DYNAMIC MICROSIMULATION OF HEALTH CARE DEMAND, HEALTH CARE FINANCE AND THE ECONOMIC IMPACT OF HEALTH BEHAVIOURS: SURVEY AND REVIEW

Martin Spielauer

Socio-Economic Analysis and Modeling Division, Statistics Canada, 100 Tunney's Pasture Driveway Ottawa   K1A 0T6 CANADA; email: Martin.Spielauer@statcan.ca

ABSTRACT: This paper reviews the issues to be faced in attempting to create a microsimulation of health care demand, health care finance and the economic impact of health behaviour. These issues identified via an in-depth review of seven dynamic microsimulation models, selected from an initial set of 27 models in order to highlight the main differences in approaches and modelling options currently adopted. After presenting a brief description of each of the seven selected models, the main modelling approaches are summarized and critically appraised using five main distinguishing criteria. These criteria are the use of alignment techniques, model complexity (as reflected in the range of variables used), theoretical foundations, type of starting population, and the extent and detail of financial issues covered. Building upon this appraisal, the paper goes on to show how the '12 SAGE lessons' apply in the field of health care microsimulation. The trade-off between complexity and predictive power is shown to be key. Finally an appendix summarises the main features of all 27 of the dynamic microsimulation models originally surveyed.

1. INTRODUCTION

There have been a number of surveys and reviews of microsimulation models (Merz, 1991; Mot, 1992; Klevmarken, 1997; O'Donoghue, 2001; Zaidi & Rake, 2001; Anderson, 1997; Cassells *et al.*, 2006) with different scopes and different purposes. In some cases these reviews have been undertaken to draw and document lessons from existing microsimulation models as a first step towards new model development. (E.g. for FAMSIM+: Spielauer, 2003; SAGE: Zaidi and Rake, 2001; APPSIM: Cassells *et al.*, 2006). The purpose of this paper is to capitalize on the expertise acquired by what is now four decades of dynamic microsimulation model development with regard to modelling health care demand, health care finance and the economic impact of health behaviour. Based on an extensive literature search, 27 dynamic microsimulation projects were identified for which documentation is available. A short description and classification of all 27 of these projects is given in the appendix of this report. Seven projects are reviewed in more detail: DYNASIM, CORSIM, DYNAMOD, LifePaths, MOSART, and NCCSO. All of these seven models include health-related variables. However, the range of health-related issues that can be studied using these models varies widely, as health is not the central focus of the majority of the models. Consequently, this review does not exclusively concentrate on the treatment of health issues in microsimulation. Rather, the selection of models was made with the intention of reviewing key differences in approach towards dynamic microsimulation, thereby raising generic issues believed to be of direct relevance to health modelling. After providing a brief description of each of the selected models, this paper summarizes and critically appraises their modelling approaches using five distinguishing criteria. These are the use of alignment techniques, the model's complexity and range of variables used, the theoretical foundation of the

model, the type of starting population used, and the extent and detail of financial issues covered.

On the basis of this appraisal, the paper concludes by identifying a series of "lessons" that can be learned from existing projects. A similar approach can be found in Zaidi and Rake (2001), who focus on the simulation of social policies in an aging society and presents 12 lessons based on a review of seven dynamic microsimulation projects. These lessons were used as template for the organization of the conclusions reached concerning the microsimulation modelling of health care finance and health behaviour.

2. REVIEW

2.1 DYNASIM

2.1.1 Overview
The DYNASIM "Dynamic Simulation of Income Model" was the first large-scale dynamic microsimulation model in social sciences. It was developed between 1969 and 1976 under the direction of Guy Orcutt at the Urban Institute (Orcutt, 1957). It was Orcutt who first proposed the concept of developing dynamic microsimulation in social sciences in 1957. Microsimulation was intended to serve as a social science research tool capable of mimicking natural experiments in economics and as a framework for integrating economic and sociological research. The early model was used to analyze Aid to Families with Dependent Children (AFDC) and Unemployment Insurance issues and to develop long-range projections of earning histories for the analysis of social security issues (Anderson, 1997).

A second version - DYNASIM2 – was developed between 1979 and 1983. The base year database was generated by matching the March 1973 Current Population Survey (CPS, n=60,000

persons) with the Social Security earning records for 1951-1972. Selected later data were incorporated until 1993. The simulation horizon is from 1973 to 2030.

A third version – DYNASIM3 – was completed in 2004 and is based on 1990-1993 Survey of Income and Program Participation panels (SIPP, n=100,000 persons). As in DYNASIM2, the main focus lies on pension simulations and issues of population ageing (Favreault and Smith, 2004). Being the model with the longest history, DYNASIM also served as "template" for various other models; its structure – which stayed basically unchanged over the three versions - is therefore explored in more detail below.

DYNASIM is organized in three sub-models that follow different approaches and simulate events of different domains. These are:

– The Family and Earnings History (FEH) model;

– The Jobs and Benefits History (JBH) Model; and

– The Cross-Sectional Imputation Model (CSIM).

The Family and Earnings History (FEH) model is a dynamic microsimulation model of demographic and labour market behaviour consisting of 14 modules corresponding to the events or characteristics simulated. It is a discrete time model with annual updates. The output of the FEH model consists of a file that contains the demographic and labour force histories for each person and cross-sectional files for every (selected) year of the simulation.

The FEH output serves as input for the Jobs and Benefits History model (JBH). Concerning the simulation approach, the Jobs and Benefits History model (JBH) follows a different order: for each individual it simulates the whole life career at once. It contains six sub-models for (1) job characteristics and pension plans, (2) pension eligibility and benefits, (3) social security eligibility and benefits, (4) individual retirement accounts and (5) retirement, and (6) Supplemental Security Income.

The JBH model produces both events, such as job changes, and detailed histories of retirement, disability, spouse and child benefits. The tax-benefit models used are highly parameterized in order to allow for the simulation of various alternative policy scenarios. Taxes and social security contributions as calculated in the last module are only determined for the last simulated year.

The CSIM Cross-Section Imputation model is a static model used to impute additional information into a single cross-sectional file for a given year generated by the other two models. Imputed variables include health status, institutionalization for persons 60+, financial assets including home

ownership, and supplemental security income.

2.1.2 Health
The CSIM sub-model treats institutionalization, disability onset and recovery as events, and also models disability benefits; however, detailed health and disability status is not dynamically modelled, but imputed for a given year. Health status is measured by the number of limitations on activities of daily life (ADLs), and limitations on instrumental activities of daily living (IADLs). The model does not take into account the financing of health care, whether from the public or private sector.

An interesting feature of DYNASIM is its modular structure. Concerning the modelling of health issues, the organization of the model in three sub-models corresponds with three general modelling options: (1) dynamic modelling allowing for interactions with other domains and behavioural response; (2) dynamic modelling on top of other dynamic models with results not feeding back into the other models; and (3) cross-sectional imputation, i.e. the use of microsimulation for the projection of cross-sectional covariates of health models. As noted by Cassells *et al.* (2006:14), one of the lessons from the DYNASIM approach is "that there is clearly a trade-off here between simplicity and the ability to model behavioral responses".

2.2 CORSIM

2.2.1 Overview
CORSIM (Strategic Forecasting, 2002) based at Cornell University and developed under the direction of Steven Caldwell, was initiated in 1987, building on the first dynamic microsimulation model DYNASIM (Caldwell was part of the team that developed DYNASIM at the Urban Institute). The project is now in its fourth generation (Corsim 4.0) and probably also the most heavily "researched" model, since this university-based model has not only been built to (1) simultaneously support basic research into fundamental socioeconomic processes, and (2) provide a platform for a broad range of policy analysis, but it has also been built (3) as a study object itself serving as platform and framework for research in microsimulation modelling. The core CORSIM modules were also widely adapted by other models, namely the Canadian DYNACAN and the Swedish SVERIGE model.

Typical applications include the estimation of welfare costs and the distribution of benefits of welfare reform proposals by Nixon, Carter, Reagan and Clinton, as well as a detailed assessment of Reagan's tax and federal benefit policies over the 1981-1983 period.

The base year database is the 1960 US Census Public Use Micro data Sample (PUMS) containing 180,000 person records. With regard to the behavioural modules, CORSIM aims at synthesizing the empirical strengths of numerous,

diverse data files of various types, including longitudinal micro data – i.e., the Longitudinal Mortality Survey – aggregate totals, cross-section micro data, vital statistics, as well as administrative statistics. CORSIM makes extensive use of grouping of the population into subgroups for which behavioural equations are separately estimated. Concerning data sources and number of equations CORSIM is among the largest microsimulation models. Individual and family behaviour is represented by approximately 1100 equations and 7000 parameters as well as dozens of algorithms. Individual behaviours include schooling, labour supply, demographic characteristics, and risk factors, such as smoking, alcohol or diabetes. Family behaviours and attributes include wealth represented by 11 asset types and 3 debt types, different taxes and benefits, demographic attributes such as family links, and economic behaviour such as consumption and savings.

