Rather, a given "view" can be seen as a demonstration of the power of contemporary computer graphics to facilitate more textured appreciations of social economies than is possible with a single index.

To start, one of the most basic demographic images is the population pyramid. Figure 1 shows such a pyramid for the base case life table population, with counts of females along the horizontal axis to the right, male counts to the left, and age to 100 along the common vertical axis. It is based on *period* (late 1980s and early 1990s) transition probability functions, which are sketched in a later section. As expected, at higher ages, the survival curve for females falls more slowly than that for males, a counterpart to (or more accurately the underlying reason for) females' higher life expectancy. (The blip in the age 99 interval reflects the fact that this is actually the age  $\geq 99$  interval.)

**Figure 1 -- LifePaths Population (person-years) by Major Activity, Age and Sex**

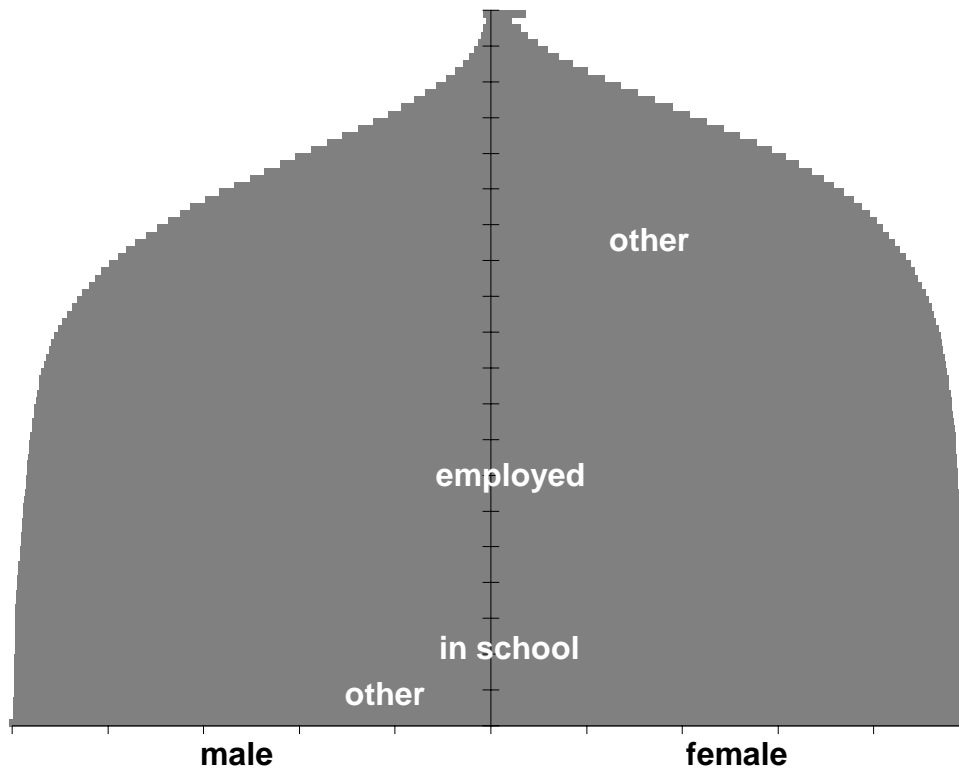
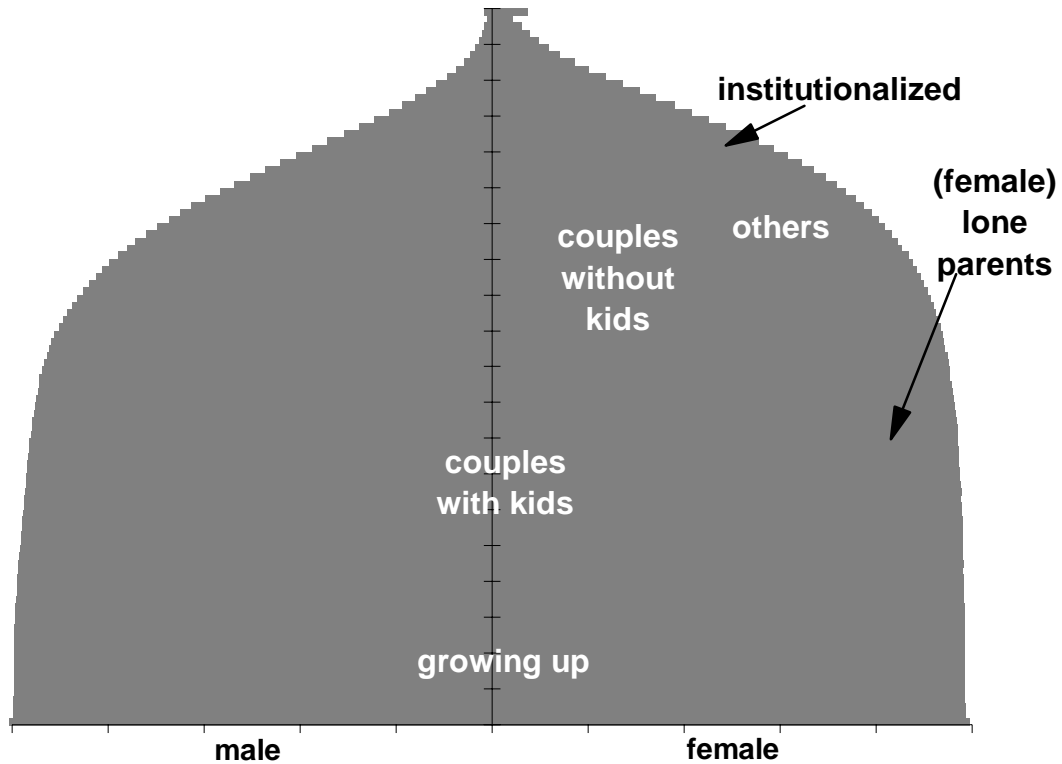


Figure 1 also shows the population broken down into three socio-economic categories -- "employed", in "school", and "other". "School" starts at grade 1, so daycare and kindergarten are part of "other". Since the LifePaths framework tracks individuals through time continuously, some arbitrary decisions have been applied in years where individuals engage in more than one activity. Specifically, to be considered "employed" in this diagram, the individual had to be working at least 15 hours per week, and the plurality of time during the year had to be spent working at this rate. Thus, someone who spent 5 months as a student, 4 months working at least 15 hours per week, and the remaining 3 months of the year working less than 15 hours per week (including not working at all) would be considered in "school" that year; while if the 5 and 4 were reversed, they would be considered "employed". (Definitions such as these are under the control of the LifePaths user.) The diagram shows that virtually everyone is in school by age 8, a few start leaving at age 16, most have left by age 20, but there is a tail of both males and females who are in school through their twenties.

No one appears to make a transition directly from school to employment, though we return to this point in a later figure. Instead, perhaps a surprising proportion of individuals are in the "other" category, which includes the unemployed as well as those not in the labour force (e.g. homemakers, the retired). As expected, males are more likely to be employed at various ages than are females. The employed portion of the population shows a dip in the age-related trend to higher participation for women in the prime child-bearing years 20-25, and then something of an acceleration in the 25-35 age range. Men show a relatively sharp decline in participation in the age 60-65 age range.

Figure 1 corresponds to Stone's "active sequence" (i.e. transitions among working and learning states), while Figure 2 gives an overview of his "passive sequence". It uses the same population pyramid graphic form, and refers to exactly the same underlying LifePaths synthetic population, but classifies individuals along a different dimension, family status. By definition, all individuals under age 18 are classified as "growing up" unless they are married or have a child. Also, whenever a marriage breaks down, any children are assumed to remain with the mother. This assumption explains why there are female, but no male lone parents. (Future versions will incorporate more realistic data on custody arrangements.)