In contrast to DYNASIM, CORSIM is a fully dynamic, single integrated simulation model. It is organized into approximately 26 behavioural modules and several rule-based accounting routines. Three modules are separable from the main model as their results do not feed back into the model: a voting module, a consumption expenditure module, and a dental module (the second "generation" version of CORSIM developed in 1990-1993 was funded by the National Institute of Dental Research).

One of the most distinctive characteristics of CORSIM is its use of an initial population dating back to more than 40 years before the present. This starting population contributes to CORSIM's usability as a study tool, allowing the modelling and analysis of the socioeconomic processes underlying contemporary cross-sectional distributions, as well as providing a useful means of evaluating the accuracy of model outputs against known historic outcomes. More controversially, CORSIM makes heavy use of alignment techniques to recalibrate its 'projections' for past years to published data. For future years, time series trends in historic alignment factors are rolled forward in order to develop alignment factors for future years. This approach makes projections into the future difficult to interpret, or, as Anderson (1997:[5]19) states, "without realigning or rebasing the data for a recent historic year, projections of future years may begin from a base that already is subject to errors accumulated over a 35 year simulation period". Even if many group and aggregate outcomes can be exactly aligned to recent data, there is no way of assuring that the joint distributions based on the 1960 data remain accurate after 35 years.

2.2.2 Health
Compared to DYNASIM, behaviour is modelled in far more detail both concerning the variables used – e.g., the inclusion of income and wealth in the modelling of fertility and mortality – and the

number of population groups built. Concerning health, CORSIM includes the modelling of four main risk factors, namely smoking, alcohol consumption, sugar consumption and diabetes. It keeps track of disability status, and models institutionalization. As the life-course projection of contributions paid by each person during their working years and the benefits received from the US Old Age Security and Disability Insurance System is one of the main applications of the model, this public system is implemented in detail; the model thus covers disability insurance. Private systems are only covered with regard to dental care, including modules for dental insurance coverage, dental condition/health and dental services and expenditures. CORSIM keeps track of kinship networks among parents and children, among spouses and ex-spouses as well as among siblings, including half and step siblings. This information is valuable for the study of future informal care supply.

2.3 DYNAMOD

2.3.1 Overview
DYNAMOD (King *et al.*, 1999), a dynamic microsimulation model of the Australian population, is designed to project characteristics of the population over a period of up to 50 years. Major elements of the model include demographics, international migration, education, the labour market and earnings.

The DYNAMOD model can be seen as the population simulation module of what was initially conceived as a two-part model, with a separate analysis module being the second part – a design following the DYNASIM2 approach to reduce computing demands. The first analysis tool corresponds to the model's first specialized application as a model for the analysis of student loans.

DYNAMOD uses a "pseudo-continuous" time framework operating in monthly steps for most demographic and labour market processes and in annual steps for education and earnings. Concerning the statistical modelling approaches used, it makes maximum use of survival functions. This design sought to make a balanced trade-off between time interval and computing demands: while it is one of the first models using months as time units, the survival functions only have to be re-evaluated if changes occurred in the characteristics incorporated in these functions. For example, the month of death is determined at birth and stored in what was called the 'crystal ball' (King *et al.*, 1999). This month is only re-evaluated if a change in the health status occurs, as, apart from year of birth, age, sex and disability status, no other variable enters the survival function used.

DYNAMOD concentrates on four broad groups of processes, namely demographics, education, labour markets and earnings. The current version is version 3 (Kelly, 2007a,b). Substantial changes

concern the improvement of alignment of demographics and labour force transitions as well as the addition of various economic modules.

2.3.2 Health

While health issues are not central topics of DYNAMOD, disability is an important variable that, in addition to the mortality function, also enters educational functions.   What makes DYNAMOD interesting concerning the modelling of health is the model approach itself, namely the pseudo-continuous time framework which allows for continuous time competing risk modelling. Overall, the development of DYNAMOD was accompanied by a continuous struggle between ambition and feasibility which is documented in Cassells *et al.* (2006) and makes the model an interesting study object itself.

2.4 LifePaths

2.4.1 Overview

LifePaths (Statistics Canada, 2002) is a dynamic microsimulation model developed at the Canadian Statistical Office that differs considerably from other existing models for four reasons:

- It operates in continuous time which (amongst other things) allows for a more accurate representation of causation and behaviour.

- It is an open model in which new individuals are created for the case of partnership formation, using a concept of "dominant individuals".

- It uses a synthetic initial database: LifePaths uses a variety of historical micro-data sources in order to create representative synthetic life histories from birth to death for all birth cohorts since 1872.

- It runs on a generic simulation language (Modgen), also developed at Statistics Canada (Statistics Canada, 2002).

LifePaths is an integrated "general purpose" model, including demographic behaviors, health, education, labour and allowing for tax-benefit and pension modelling. LifePaths is structured with an explicit event orientation. Behavioural equations together with their stochastic components determine the distribution of waiting times to events. A LifePaths simulation consists of a set of mutually independent cases. Each case contains exactly one dominant individual in the first generation. The spouse and children of the dominant individual are simulated as part of the case and are created to satisfy the marriage and fertility equations. This approach also determines the order of the simulation: LifePaths simulates the completion of one case before going on to the next. Not starting from a cross-sectional database, the number of cases can be set by the user. This allows LifePaths to simulate large numbers of cases – usually millions – which helps to reduce the statistical noise in outputs caused inherent in

any model that uses Monte Carlo sampling.

The continuous time framework allows for a broad palette of behavioural models to be used. While the model closely reproduces past census information and is aligned to central scenarios of future demographic trends, at the same time it includes more 'behavioural modelling' than to be found in many other models. This is visible, for example, in the modelling of fertility. Births are simulated as a sequence of fertility decisions. Each decision is modelled in two parts: first a decision of whether to have a child is taken, and in the positive case a waiting time is generated.

2.4.2 Health

The modelling of disability status is a recent addition to the model and was implemented mainly to allow for studies of future home care services (Keefe *et al.*, 2005; Carriere *et al.*, 2007) and informal support (Wolfson and Rowe, 2004). LifePaths uses its own definition of disability developed specifically for the analysis of future care needs. Four disability states are recognised ranging from no disability through to severe disability; movement between these states is modelled as a set of competing risks. A fifth and terminal state is institutionalization. Covariates of the hazard models are age, educational attainment, living arrangements, age at immigration and recent disability history. The inclusion of health in LifePaths makes it a tool for the study of both the demand and the supply of support (i.e. the existence and composition of family networks including individual characteristics of network members, e.g. labour market participation).

With regards to health issues, LifePaths has proven to be a successful modelling approach in a number of ways: (1) it is flexible enough to have dynamic health models added as demanded; (2), its detailed output may be used as input for cross-sectional imputation models in studies of future care demand and supply; and (3) a simplified version of LifePaths (stripped off the tax benefit models etc.) serves as a starting point for the family of POHEM health models introduced in the following section. What also makes LifePaths interesting as a study object is the public availability of not only the model itself, but its source code, written in the public domain Modgen programming language.

2.5 POHEM

POHEM (Population Health Model) started as a sister model of LifePaths and over time developed into a family of different model variants for different health-related applications. The original POHEM was built on top of the demographic modules of LifePaths, replacing the mortality equations with a highly detailed model of morbidity and mortality. Like LifePaths it is developed at Statistics Canada using the generic microsimulation programming language Modgen. The development of POHEM is project oriented, with projects not just corresponding to new

modules added to the existing model, but also to the creation of new model variants, both building upon existing modules, and incorporating model-specific variations and approaches. This is apparent, for example, in the different types of starting populations used in the different POHEM applications, including both the original LifePaths approach of synthetic populations and the more latterly, survey-based cross-sectional starting populations, built from sources such as the Canadian Community Health Survey.

In contrast to the other dynamic microsimulation models surveyed in this paper, POHEM is a dedicated health model. Typical output includes measures of cost effectiveness, cost per year of life gained, cost per health-adjusted year of life gained and measures on the impact of interventions on incidence and survival. In addition, disease-specific survival curves and costs of care for co-morbidities can be established.

POHEM is used both to assess the economic costs of alternative health intervention, as also the potential impact of those interventions on disease incidence and progression. POHEM models three aspects of health: (1) the progression of risk factors, (2) disease onset and progression; and (3) public health interventions. Risk factors include weight gain, smoking, total cholesterol, blood pressure, and alcohol consumption. Diseases modelled include breast, lung, and colorectal cancers, and acute myocardial infarction (Flanagan e*t al.*, 2003; Maroun *et al.*, 2003; Will *et al.*, 1999, 2000 and 2001).

2.6 MOSART

2.6.1 Overview
MOSART (Andreassen *et al.*, 1994; Fredriksen, 1998, 2003) is a dynamic microsimulation model for Norway developed by Statistics Norway to investigate policy options with regard to financing public expenditure. In its first version, developed between 1988 and 1990, MOSART focused on demographic behaviour, education and labour force participation in order to study the impact of demographic change on labour force and education attainment. A second version extended the model to include the modelling of pensions. Currently MOSART is in a third version that includes more detailed behavioural modules of household formation and disability. MOSART is mostly based upon a set of administrative data and records representing 12% of the Norwegian population. The detailed administrative data available in Norway allowed for a construction of a longitudinal database that contains rich retrospective information on many variables dating back to 1985, and to 1967 for labour income and pension entitlement.

Most events in MOSART are represented by time-invariant transition matrices and logit relationships assuming constant behaviour over time. Time-invariant transition tables are used for leaving home, institutionalization, marriage and cohabitation, matching couples and couple dissolution. Fertility is also assumed to be time-invariant, determined by age of mother, age of youngest child and parity. The only exception currently is mortality rates, which are assumed to continue decreasing over time.

2.6.2 Health
MOSART models a limited set of health-related behaviours, namely moving into or out of old age care institutions, disability and rehabilitation, and public disability pensions. It also models living arrangements. Like other models producing such projections, MOSART it a potential data source for more extended care models. Compared to other models, MOSART distinguishes itself by its simplicity of modelling approach and the use of administrative data. This combination of reliable data and easy to understand models and scenarios made it a trusted tool for policy analysis in Norway.

2.7 NCCSU Long-Term Care Model
A microsimulation model of long-term care charging has been developed at the Nuffield Community Care Studies Unit (NCCSU) at the University of Leicester (Hancock 2000; Hancock *et al.*, 2006). The NCCSU model is the microsimulation component of a micro-macro model and an example of a very different application of microsimulation in the health field. The model is highly specialized in terms of both the population and the individual characteristics simulated: mainly the income and assets of persons aged 65+. In the context of the means-tested long-term care system of the UK, the NCCSU model assesses the future public-private share of care expenditures under alternative policy scenarios.

The NCCSU model is based on data on older participants of the Family Resources Survey 1997 (FRS), a representative sample of British households (n=6,400 individuals 65+). It contains detailed information on incomes, wealth and housing for the non-institutionalized population, i.e. the population representing the future potential entrants to care homes. The NCCSU model simulates alternative policies taking into account the income and wealth position of individuals in means-tested programs. The model simulates the incomes and assets of future cohorts of older people and their ability to contribute towards care home fees, should they need to be cared for in such settings.

In order to project future health care costs, transitions concerning health care needs have to be modelled. The model concentrates on the cost incidence – the simulation of means tests, etc. – and uses exogenous scenarios from macro-projections in modelling future demands. This is done by linking the microsimulation model with the PSSRU (Personal Social Services Research Unit, University of Kent) cell-based long-term care macro-model. PSSRU projects three key variables: the size and age distribution of the future

population 65+ with dependency, their demand for long-term care services, and the cost of those services. When linking the models, this information is used to weight the micro simulated population, thereby assigning the right number of people distinguished by age and other characteristics to different types of residential and nursing homes. Dynamic microsimulation is then used to project eligibility of the institutionalized population for means-tested state support, including simulation of the "running down" of personal assets due to costs associated with care needs, death and the effect of widowhood.

The long-term care micro-macro model has been used to study both the sensitivity of the current care and funding arrangements to alternative assumptions on population ageing and dependency rates, and for the study of policy options and their distributional implications.

3.   A   COMPARISON   OF   APPROACHES: OPTIONS FOR HEALTH CARE MODELS

3.2 Overview
Health models produce or incorporate external projections of various types, including

–   demographic projections in order to account for demographic changes;

–   earnings projections in order to be able to calculate health care contributions;

–   care need projections for different age and risk groups split up for different types of health care needs ranging from medication to personal care;

–   projections of unit costs of health care per type of health care;

–   projections of available informal care, typically taking account of changing demographic trends and levels of (female) labour market participation upon available kinship support networks.

This section focuses instead upon five key aspects of difference between microsimulation models:

–   Models differ in the extent to which the aggregate outcome of the microsimulation projections are trusted, or to which degree these projects are aligned to other (macro) projections.

–   Models differ in the degree of detail and number of variables in general and specifically concerning health-related variables.

–   Models differ in the degree of (explicit) behavioural modelling, i.e., to which extent the models are based on theory versus statistical "black-box" models. Behavioural modelling in the context of health will typically include the exposure to different risk factors (such as smoking and alcohol) and take into account the whole individual health history when determining specific health risks.

–   Models differ in the way the base population is created. It can be derived from a cross-sectional sample or by creating a synthetic population from other sources of information.

–   Models differ in the extent to which they include health care finance issues, including social security policies and accounting detail.

In addition, health models usually include policy simulations of benefit contribution rates and means-tested deductibles, as well as a wide range of accounting routines, although the microsimulation projects surveyed above differ considerably with regard to the extent to which they produce these forecasts themselves or incorporate external information and forecasts. A similar comment may be made concerning the degree to which each of the models considers interactions between demographic, health and economic processes are considered.

3.2 Alignment to other models
The first point concerns the extent to which internally produced projections are used (and trusted) with regard to the projected aggregates. Many models, like DYNASIM, CORSIM and its various successors, make heavy use of alignment methods in order to align the model's aggregate projections to external forecasts (or to historic numbers if simulation starts in the past). The latter is the case, for example in COSRSIM, that still uses a 1960s population sample as its starting population. Over the years, time series of adjustment factors were added and, in order to use the model for forecasts, research has concentrated rather on the prediction of alignment factors from this time series data than on changing the model (Anderson, 1997). In the context of aligned outputs, the 'internal' behavioural equations are therefore used to illustrate the socio-structural effects and distributions of events across socio-demographic groups, while the aggregate results are aligned to external forecasts. There are various reasons for pursuing this approach. The first is 'specification randomness' (Van Imhoff and Post, 1998). The greater the number of explanatory variables for a given attribute or event, the less statistically robust the associated empirical estimate of the predictor function. Specification randomness reduces the aggregate prediction power of a model (in a trade-off against misspecification errors due to the omission of important variables). A way out of this dilemma is frequently seen in the alignment of detailed micro-models to macro-models, which are believed to have a higher predictive power. But the reason for aligning outputs does not always lie in the lack of trust in the model predictions; it can also result from the demands of policy makers who are interested in 'what if studies', the 'if' being the aggregate

output the model is expected to reproduce. A typical example is 'official central scenario' population forecasts. In this respect, the alignment of outcomes is also needed and used to make model results comparable to other modelling approaches by using the same population scenario. As a result of these pressures, alignment methods have also been incorporated into microsimulation models which had initially not been planned to follow this road – the Australian DYNAMOD being a good example. Less complex models, like MOSART, are able to reproduce pre-set scenarios more easily as they apply methods and variables that come closer to conventional cell-based macro models. Thus they can 'internally' reproduce given scenarios by corresponding parameterization.

3.3 Complexity of the models and number of variables used

One of the features that make microsimulation especially attractive, namely the large number of variables models can include, comes at a price. While models that are based on rather simple behavioural models are often more likely to reproduce 'trusted' aggregate projections, they are not only rather weak with regard to explanation, but also limit the analysis of behaviour to this reduced set of variables. Simple models do not allow for the inclusion of many of the socio-economic variables regarded as important in the modelling of health behaviour. As the excluded socio-economic characteristics must therefore be assumed independently of health, these models might produce quite biased joint distributions when these additional characteristics are included into the analyses. This is especially a problem in tax-benefit and health care finance analyses, since health is highly related with a series of other socioeconomic variables including education and income. This might not be a problem for forecasts in the short term if the base population comes from a recent representative sample, but it generates a trade-off between good aggregate predictions and a good prediction concerning distributional issues in the long run. This leads back to the heavy reliance on alignment techniques as used in the models mainly built for policy analysis like CORSIM, DYNACAN or DYNAMOD. In this respect, the NCCSO model represents the 'extreme case' of specialization, as it entirely concentrates on the projection of income and wealth distribution of pensioners (used for the means test of care policies), but leaves the modelling of population numbers by age and care need to a cell-based macro-model which the results of the microsimulation model are fed into.

Independent of model detail and the number of variables integrated in the microsimulation model, socioeconomic long-term projections still require exogenous assumptions. This is especially true for macroeconomic variables like future wages and prices. Various approaches exist in order to link micro-models of the household sector with macro-economic models, the German Darmstadt Micro

Macro Simulator (DMMS) being an early example (Heike *et al.*, 1994). But in practice, combining the strengths of micro- and macro-approaches by developing integrated 'all purpose' micro-macro models has often turned out to be expensive, both concerning development costs and model transparency. In the development of DYNAMOD such plans have been dropped. In contrast, the much simpler MOSART model was recently linked to a large-scale macroeconomic Computational General Equilibrium (CGE) model (Fredriksen and Stolen, 2007).