**Figure 2 -- LifePaths Population (person-years) by Family Status, Age, and Sex**

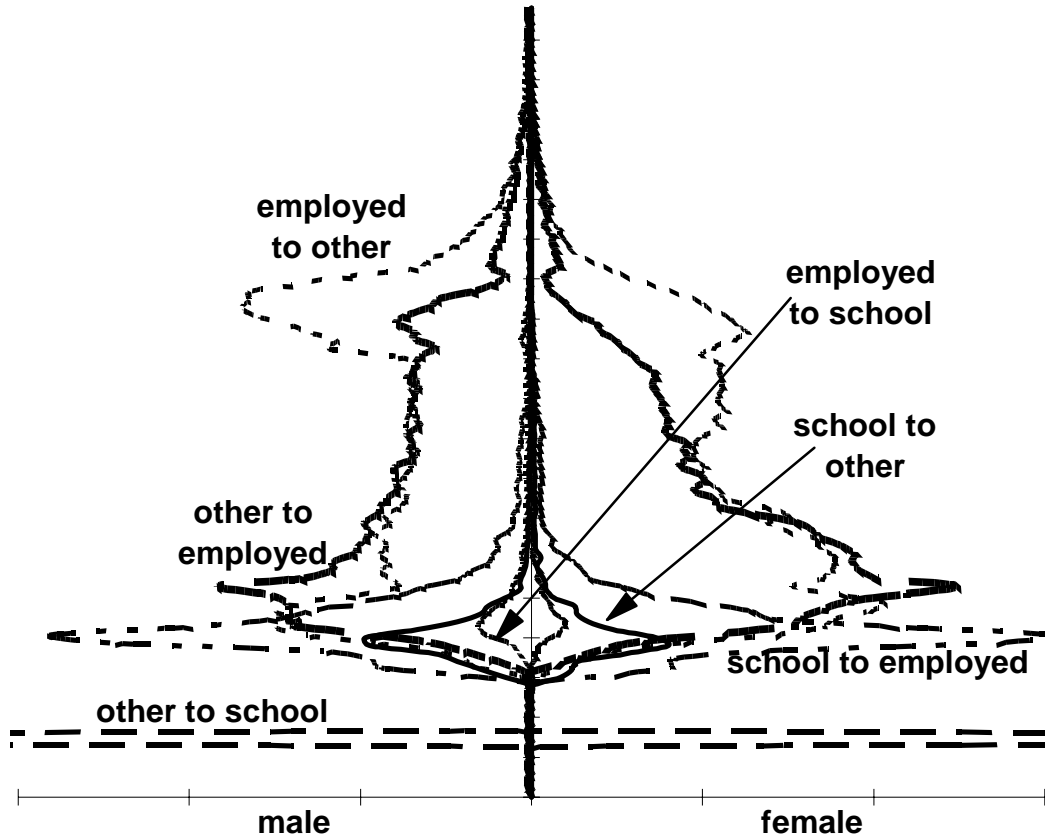


Comparing the male and female curves for the married states (couples with and without children) shows the male curves displaced a few years toward higher ages. This is a reflection of the general pattern where husbands tend to be a few years older than their wives. The diagram also shows there are many more widows than widowers. This is a consequence of both the positive average age difference between husbands and wives, and the greater life expectancy of women. Finally, the diagram indicates the much higher rates of institutionalization of women (principally in nursing or chronic care facilities), due in turn to their greater longevity and higher prevalence of health problems at older ages, and the fact that similarly incapacitated males more often have a wife who can care for them at home.

Figures 1 and 2 show only the beginnings of the LifePaths framework; they are simply two "views" (in this case cross-tabulations) of the full underlying microcosm -- a longitudinal micro data set for a synthetic "early 1990s" period birth cohort. Exactly this same underlying longitudinal micro data set can be tabulated to generate the view in Figure 3, which shows *flows* between states rather than *stocks* of individuals within each state. In this case, Figure 3 graphs the flows corresponding to the stocks in Figure 1. The horizontal axis shows the number of individuals making each kind of transition each year, again in population pyramid style with age along the common vertical axis, females on the right horizontal axis, and males on the left. (The extremes of the horizontal axis span 18% of the population, so that for a cohort of 100,000 the maximal male and female flows shown are each 9,000 per year.)

The first transition is from “other” (early childhood or pre-school) to “school”. Figure 1 indicates that all male and female children make this transition by ages 6 and 7. The next major transition is at the end of “school”, where the peak flow rate to “employed” occurs around age 20 for both males and females. A smaller number, also peaking at about age 20, move from school to “other” activity. Recall that the “other” category is any person-year where the plurality of the year (i.e. at least a tiny bit more than one-third) was spent neither as a student nor working more than 15 hours per week.

**Figure 3 -- LifePaths Gross Flows Between Major Activities  
(persons per year) by Age and Sex**

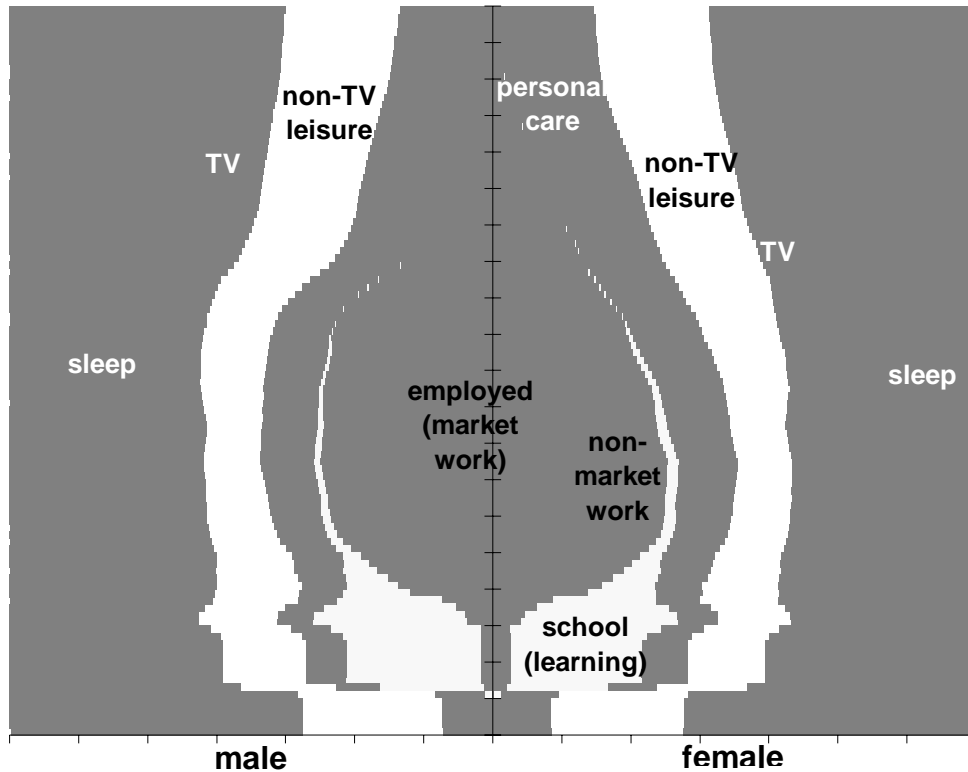


From early adult ages to the 60s, the main flows are between the “employed” and “other” categories. Note that all these flows are gross rather than net. It is notable that the net flow between employed and other (based on comparing the gross flows) shifts direction toward “other” in the 40-45 age range for females, but remains quite small for males through age 50. This is followed by retirement peaks in the 55-65 age range, the one for males being more pronounced.

In addition to stocks and flows of individuals in various categories of activity, the LifePaths framework also supports data views showing sojourn times -- lengths of time individuals spend in various states. Such sojourn times have already been illustrated in Table 1 above, giving earlier estimates of working life expectancy. A major additional capability in LifePaths, given its explicit micro data foundations, is views of uni- or bivariate distributions of durations or sojourn times across the population -- for example the joint distribution of years of school and employment for males and females. (Space limits preclude showing any of these graphs.)