3.4 Theoretical foundation of behavioural models

The weak theoretical foundation of many microsimulation models is a common source of critique (Klevemarken, 1997). This topic is closely related to the intended use of a model – prediction versus explanation – as a good theoretical foundation usually does not go hand-in-hand with the predictive power of a model. Discrete time models typically use either transition tables or (usually logistic) regression models. All of these models can be regarded as typical "blackbox" models, as, apart from the selection of appropriate explanatory variables, little or no theoretical foundation is given. LifePaths deviates from this approach, as it introduces more 'behaviour' in its modelling of fertility which is modelled as a sequence of fertility decisions, as distinguished from the statistical modelling of the waiting time until birth after a decision was made. This might be a very useful departure point in order to introduce agent-based behaviour, such as goal orientation, and explicit models of decision making into microsimulation (Vencataswawmy, 2002). Generally, the inclusion of explicit behaviour is supported by time-frameworks that avoid reliance solely upon transition models. The Australian DYNAMOD model gives a very interesting example in this respect. With its pseudo-continuous time framework (of monthly steps) and the ability to store future events (in what is called the "crystal ball"), whose effective occurrence might be reassessed as circumstances change, a variety of ways of modelling behaviour are opened up. This, of course, holds true also for continuous time models like LifePaths and POHEM.

3.5 The starting population

Dynamic microsimulation models usually take as a starting population a cross-sectional database representation of the population simulated. Another approach is to simulate all members of the population from birth. A synthetic generation of a full population can be found in the LifePaths model. This approach is typically chosen when estimates of population characteristics are required that are not contained in survey information. Kinship patterns are a good example, with the work of Wachter (Wachter, 1995, 1998; Wachter *et al.*, 1998), who restored the kinship patterns of the US population using the SocSim software, being perhaps the most prominent example. Other applications using this approach are the simulation of wealth accumulation and

distribution including bequests.

Other than LifePaths and a few POHEM applications, all of the models surveyed start with a population derived from a survey or – as in the case of MOSART – from administrative data. Depending on the retrospective detail of the data used for the generation of the starting population, it can still be necessary to restore missing information by simulation and/or by commencing the simulation at a point in time well before the present day. This is done in CORSIM, which is based on 1960 data and simulates earning and other histories from this year onward.

A special situation arises when modelling a newly introduced health care system that (initially) does not cover the whole population. As microsimulation allows for individual accounting, it might be especially useful to study sustainability issues in the presence of transition dynamics – that is, in situations in which the health risk patterns in the initial phase of the transition might differ considerably from the long-term pattern. While the starting population (or the population from which individuals enter the social security system) will typically be generated from survey data in the initial phase, individual data records might be successively replaced as hard data become available.

3.6 Policy simulation and accounting
Given sufficient micro-level information microsimulation allows the modelling of policies at any level of detail, which is vital if attempting to model policies that link taxes and benefits in a non-linear way to individuals, or if individual contribution histories are relevant, as in the calculation of pensions. This also applies when calculating benefits and taxes that depend in part upon spousal and family characteristics, such as the calculation of survivors' pensions which may depend in part upon the contribution history of the deceased partner. In the case of health care studies, individual and familial income and accumulated financial assets are often of great relevance to means-tested care policies. But of central importance is the availability of kin, drawn from both within and beyond the household, to fulfil the role of informal care-giver.

Policy microsimulation is both a modelling and an accounting exercise. In its static dimension modelling involves the tax-benefit regime itself and some behavioural aspects like take-up rates of benefits. It concentrates on the first-order effects of policies, which might include the calculation of some measures of 'pressure on behaviour', such marginal tax-rates. (Immervoll and O'Donoghue, 2001). Dynamic microsimulation can additionally include behavioural models of policy response, ranging from changes in labour supply to decisions concerning the timing of retirement or durations of parental leave, to give some examples.

Being based on micro-units is not only the major

characteristic but also the major advantage of microsimulation models for social and economic policy analysis as they produce results which can then be analyzed at the individual level. Thus, the distributional impact of a policy measure across different types of families or different geographical regions can be assessed. At the same time, estimates of the aggregate outcomes can still be derived easily, by summing the individual results.

National tax-benefit models became standard tools for the calculation of costs and the (distributional) impact of policies in many countries; in recent years also efforts were made to build models that allow comparative analysis across countries, the EUROMOD project (Sutherland, 2001) being a prominent example. Economic and policy applications of dynamic microsimulation models usually also include cross-sectional analysis with static microsimulation being one "dimension" of these models. This view of static microsimulation as being one dimension of dynamic models might be justified on the basis of the various attempts to extend existing static models to dynamic models, both by including feedback behaviour – where the calculation of "pressure on behaviour" is one initial step in this direction – and behaviour over time. An example of combining a dynamic microsimulation model with an existing static tax-benefit model is the dynamic microsimulation model developed by O'Donoghue (2001) for Ireland which can "communicate" with the static EUROMOD model.

All surveyed microsimulation models include policy simulations to various degrees. The main distinction can be found concerning the availability of individual information for each single period. DYNASIM is a typical example of a model in which not all information is generated and available dynamically, but some variables are imputed to a simulated cross-section of a given year. Integrated models like CORSIM have the added benefit that they can include routines to calculate internal returns to contributions and can therefore serve to assess distributional issues over the whole life-course and between generations. In the health field, this longitudinal dimension is often of key importance, as the policy impact of interventions on incidence and survival is often long-term in nature. A model which makes wide use of these strengths of dynamic microsimulation for the assessment of the cost-efficiency of alternative health interventions in POHEM, for example by calculating measures such as the cost per year of life gained.

4. THE 12 'SAGE LESSONS'

The SAGE research group (Zaidi and Rake, 2001) have drawn up '12 lessons' for microsimulation modellers engaged in the creation of a new microsimulation model, with a particular focus on the simulation of social policies in an aging society. Building upon the in-depth survey of seven dynamic models presented above, this

paper now revisits these lessons, having particular regard to their applicability to health care modelling.

The surveyed models differ considerably in the number of processes that have been modelled and therefore in comprehensiveness. Comprehensiveness and complexity comes at the price of making it difficult to interpret results and to separate out the impact of individual processes. Zaidi and Rake conclude in this context that the effectiveness and suitability of a dynamic microsimulation model has to be judged in relation to the purpose for which the model was built; they summarize this statement in the first of their 12 lessons:

> "A successful model requires clear objectives. From these objectives, model builders can identify the processes which are essential to the model and design a developmental strategy for the model, whereby other processes are incorporated over the longer term." (Zaidi and Rake, 2001:18).

Dynamic microsimulation in the field of health care studies can be seen, or should be designed as, a tool for the investigation of health related processes, supporting the conceptualization of these processes and the study of their determinants and consequences. Consequently, such a model has to include the modelling of changing health attributes over time alongside the modelling of core demographic processes. If the objective is to project future health care demands, the model should also ideally be able to produce its own forecasts of future population aggregates without being aligned to other projections. Health care studies are concerned with and have to take into account a wide field of social and economic changes that have a strong impact on health issues. In order to design a dynamic microsimulation model as an appropriate tool in health care studies, it has to include additional relevant processes and variables in a way that makes it either a comprehensive model or a model that produces a detailed and adequate population input for other models. In both cases microsimulation can be the appropriate modelling approach, as it adds flexibility in the modelling of dynamics and increases the range of variables compared to cell-based models.

Comparing the surveyed projects, a clear trade-off can be observed between the socioeconomic detail included to carry out detailed tax-benefit calculations and the predictive power of the models in the long run. This can also be seen as a trade-off between detail in cross-sectional analysis and the suitability and transparency of a model for the study of (health related) processes in the long term. Health care studies focus on both distributions and processes, especially with regard to policies, as a detailed calculation of the costs and distributional impacts of health care policies at a given point in time might be equally as important as the study of long-term effects. A possible way to avoid such a trade-off might be to design a microsimulation model as a modelling platform rather than as one single model, suitable to include different degrees of detail depending on the projection horizon.

A problem of all data-based microsimulation models is the availability of data. In this respect, the "model builders need to be sensitive to the shortcomings of data […]" (Zaidi and Rake, 2001:18), and "the model should be flexible enough to incorporate the most recent and robust data" (Zaidi and Rake, 2001:18) – are essentially lessons 2 and 3 from Zaidi and Rake. Concerning data, health care models will typically make use of a wide range of data sources. In this process, available survey data may often be only a starting point, especially when modelling social security funds, as social insurance agencies maintain huge databases of individual contribution and spending histories. Similarly, the modelling of specific health attributes/processes may draw upon aggregates or functions developed using clinical records collected as part of detailed medical research projects. Maintaining a high flexibility in order to allow for the incorporation of most recent data is a key requirement in health care modelling, allowing the integration of the latest, and possibly more detailed, data from cross-sectional and longitudinal surveys.