Figure 4 gives one more image from the basic LifePaths simulation -- but this time showing another classification of activities, and using a different horizontal axis. Instead of person-years from a period life table birth cohort as in Figures 1 and 2, in Figure 4 the horizontal axis shows major activities in terms of the number of hours spent in each activity during an average week (i.e. 168 hours) -- for each sex and single year of age. Thus, for example, market work here is shown averaging about 40 hours per week for males age 30 to 55.

**Figure 4 -- LifePaths Time Use (average hours per week) by Major Activity, Age and Sex**



Superficially, Figure 4 looks exactly like data that could be produced directly from a time use survey, and as a matter of validation, it should be very close. However, it was generated by the LifePaths simulation, and differs somewhat from the underlying time use survey data principally because the data have been made coherent. For example, the underlying annual labour force participation rates by age and sex in Figure 4 are consistent with those underlying Figure 1, and the demographic patterns with Figure 2 -- by construction.

One impression left by the diagram is the relatively small proportion of average male and female lifetimes spent in "market work" -- the ostensible domain of the SNA. When viewed from the perspective of average weekly hours (rather than whether more than a third of the year was spent working more than 15 hours per week, as in Figure 1), market work is a very small portion of a total (or even waking) lifetime. Of course, non-market work and the consumption aspects of personal care and use of leisure time also have important economic aspects, but they are not captured in the SNA beyond aggregate dollar measures of personal consumption by commodity.

This figure also indicates the limitations of conventional demographic dependency ratios -- which use raw counts of individuals of working age (e.g. age 20 to 64) as the denominator. In the context of Figure 4, such ratios clearly understate the degree of economic dependence of many individuals in society. The diagram also suggests the need to represent more explicitly the mechanisms by which purchasing power, generated principally by time spent in "market work", is made available to the rest of the population. These mechanisms include intra-family transfers, and government tax/transfer programs. More generally, this combination of a time use with the more conventional demographic framework in LifePaths offers the

opportunity to construct a coherent series of statistical views that provide a much more comprehensive accounting of social and economic activity.

LifePaths images such as Figure 4 clearly show there is much more to life than is captured in the market economy focus of the SNA. It follows that regular publication of this kind of statistical framework could have an important effect on public policy discussion. It would place economic factors in a broader context, and draw attention to a much wider range of impacts of policies directed toward unemployment, retirement, income redistribution, education, childcare, de-institutionalization, and the work week -- to name a few.

It should be emphasized again that these results from the LifePaths statistical framework are still substantially illustrative. The underlying synthetic longitudinal micro data set is still under development. As will be described in the next section, these underlying data are based on a range of recent surveys and analyses -- i.e. real data. But the underlying analyses in part still involve preliminary results.

## 7. Underlying Methods

The LifePaths framework just illustrated draws particularly on two recent data sets, and almost a decade of development of related microsimulation models. The recent data sets are the 1992 General Social Survey (GSS), which includes detailed questions on time use based on 24 hour recall, and the Labour Market Activities Survey (LMAS), which provides detailed longitudinal data on labour market dynamics over the 1988 to 1990 period. The LifePaths microsimulation model in turn is a combination of the results of the GSS and LMAS analyses, the DEMOGEN microsimulation model (Wolfson, 1989) as it has recently been re-implemented in the newly created ModGen C++ microsimulation software environment, and the new post-secondary education Income Contingent Repayment Loan (ICL) model being developed for the Human Resources Development Ministry of the Government of Canada.

This section gives a very brief overview of the processes involved in synthesizing a LifePaths birth cohort, the core of the LifePaths statistical framework. Generally, the synthesis process involves an overall architecture connecting a series of economic and socio-demographic processes, and detailed data analysis to develop empirically based statistical descriptions of each process (i.e. behaviour dynamics).

As in a conventional life table, LifePaths starts with a specified population of individuals, say 100,000 births. Unlike a life table, however, each individual is followed over time until his or her death. (A life table, in contrast, follows groups of individuals, all of whom are considered homogeneous.) At any moment in time, an individual faces a chance of making a transition. Depending on his or her current state or set of attributes, this could be a transition into the labour force, or into a marital union. Which transitions are possible depends on the range of states that are explicitly considered. In the current version of LifePaths, individuals are jointly characterized by the following basic attributes at each point in their lives:

- age -- as a continuous variable
- fertility -- ages at the birth of children, presence of children in the familial home
- nuptiality -- unattached, in a common-law or marital union, separated, or divorced
- work status -- including labour force participation and employment status (hours per week, weeks in the year)
- school status -- grade and type of institution if attending, educational attainment
- work income -- hourly rate, weekly and annual earnings
- time use -- categories shown in Figure 4 plus finer disaggregations
- program participation -- including welfare, unemployment insurance, public pensions
- spouse attributes -- including age, educational attainment, labour market experience

In addition, a wide range of derived attributes can be constructed from these basic attributes such as the variables shown in Figures 1 to 4.



Given this listing of attributes, the next step in describing LifePaths is the processes by which the trajectories for each attribute is generated. A brief sketch is given in the following paragraphs.

Demography -- Fertility is modeled as the sequel to conception, which in turn is modeled as a series of piecewise constant hazard rates, conditional on age, marital status, and number of previous live births. The main data source is birth registrations, supplemented by data from the 1983 Family History Survey to account for biases arising from conception while single or in a common law union, followed by marriage before the birth of the child. Mortality rates are conditional on age, sex and marital status, and are based on death registrations. In both cases, the population census provides the denominators.

Union formation and dissolution are represented by a series of hazard functions. From the single state, there are competing risks of entering a common-law union or a legal marriage. Marriage breakdown involves risks of separation and subsequent divorce. These hazards have been separately estimated for men and women, and depend in a complex way on previous history. For example, females' "risk" of entry to a union is positively related to being pregnant, and is highest shortly following labour force entry. Risks of separation for females are higher if there are no young children at home, if the woman was a teenage bride, and if the woman has recent work experience.

Educational Progression -- Transition rates for progression through elementary and secondary school were constructed to be as close to jointly consistent as possible with the 1986 and 1991 population census data on the school attendance rates of children of the relevant ages. Progression through post-secondary institutions (colleges, trade schools, universities) is based on hazard rates jointly estimated from the National Graduates Survey (NGS), administrative data on school enrollments, and the Labour Market Activities Survey (LMAS) for cases where young people quit work to return to and continue their studies.

Labour Market -- Labour market experience is simulated in two main parts -- whether or not employed, and earnings from employment. The first of these, transitions into and out of employment, is estimated from the LMAS separately for males and females, and also separately for first entry, second and subsequent entry, and exit from employment. First entry is represented by waiting time distributions, while the other transitions are represented by multivariate hazard functions. Sex and educational attainment are important determinants of the waiting time to first employment. Re-entry hazards depend on sex, educational attainment, and duration of the current spell of non-employment, and for women the presence of infant children has an additional depressing effect.

Earnings are in turn based on employment status as just described, and separate models for weekly hours of work, and hourly wages. Upon first entry to employment, a weekly hours value is randomly assigned drawn from an age-, sex- and educational attainment-specific distribution, in turn based on data from a combination of the NGS, LMAS, and the Survey of Consumer Finances (SCF -- the annual household income distribution survey). Subsequently, the weekly hours variable is updated as a function of age, sex, last year's weekly hours, and educational attainment. At the same time that weekly hours is assigned, each individual is assigned a percentile rank for hourly earnings. The hourly earnings rate is then "looked up" from age-, sex- and educational attainment-specific distributions. Percentile ranks are adjusted from year to year based on estimates of rank order "churning" from the LMAS.