Zaidi and Rake's lesson 4 is superficially intuitive, but repays closer attention. "Innovation in model building may be desirable, although it involves taking risks, with parts of the model building process having unknown rewards and pitfalls." (Zaidi and Rake, 2001:19)

A topic related to the comprehensiveness of models, as discussed above, is whether models are used and designed to produce input to other models and if so, whether this combination of models involves feedback reactions. In the wide area of health care studies, microsimulation can be useful in each of three cases: (1) as a 'stand-alone' tool to study and project population and health dynamics including disease episodes, care seeking and insurance claims; (2) as a method that can produce a more detailed population input to other models, which could be achieved by the cohort-component method; or (3) as one side of an integrated micro-macro-model where results of one side feed into the other and vice versa, such as in a model of the interactions between population and environment. Although this third option of creating an integrated micro-macro model is potentially attractive in many ways, design of such models has in reality turned out to be expensive, with regards to both development costs and model transparency. The experience of DYNAMOD can serve as an example of model builders ultimately preferring to allow for flexibility in specifying external aggregates. Zaidi and Rake conclude in their fifth "lesson" that "[...] Simpler solutions, in the form of taking macroeconomic indicators from external sources and performing sensitivity analysis may be preferable in the

short/medium term." (Zaidi and Rake, 2001:19) This might equally apply to health care studies, at least as long as feedback reactions are not the central focus of the analysis itself.

Models that include the projection of future unit costs and other financial outcomes will typically result from a combination of micro- and macro-sub-models or modules. An appropriate model composition might consist of three such sub-models, namely

– an accounting module reproducing the balance sheet of health care providers and the insurance fund;

– a macro-economic module, incorporating macro models such as general equilibrium models, 'fixed-coefficient' social accounting models and hybrid approaches in which some prices adjust but others do not; and

– a household module, based upon a microsimulation model employed to produce a population of individuals characterized by individual-level behaviours related to demography, labour supply, education, morbidity, health care seeking and making insurance claims and insurance contributions.

When designing a microsimulation model as an integral part of a health care finance model, clear interfaces to other modules or models should be defined from the very beginning, allowing to include external information in a defined way, independent of how this information was generated, whether by model or by assumption. Rather than opting for 'simpler solutions' with regard to the determination of macroeconomic variables, the development of the microsimulation model should be made as independent as possible of the modelling choices made with respect to these macroeconomic indicators through the provision of clear interfaces for the data exchange between the modules.

Another issue of concern is the appropriate time frame to use. Zaidi and Rake conclude in their sixth lesson:

> "Limits of data, and the difficulties of modelling 'continuous time' mean that a traditional structure may be preferable. However, it may bring dividends to introduce innovations into a traditional structure. For example, the feasibility of looking at certain events on a shorter timescale (e.g. monthly) should be explored. In addition, hazard rates and survival functions should be examined" (Zaidi and Rake, 2001:20).

The latter has been done by various authors including Galler (1997) and Vencatasawmy (2002). With regard to the use of microsimulation in health studies, a continuous or pseudo-continuous timeframe (e.g. of monthly steps)

might be the most appropriate choice, as it allows for various modelling approaches also including hazard rates and survival functions. In this respect, both the Canadian LifePaths/POHEM and the Australian DYNAMOD project can serve as interesting examples, as already remarked in Section 3. It should be noted that many design choices for a yearly timeframe have not been made with respect to data availability or modelling considerations, but rather in order to avoid the high computational demands of shorter time intervals, limitations that might already have been removed given advances in computing hardware. Concerning the surveyed models, the models that focus most on demographic processes and health, LifePaths and POHEM, use a continuous timeframe.

The second time dimension refers to the period over which models operate. Zaidi and Rake conclude in their seventh lesson:

> "Producing output that covers the short and the medium term as well as the longer term is an essential way of ensuring that the model remains credible. In setting the end date of the model attention needs to be paid to known demographic transitions and the life-span of policy reforms in order to show its full impact." (Zaidi and Rake 2001:20).

Most processes that include and result from demographic changes evolve over many decades rather than years, and projections of 50 to 100 years are quite common in social security and health care projections. As many phenomena that can be observed today are the result of past dynamics, one frequently also has to look back in time. Microsimulation in this respect can also serve as tool to 'restore the past'. A historical starting date as used in CORSIM may be chosen both as a way of validating the model and as (sometimes the only) way to impute characteristics of today's population otherwise not available, such as kinship networks and histories of past contributions to social security systems. In demographic research, microsimulation has also been used to restore historic populations (Wachter, 1995, 1998; Wachter *et al.*, 1998).

Lesson eight is again derived from data considerations, stating that the representativeness of the base data is of greater importance than its detail. The choice of the appropriate model complexity and detail is a difficult one. Model builders often tend to produce over-ambitious models based on detailed – but small – surveys with model results then not always 'trusted' in actual policy debate. This also applies to health care modes. Overall, data availability and quality improved considerably over the last decades, especially concerning longitudinal data, what makes microsimulation an increasingly promising modelling option in the health field, given a right choice of data and model detail.

The next two lessons deal with model validation, rather generally stating that "[...] sensitivity analysis as a way of estimating the impact of specific parameters on model output and is a first step in validating a model" (Zaidi and Rake, 2001:21) and that "[...] operating a retrospective microsimulation model is one attractive, although not complete, way of establishing its validity." (Zaidi and Rake, 2001:22). This is definitely true also for microsimulation applied in health care studies, as are the following and last two lessons, the first highlighting the necessity for thorough and clear model documentation and the last specifying the need for a computing strategy "to be developed alongside the microsimulation strategy." In this respect, LifePaths clearly distinguishes itself, as the development of this model drove the development of the microsimulation language Modgen (Statistics Canada, 2007). Subsequently, Modgen has grown into a generic microsimulation language (technically a superset of the C++ programming language) supporting most modelling approaches, including continuous and discrete time models as well as case-based and time-based models. This makes Modgen an interesting programming option for new model developments.

5. SUMMARY

This paper has provided an in-depth review of a selected sample of dynamic microsimulation models, with a view to identifying the modelling approaches and options most applicable to the microsimulation of health care demand, health care finance and the economic impact of health behaviour. The surveyed models differ considerably in scope, complexity, comprehensiveness, theoretical foundation and predictive power, as well as in accounting detail. None capture fully all aspects of health care, which include both demand (disease burden and progression) and supply (the financing and provision of health care). Key differences in approach centre around the use of alignment techniques, model complexity (range of variables used), theoretical foundation, the type of starting population and the extent and detail of financial issues covered.

With regards to the modelling of health-related issues, it is clear that a series of tradeoffs have to be considered, in particular between the detail of the model and its overall predictive power. It has been noted that 'large' and 'general' models, such as DYNASIM and its successors, make heavy use of alignment methods. More specialized models developed and designed for the study of demographic and health-related processes, such as LIFEMOD and POHEM, differ considerably from these general models with regard to the types of behavioural models used, the treatment of time and the underlying data. The application of the 12 'SAGE lessons' to the problem of modelling health care reinforces the importance and nature of this trade-off, highlighting in addition the need for model modularity and the adoption of an appropriate time-frame. The ideal microsimulation health care model remains to be created, but between them the approaches adopted by POHEM and DYNAMOD perhaps offer the best clues for the way forward.

ACKNOWLEDGEMENTS

REFERENCES

Anderson J M (1997) Models for Retirement Policy Analysis, Report to the Society of Actuaries. (http://www.soa.org/research/pension/research-models-for-retirement-policy-analysis.aspx) [accessed 1 September 2007]

Andreassen L, Fredriksen D and Ljones O (1994) 'The Future Burden of Public Pension Benefits A Microsimulation Study', Discussion Papers 115, Statistics Norway.

Blanchet D and Crenner E (2006) 'Simulating pensions in France : the DESTINIE model', Paper presented at the Colloquium of the AIM (adequacy of old age income) project, CEPS, Brussels, Nov 6-7.

Brouwers L (2005) 'MicroPox: a large-scale and spatially explicit microsimulation model for smallpox planning', in Ingalls V (ed) *The Proceedings of the 15th International Conference on Health Sciences Simulation*, San Diego: SCS, 70–76.

Butrica B A, Iams H M, Moore J H and Waid N D (2001) 'Methods in Modeling Income in the Near Term (MINT I)', ORES Working Paper Series No 91, Social Security Administration, Office of Policy, Washington DC.

Carriere Y, Keefe J, Legare J, Lin X and Rowe G (forthcoming) 'Population aging and immediate family composition: implications for future home care services', *GENUS*, LXIII(1).

Cassells R, Harding A and Kelly S (2006) 'Problems and Prospects for Dynamic Microsimulation: A Review and Lessons for APPSIM', NATSEM Discussion Paper 63, National Centre for Social and Economic Modelling, University of Canberra.

Ericson P and Hussénius J (2000) 'Studiebidragen i det långa loppet', Rapport till Expertgruppen för studier i offentlig ekonomi, Ds 2000: 19.

Favreault M and Smith K (2004) 'A Primer on the Dynamic Simulation of Income Model

(DYNASIM3)', The Retirement Project, Urban Institute, Washington.

Flanagan W, Le Petit C, Berthelot J-M, White K J, Coombs B A and Jones-McLean E (2003) 'Potential impact of population-based colorectal cancer screening in Canada', *Chronic Diseases in Canada*, 24(4), 81-88.

Flood L (2004) 'Formation of Wealth, income of capital and cost of housing in SESIM', SESIM working paper, Ministry of Finance. (http://www.sesim.org/Documents/Wealth.pdf) [Accessed 1 September 2007]

Flood L, Jansson F, Pettersson T, Pettersson T, Sundberg O and Westerberg A (2005) 'SESIM III - a Swedish dynamic micro simulation model', (http://www.sesim.org/Documents/Handbook%20of%20SESIM.pdf) [accessed 1 September 2007]

Fredriksen D (1998) 'Projections of Population, Education, Labour Supply and Public Pension Benefits - Analyses with the Dynamic Microsimulation Model MOSART', *Social and Economic Studies*, 101, Statistics Norway. (http://www.ssb.no/emner/02/03/sos101/sos101.pdf) [accessed 1 September 2007]

Fredriksen D (2003) 'The MOSART model - a short technical documentation', Paper presented at the International Conference on Population, Ageing and Health: Modelling Our Future, Canberra, December 2003. (https://guard.canberra.edu.au/natsem/conference2003/papers/pdf/stoelen_nils-2.pdf) [accessed 1 September 2007)

Fredriksen D and Stolen N (2007) 'MOSART', in Gupta A and A Harding (eds.) *Modelling our Future: Population Ageing, Health and Aged Care*, International Symposia in Economics, Volume 16, North-Holland, Amsterdam, 433-437.

Galler H P (1997) 'Discrete-Time and Continuous-Time Approaches to Dynamic Microsimulation Reconsidered', NATSEM Technical Paper 13, National Centre for Social and Economic Modelling, University of Canberra.

Hancock R (2000) 'Charging for Care in Later Life: Analyzing the Effects of Reforming the Means Test', Working Paper NF86, Nuffield Community Care Studies Unit, University of Leicester.

Hancock R, Wittenberg R, Pickard L, Comas-Herrera A *et al.* (2006) 'Paying for Long-Term Care for Older People in the UK: Modelling the Costs and Distributional Effects of a Range of Options', Discussion Paper 2336, London School of Economics.

Heike H-D, Beckmann K, Kaufmann A and Sauerbier T (1994) 'Der Darmstädter Mikro-Makro-Simulator - Modellierung, Software Architektur und Optimierung' in Faulbaum F (ed.) *SoftStat'93 - Advances in Statistical Software 4*, Fischer, Stuttgart, 161-169.

Holm E, Holme K, Mäkilä K, Mattson-Kauppi M and Mörtvik G (2002) 'The SVERIGE spatial microsimulation model: content, validation and example applications', Technical report, CERUM Kulturgeografi, Kulturgeografiska Institutionen/SMC, Umeå University.

Holm E, Lindgren U, Lundevaller E and Strömgren M (2006) 'The SVERIGE spatial micro simulation model', Paper presented at the 8th Nordic Seminar on Microsimulation Models, Oslo, 8-9 June 2006. (http://www.ssb.no/misi/papers/holm_einar_svesim.doc) [accessed 1 September 2007]

Holmer M, Janney A and Cohen B (2007). 'PENSIM Overview', Policy Simulation Group, Washington D.C.. (http://www.polsim.com/overview.pdf#search='overview%20of%20PENSIM) [accessed 1 September 2007]

Immervoll H and O'Donoghue C (2001) 'Towards a Multi-Purpose Framework for Tax-Benefit Microsimulation', EUROMOD Working Paper EM2/01, Institute for Social and Economic Research, University of Essex. (http://www.iser.essex.ac.uk/msu/emod/workingpapers/em201_cov.pdf) [accessed 1 September 2007].

Keefe J, Legare J and Carriere Y (2005) 'Developing new strategies to support future caregivers of the aged in Canada: projections of need and their policy implications', Paper presented at the XXV IUSSP International Population conference, Tours, 18-23 July. (http://iussp2005.princeton.edu/download.aspx?submissionId=52430) [accessed 1 September 2007]

Kelly S (2007a) 'DYNAMOD' in Gupta A and A Harding (eds.) *Modelling our Future: Population Ageing, Health and Aged Care*, International Symposia in Economics, Volume 16, North-Holland, Amsterdam, 439-442

Kelly S (2007b) 'APPSIM – Objectives and Requirements', NATSEM Working Paper 1, National Centre for Social and Economic Modelling, University of Canberra. (https://guard.canberra.edu.au/natsem/index.php?mode=download&file_id=741) [accessed 1 September 2007]

King A, Baekgaard H and Robinson M (1999) 'DYNAMOD-2: An Overview', NATSEM Technical Paper 19, National Centre for Social and Economic Modelling, University of Canberra. (https://guard.canberra.edu.au/natsem/index.php?mode=download&file_id=560) [accessed 1 September 2007]

Klevmarken N A (1994) 'The impact of demographic changes on the income distribution: Experiments in microsimulation', Paper presented at the 8th annual conference of the European Society for Population Economics, Tilburg, Holland, June 2-4 1994.

Klevmarken N A (1997) 'Behavioral Modeling in Micro Simulation Models. A Survey.', Working Paper 1997:31, Department of Economics, Uppsala University.

Klevmarken N A (2001) 'Microsimulation - a tool for economic analysis', Working Paper 2001:13, Department of Economics, Uppsala University.

Klevmarken N A and Olovsson P (1996) 'Direct and behavioral effects of income tax changes – Simulations with the Swedish model MICROHUS', in Harding A (Ed.) *Microsimulation and Public Policy,* Contributions to Economic

Analysis 232, North-Holland, Amsterdam, 203-229.

Knickman J R and Snell E K (2002) 'The 2030 Problem: Caring for Aging Baby Boomers', *Health Services Research*, 37(4), 849-884.

Maroun J, Ng E, Berthelot J-M, Le Petit C *et al.* (2003) 'Lifetime costs of colon and rectal cancer management in Canada', *Chronic Disease in Canada*, 24(4), 91-101.

Merz J (1991) 'Microsimulation - a survey of principles, developments and applications', *International Journal of Forecasting*, 7(1), 77-104.

Dussault B (2000) 'Overview of DYNACAN - a full-fledged Canadian actuarial stochastic model designed for the fiscal and policy analysis of social security schemes', Report, International Actuaries Association, Ontario. (http://www.actuaries.org/CTTEES_SOCSEC/Documents/dynacan.pdf) [accessed 1 September 2007]

Mot E S (1992) 'Survey of Microsimulation Models', Social Security Research Committee, VUGA: The Hague.

O'Donoghue C (2001) 'Redistribution in the Irish Tax-Benefit System', Unpublished PhD thesis, London School of Economics.

Orcutt G (1957) 'A new type of socio-economic system', *Review of Economics and Statistics*, 58, 773-797.

Robert-Bobee I (2001) 'Modelling demographic behaviours in the French microsimulation model Destinie: An analysis of future change in completed fertility', INSEE N° G2001/14.

Spielauer M (2003) 'A dynamic microsimulation model for Austria: general framework and application for educational projections', Unpublished Doctoral Thesis, University of Vienna.

Statistics Canada (2002) 'The LifePaths Microsimulation Model - An Overview', Statistics Canada, Ottawa. (http://www.statcan.ca/english/spsd/LifePathsOverview_E.pdf) [accessed 1 September 2007]

Statistics Canada (2007) 'Modgen Developer's guide', Statistics Canada, Ottawa. (http://www.statcan.ca/english/spsd/ModgenDev_EN.pdf) [accessed 1 September 2007]

Strategic Forecasting (2002) 'CORSIM: Analyst Documentation', Strategic Forecasting, Cornell. (http://www.strategicforecasting.com/docs/corsim_analyst.pdf) [Accessed 1 September 2007]

Sutherland H (ed) (2001) 'EUROMOD: An Integrated European Benefit-Tax Model, Final Report', EUROMOD Working Paper No. EM9/01, Institute for Social and Economic Research, University of Essex. (http://www.iser.essex.ac.uk/msu/emod/workingpapers/em901_cov.pdf) [accessed 1 September 2007].

Toder E, Uccello C, O'Hare J, Favreault M *et al.* (1999) 'Modeling Income in the Near Term – Projections of Retirement Income Through

2020 for the 1931-60 Birth Cohorts. Final Report.', Urban Institute, Washington DC. (http://www.urban.org/UploadedPDF/MINT_full.pdf) [accessed 1 September 2007]

Van Imhoff E and Post W (1998) 'Microsimulation methods for population projection', *Population: An English Selection*, 10(1), 97-138.