Daily Time Use -- The 1992 General Social Survey (GSS) collected 24 hour time use diary data for about 9,000 individuals, evenly distributed by age, sex, day of the week, and month of the year. The GSS also collected basic data on educational attainment, employment status, and family status. After extensive analysis of these data, a LifePaths module was created which imputes to every simulated person-day one vector of time spent over a 24 hour period in each of a series of activities, including at the highest level of aggregation the categories shown in Figure 4. (Special assumptions have been made for children under age 15 and those elderly living in institutions, since they were not covered by the GSS.)

The statistical analysis indicated that age, sex, day of the week, marital status, presence of young children, educational attainment, and main activity (i.e. student, employed or self-employed, other) were all significantly associated with these vector patterns. Thus, all of these attributes, as generated by other LifePaths processes, were used in the imputation. The imputation process was also designed to reproduce

the observed variability in time use patterns amongst individuals with the same attributes, essentially by using the distribution of vector residuals from a multivariate regression analysis.

## **8. Validation and Data Quality Concerns**

Validating the LifePaths model is fundamentally impossible. The reason, simply, is that its intent is to create an instance of a sample from a hypothetical birth cohort. Thus, no comparison with “reality” is ever possible. However, the synthetic microcosm of individual life paths should, by construction, reproduce the major marginal joint distributions from which it was built -- for example, labour force participation rates, fertility rates, mortality rates, union formation and dissolution rates, educational enrollment rates, and age/sex-specific distributions of labour market earnings.

During the course of constructing the LifePaths prototype described in this paper, all these comparisons have been continually checked. By and large, agreement is good. The main instances of disagreement arise when the underlying data sources are not themselves consistent with each other. If anything, this is a signal of error in the source data. In effect, LifePaths has provided a framework for socio-economic micro data, in part analogous to the SNA framework, wherein data from diverse sources are rendered coherent, and inconsistencies thereby highlighted.

## **9. Concluding Comments**

This paper started with users’ needs for more comprehensive and coherent socio-economic statistical information, and offered a diagnosis of the failures of earlier international efforts to address these needs. A new approach is suggested, based on much more extensive use of a range of multivariate micro data sets, and microsimulation methods. Initial features of the approach -- particularly comprehensiveness and coherence -- have been illustrated with preliminary results from the LifePaths model under development at Statistics Canada.

Space has not permitted other features to be graphically illustrated, such as the explicit micro data foundations and hence the capacity to display variety. Further work is required to illustrate other key features such as summary indicators (e.g. lifetime income distributions), and “what if” simulations. Still, the results presented constitute a substantial “proof by construction” of the practical and technical feasibility of the approach.

At the same time, the approach highlights gaps and weaknesses in existing socio-economic statistical data, particularly from a microanalytic perspective. The LifePaths approach would place much stronger demands on the coherence and quality of underlying socio-economic surveys and data collection systems. Given a measure of acceptance of the benefits for socio-economic statistical reporting of something like the LifePaths approach, it can provide the basis for strategic planning in national statistical agencies.