Vencatasawmy C (2002) 'Modelling fertility in a life course context: some issues', Working Paper 16, Austrian Institute for Family Studies, Universität Wien. (http://62.116.39.195/ftp/projekte/famsim/wp_16_famsim_fertility.pdf) [accessed 1 September 2007]

Wachter K W (1998) 'Kinship Resources for the Elderly: An Update', Working Paper, Dept. Of Demography, University of California. Berkeley. (http://www.demog.berkeley.edu/~wachter/WorkingPapers/terrace.pdf) [accessed 1 September 2007]

Wachter K W (1995), 2030's Seniors: Kin and Step-Kin', Working Paper, Dept. Of Demography, University of California, Berkeley. (http://www.demog.berkeley.edu/~wachter/WorkingPapers/kinpaper.pdf) [accessed 1 September]

Wachter K, Blackwell D and Hammel E A (1998) 'Testing the Validity of Kinship Microsimulation: An Update', Working Paper, University of California. (http://www.demog.berkeley.edu/~wachter/WorkingPapers/valpaper.pdf) [accessed 1 September 2007]

Will B P, Le Petit C, Berthelot J-M, Tomiak E M *et al.* (1999) 'Diagnostic and therapeutic approaches for nonmetastatic breast cancer in Canada, and their associated costs', *British Journal of Cancer*, 79(9-10), 1428-1436.

Will B P, Berthelot J-M, Le Petit C, Tomiak E et al. (2000) 'Estimates of the lifetime costs of breast cancer treatment in Canada', *European Journal of Cancer*, 36, 724-35.

Will B P, Berthelot J-M, Nobrega K M, Flanagan W and Evans W K (2001) 'Canada's Population Health Model (POHEM): A tool for performing economic evaluations of cancer control interventions', *European Journal of Cancer*, 37, 1797-1804.

Wolfson M and Rowe G (2004) 'Disability and informal support: Prospects for Canada', in Cohen S B and J M Lepkowski (Eds.) [Proceedings of the] *Eighth Conference on Health Survey Research Methods*, Hyattsville, MD: National Center for Health Statistics, 15-22.

Zaidi A and Rake K (2001) 'Dynamic Microsimulation Models: A Review and Some Lessons for SAGE', SAGE Discussion Paper 02, ESRC SAGE Research Group, London School of Economics. (http://www.lse.ac.uk/collections/SAGE/pdf/SAGE_DP2.pdf) [accessed 1 September 2007]

APPENDIX: DYNAMIC MICROSIMULATION MODELS

Table A.1 Summary model descriptions

Model	Institution / Country	Description	Data	Time period	Population	Link
APPSIM	Australia – NATSEM	APPSIM (Australian Population and Policy Simulation Model) is the second generation dynamic simulation model built at NATSEM It aims at providing policymakers with the right toolbox to assess and prepare for the social and fiscal challenges ahead that are associated with the aging of the Australian population. (Cassells *et al.*, 2006) The APPSIM improvements over previous models are the easy replacement of modules, the fully integrated alignment and aggregate calibration from the start for every module, the ability to change the order of simulation, the flexibility in output, and the clear focus on user needs and requirements. (Kelly, 2007a,b)	HILDA	discrete	closed	http://www.canberra.edu.au/centres/natsem
DYNAMOD	Australia - NATSEM	The DYNAMOD model, a dynamic microsimulation model of the Australian population, is designed to project characteristics of the population over a period of up to 50 years. (King *et al.*, 1999) Major elements of the model include demographics, international migration, education, the labour market and earnings. DYNAMOD uses a "pseudo-continuous" time framework operating in monthly steps for most demographic and labour market processes and in annual steps for education and earnings. With regard to the statistical modelling approaches used, it makes maximum use of survival functions.	n=150,000 sample of census	discrete (months)	closed	http://www.canberra.edu.au/centres/natsem
HealthMod	Australia - NATSEM	This is a model currently being developed that simulates the use and costs of medical and related services by Australian families. The objective is to be able to analyse distributional impacts of potential policy changes as well as the revenue and expenditure implications for government. The unit of analysis is the family. The data is based on the Australian Bureau of Statistics (ABS) 2001 National Health Survey (NHS01). This dataset needed to be enriched by statistical matching techniques in order to additionally be able to model at the family level rather than solely at the individual level.	2001 National Health Survey	discrete	closed	http://www.canberra.edu.au/centres/natsem
FAMSIM+	Austria	FAMSIM+ is a dynamic microsimulation modelling platform for Austria developed alongside a PhD thesis (Spielauer, 2003). It contains basic demographics like fertility and partner matching by education, as well as a more detailed educational module for the study of the intergenerational transmission of education in Austria. FAMSIM+ is an example of a simple and focused model. Technically it follows an object oriented programming approach. Its simplicity together with the programming approach made it also a teaching tool.	1996 Micro Census special program on education and fertility	discrete (months)	closed	

DYNACAN	Canada - Canada Pension Plan	DYNACAN is a Canadian stochastic dynamic microsimulation model for fiscal and policy-oriented analysis of the Social Security System in Canada and mainly focusing on the Canada Pension Plan (CPP). The model is made up of three main components (Morrison 1998/ Dussault 2000): DYNACAN A assembles the initial representative data base from various data sources; DYNACAN B simulates the demographics and employment earnings over time. This module also includes the modelling of disability and recovering from disability; DYNACAN C calculates the CPP contributions and benefits.	n=212,000 sample from 1971 Census	discrete	closed	
LifePaths	Canada - Statistics Canada	LifePaths is an integrated "general purpose" model, including demographic behaviours, health, education, labour and allowing for tax-benefit and pension modelling. It operates in continuous time which (amongst other things) allows for a more accurate representation of causation and behaviour. LifePaths uses a synthetic initial database based on a variety of historical micro-data sources in order to create representative synthetic life histories from birth to death for all birth cohorts since 1872. The model is programmed in the generic microsimulation language Modgen.	synthetic population (sample size set by user)	continuous	open	http://dissemination.statcan.ca/english/spsd/LifePaths.htm
POHEM	Canada - Statistics Canada	POHEM (Population Health Model) started as a sister model of LifePaths and over time developed to a family of different model variants for different health-related applications. POHEM is a dedicated health model. Typical output includes measures of cost effectiveness, cost per year of life gained, cost per health-adjusted year of life gained and measures on the impact of interventions on incidence and survival. In addition, disease-specific survival curves and costs of care for co-morbidities can be established. The model is programmed in the generic microsimulation language Modgen.	(1) synthetic population (sample size set by user) or (2) Canadian Community Health Survey	continuous	open	http://dissemination.statcan.ca/english/spsd/Pohem.htm
DESTINIE	France - INSEE	DESTINIE is an analytical tool for pension policy which produces pension projections to the year 2040. The demographic transition probabilities are based on the "school living age", which is a major covariate of the model, but the labour market transition probabilities are based on age, gender and education levels. Wage trajectories of individuals from the same cohort, gender and educational attainment are represented using autoregressive random components to account for additional dispersion (Blanchet and Crenner 2006).	n=50,000 Financial Assets Survey 1991, 1997, 2003	discrete	closed	http://www.insee.fr
ANAC, DYNAMITE	Italy - Bank of Italy	The ANAC-model was developed at the University of Pennsylvania by Alberto Ando and Sergei Nicoletti Altimari to examine the Italian savings behaviour. DYNAMITE was developed by the Bank of Italy (Banca d'Italia) based on ANAC. The aim of this model is to examine microeconomic questions at the household level and the impact of macroeconomic factors and institutional changes on distributional issues. The program contains three special modules. A labour market model that utilizes insight from search theory, a tax calculation module, and a retirement decision module that determines the retirement decision endogenously.	n=24,000 1993 Survey on Household Income and Wealth	discrete	closed	