## References

- Bordt, M., G. Cameron, S. Gribble, B. Murphy, G. Rowe, and M. Wolfson (1990), "The Social Policy Simulation Database and Model: An Integrated Tool for Tax/Transfer Policy Analysis", Canadian Tax Journal, 38:48-65.
- Citro, C.F. and E.A. Hanushek (1991), Improving Information for Social Policy Decisions. The Uses of Microsimulation Modeling, National Academy Press, Washington, D.C.
- Easton, G.S. and R.E. McCulloch (1990), "A Multivariate Generalization of Quantile-Quantile Plots", Journal of the American Statistical Association, June, vol. 88, No. 410, Theory and Methods, pp376-386.
- Garonna, P. (1994), "Statistics facing the concerns of a changing society", Statistical Journal of the United Nations ECE, Vol 11, No. 2, pp147-156
- Gnanasekaran, K.S. and G. Montigny (1975), Working life tables for males in Canada and Provinces, 1971, Statistics Canada Catalogue 71-524E Occasional, Ottawa.
- Juster, F.T. and K.C. Land (1981), "Social Accounting Systems: An Overview" in F.T. Juster and K.C. Land (Eds), Social Accounting Systems -- Essays in the State of the Art, Academic Press, New York.
- Juster, F.T., P.N. Courant and G.K. Dow (1981), "The theory and measurement of Well-Being: A Suggested Framework for Accounting and Analysis" in F.T. Juster and K.C. Land (Eds), Social Accounting Systems -- Essays in the State of the Art, Academic Press, New York.
- Mathers, C. and J-M Robine (1993), "Health expectancy indicators: a review of the work of REVES to date", in J-M Robine, C.D. Mathers, M.B. Bone, I. Romieu (Eds), Calculation of Health Expectancies: Harmonization, Consensus Achieved and Future Perspectives, INSERM / John Libby Eurotext Ltd., Vol. 226.
- Moser, Sir C. (1973), "Social Indicators -- Systems, Methods and Problems", Review of Income and Wealth, Series 19, No.2, June, pp133-141.
- OECD (1976), Measuring Social Well-Being, Paris.
- OECD (1977), "Basic Disaggregations of Main Social Indicators", D.F. Johnston, Special Studies No. 4. The OECD Social Indicator Development Programme, Paris
- OECD (1982), The OECD List of Social Indicators, Paris.
- Pommier, P. (1981), "Social Expenditure: Socialization Expenditure? The French Experience with Satellite Accounts", Review of Income and Wealth, December.
- Pyatt (1990), "Accounting for Time Use", Review of Income and Wealth, Series 36, No. 1, March, pp 33-52
- Rowe, G. and S. Gribble (1994), "Income Statistics from Survey Data: Effects of Respondent Rounding", forthcoming in Proceedings of the American Statistical Association, Section on Government Statistics.
- Ruggles, N. and R. Ruggles (1973), "A Proposal for a System of Economic and Social Accounts", in M. Moss (ed.), The Measurement of Economic and Social Performance, National Bureau of Economic Research, New York.
- Ruggles, R. (1981), "The Conceptual and Empirical Strengths and Limitations of Demographic and Time-Based Accounts", in F.T. Juster and K.C. Land (Eds), Social Accounting Systems -- Essays in the State of the Art, Academic Press, New York.
- Stone, R. (1973), "A System of Social Matrices", Review of Income and Wealth, Series 19, No.2, June, pp143-166..
- United Nations (1975), Towards a System of Social and Demographic Statistics (SSDS), Studies in Methods, Series F, No. 18, ST/ESA/STAT/SER F/18, New York.

- United Nations (1979), The Development of Integrated Data Bases for Social, Economic, and Demographic Statistics (IDBs), Studies in Methods, Series F, No. 27, ST/ESA/STAT/SER F/27, New York.
- Vanoli, A. (1994), "Extension of National Accounts: opportunities provided by the implementation of the 1993 SNA", Statistical Journal of the United Nations ECE, Vol 11, No. 3, pp183-191
- Wilk, M.B. (1987) "The Concept of Error in Statistical and Scientific Work", paper presented to the U.S. Bureau of the Census Third Annual Research Conference, Baltimore.
- Wolfson, M.C. (1979), "Saving for Retirement: How Much is Required" , Volume II, Appendix 18 in The Retirement Income System in Canada: Problems and Alternative Policies for Reform, Task Force on Retirement Income Policy, Ministry of Finance, Ottawa.
- Wolfson, M.C. (1989), "Divorce, Homemaker Pensions, and Lifecycle Analysis", Population Research and Policy Review, 8: 25-54.
- Wolfson, M.C., S.Gribble, M.Bordt, B.Murphy, and G.Rowe (1989), "The Social Policy Simulation Database and Model: An Example of Survey and Administrative Data Integration", Survey of Current Business, 69, 36-40.
- Wolfson, M.C. (1994), "Implications of Evolutionary Economics for Measurement in the SNA, Towards a System of Social and Economic Statistics", paper prepared for the Twenty-Third General Conference of the International Association for Research in Income and Wealth St. Andrews, New Brunswick, August, 21-27, 1994, mimeo, Statistics Canada, Ottawa.