MIND	Italy - University of Pisa	MIND (Micro Italy National Dynamics) simulates the demography and the socio-economic structure of Italy, by taking into account the three different macro-regions North, Centre and South. Model development started in 1998 at the University of Pisa. The current version of the model has been used to evaluate the long term effects of the Italian national pension system and the short term effects of the Italian personal tax system. Economic equality is not studied cross-sectional only by estimating the distribution of disposable income in a given year, but also for prolonged periods to cover several years, or the entire lifetime of different generations.	1995 starting population constructed from various surveys	discrete / closed	http://www.unipr.it/arpa/defi/microsim/home.html
MIDAS	New Zealand	The Microsimulation of Income Dynamics and Accumulation of Savings (MIDAS) was created to examine wealth accumulation and distribution in New Zealand. The project was commissioned by the Department of the Prime Minister and Cabinet on behalf of the Task Force on Private Provision for Retirement. Accordingly, its focus was on estimating the proportion of people who could provide for their own retirement by the age of 65.	n=10,000; sample from census 1991	discrete / closed	http://www.retirement.org.nz/midas_model.html
MOSART	Norway - Statistics Norway	MOSART is a dynamic microsimulation model for Norway developed at Statistics Norway to investigate policy options with regard to financing public expenditure. In its first version developed between 1988 and 1990 it focused on demographic behaviour, education and labour force participation in order to study the impact of demographic change on labour force and education attainment (Fredriksen 1998). The second version (1991-93) extends the model, allowing for pension modeling. Currently MOSART exists in its third version (since 1996) that includes more detailed modules with respect to household formation, income, taxation, savings, wealth (Fredriksen 2003) and disability.	Register data representing 12% of the population	discrete / closed	http://www.ssb.no/emner/02/03/sos101
MicroPox	Sweden - individual PhD thesis by Brouwers (2005)	The probabilistic dynamic microsimulation model MicroPox is based on the SVERIGE model. It is spatially explicit on 100x100m square grid. The data set for the model is extracted from anonymized administrative registers of the whole Swedish population from 1998 which includes information on family relations and workplaces containing location. The model is run with hourly changes. Places and persons are the fundamental entries. The distinguishing features from other epidemiological models are the inclusion of contact patterns and detailed spatial information. The purpose of this transmission model is to analyze and compare different intervention possibilities to prevent a spread of the infectious disease smallpox (Brouwers 2005).	1998 register data of the Swedish population	discrete (hours) / closed	
SESIM	Sweden - Ministry of Finance	SESIM is a dynamic and stochastic microsimulation model developed by the Swedish Ministry of Finance. The first version was developed in 1997 as a tool to assess the Swedish education financing system (Ericson and Hussénius, 2000). The second version (Flood 2004) has been developed since the year 2000 to "evaluate the financial sustainability of the new Swedish pension system" . SESIM III is the current version: the major purpose is still pensions, but "the new version extends the analyses of health issues amongst elderly" (Flood *et al.*, 2005).	n=100,000 1999 year wave of LINDA.	discrete / closed	http://www.sesim.org

SVERIGE, LISA, SVESIM	Sweden - SMC	SVERIGE is the first of a family of models developed by the Spatial Modelling Centre in Kiruna (Sweden) and is based on the CORSIM model extended by a spatial dimension locating individuals on a 100m2 grid. Different to most other models, it is not based on a sample but simulates the entire Swedish population based on detailed longitudinal administrative data. It is mainly used to evaluate human ecodynamics (Holm *et al.*, 2002). While health issues are not modelled in SVERIGE, the model is the basis for the development of an epidemiological model MicroPox (Brouwers, 2005). LISA is a later simplified version of the original model with updated equations and an incorporation of individual income sources targeting labour supply issues. It only looks at individuals, not at families, and locates migrants on a larger scale of labour market regions (Holm *et al.*, 2006). The most recent development is the model SVESIM which has the detail of the original model and further includes labour demand and individual labour market clearing (Holm *et al.,* 2006).	n=9,000,000 administrative data 1985-95	discrete	closed	http://www3.umu.se/ soc_econ_geography/ smc/
MICROHUS	Sweden - Uppsala University	MICROHUS is a dynamic micro simulation model for the Swedish household sector including tax and transfer system and people's behavioural "adjustments to policy changes" (Klevmarken, 2001:20). The simulation model contains the following model "blocks": demographics, geographical mobility, housing, labour market, wealth, income and taxes, childcare, price levels. MICROHUS has been used in an evaluation of the 1990/91 tax reform in Sweden (Klevmarken and Olovsson, 1996). Furthermore the impact of demographic changes on the income distribution has been studied with this microsimulation model (Klevmarken, 1994).	Swedish household sample HUS 1984	continuous	closed	
SAGEMOD	UK - London School of Economics, King's College, University of Southampton	The Simulating Social Policy in an Ageing Society (SAGE) project, funded by the Economic and Social Research Council (ESRC) started in 1999. Its main goal is the assessment of ageing population's impact on future pension, health and long-term care policies. Along with the series of static simulation models for policies' impact evaluation, SAGE has developed a dynamic microsimulation model (SAGEMOD) for projecting the British population until 2031. Incorporating census, panel and cross-sectional survey data, it simulates the following events: birth, death, education, marriage, divorce, labour force participation, earnings, health, retirement, disability and informal care.	n=54,000 0.1% of 1991 Census	discrete	closed	http://www.lse.ac.uk/c ollections/SAGE/resear ch/modellingForPolicy/ SAGEMOD/Default.htm
NCCSU	UK - University of Leicester ; London School of Economics	The NCCSU Micro-simulation Model for Long-Term Care Charging simulates the incomes and assets of future cohorts of older people and their ability to contribute towards care costs. It was developed Nuffield Community Care Studies Unit (NCCSU) at the University of Leicester and serves as the microsimulation component in a long-term-care micro-macro model. The macro component is the PSSRU Long-Term Care Financing Model that aims to make projections of three key variables: the expected number of older people (aged 65 and over) with dependency, their likely level of demand for long-term care services and the costs associated with meeting this demand.	British Family Resources Survey (FRS)	discrete	closed	http://www2.le.ac.uk/ departments/health-sciences http://www.pssru.ac.uk

CBOLT	US - Congressional Budget Office	The Congressional Budget Office Long-Term (CBOLT) policy simulation model is an integrated micro/macro policy simulation model. It was developed to answer budgetary and distributional questions about Social Security, Medicare, and other long-term policy issues. The model uses dynamic microsimulation within an aggregate budgetary and macroeconomic framework to generate long-run projections and policy analysis for aging-related programs like Social Security and Medicare. CBOLT is used to analyze how sensitive theses programs are to demographic, economic, and behavioural assumptions. The model is designed to examine the effects of various federal budget policy options, such as alternative Social Security tax and benefit rules, Medicare eligibility and cost growth assumptions but also to simulate the effects of various fundamental policy changes such as private accounts, trust fund investment.	Continuous Wage History Survey (SSA) and Center for Medicare and Medicaid Studies (CMS)	discrete closed	http://www.cbolt.com
CORSIM	US - Cornell University; Strategic Forecasting	The Cornell Dynamic Population Microsimulation Model (CORSIM) started its development at Cornell University (U.S) in 1987. The fourth and most recent version of it was released in September 2002. This model has been built for research into fundamental socioeconomic process and policy analysis. CORSIM included information on both the individual and family levels and has been applied to investigate changes occurring in family networks, wealth accumulation, patterns of intergenerational mobility, and to assess policy reforms of several US administrations. CORSIM has served as template for other models, e.g. DYNACAN built in 1995 (CANADA) and SVERIGE in 1997 (SWEDEN).	n=180,000 US Census Public Use Micro data Sample (PUMS)	discrete closed	http://www.strategicforecasting.com/
PENSIM/2	US - Department of Labour	PENSIM/2 is geared toward the analysis of employer-sponsored pension plans (Holmer et al, 2007) It is a dynamic simulation model that simulates life histories for a sample of individuals born in a particular year starting from 1935. PENSIM uses waiting time survival models to model the events in life histories. PENSIM is a very rich simulation, it simulates events ranging from timing of marriage, schooling events, child birth for women, migration, disability, jobs characteristics and to pensions plans and coverage.	Survey of Income and Program Participation (SIPP), Panel Study of Income Dynamics (PSID)	continuous open	
PRISM	US - ICF/Lewin Inc.	The current version of the Projected Retirement Income Simulation Model (PRISM) is the second major revision of the model that was first developed together by Lewin-ICF and the Brookings Institution in 1986 (Knickman and Snell, 2002). PRISM is now also part of a Long-term Care Model. This model simulates the utilization and financing of long-term care services for elderly individuals through 2050. PRISM simulates for each person future demographic characteristics, labour force participation, income and assets of the elderly. The Long-term Care Financing Model simulates disability, utilization of long term care, and methods of financing long-term care.	n=28,000, various sources	discrete closed	
MINT	US - Social Security Administration	The microsimulation model MINT has been developed by the US Social Security Administration. The model was mainly developed for the projection of retirement income for current and future benefits of an individual or a couple of the birth cohorts 1931 - 1960 (Toder *et al.*, 1999). The model makes independent projections for each retiree's income from Social Security benefits, pensions, assets, and earnings (Butrica *et al.*, 2001:3)	n=113.500 1990- 1993 Survey of Income and Program Participation (SIPP)	discrete closed	

| DYNASIM | US - Urban Institute | The DYNASIM "Dynamic Simulation of Income Model" was the first large-scale dynamic microsimulation model in social sciences. It was initially developed between 1969 and 1976 under the direction of Guy Orcutt at the Urban Institute (Orcutt, 1957). A second version - DYNASIM2 – was developed between 1979 and 1983. A third version – DYNASIM3 – was completed in 2004 and is based on 1990-1993 Survey of Income and Program Participation panels. Like in DYNASIM2, the main focus lies on pension simulations and issues of population aging. | DYNASIM3: n=100.000 1990-1993 Survey of Income and Program Participation (SIPP) | discrete | closed | http://www.urban.org/ |