Session Number: 4C
Session Title: Improving Estimates from Survey Data
Session Organizer: Stephen Jenkins, University of Essex, Colchester, UK
Session Chair: Stephen Jenkins

Paper Prepared for the 29th General Conference of
The International Association for Research in Income and Wealth

Joensuu, Finland, August 20 – 26, 2006

Enhancing the Australian National Health Survey Data for Use in a Microsimulation Model

Annie Abello, Sharyn Lymer, Ben Phillips, Laurie Brown and Ann Harding

For additional information please contact:

Annie Abello
NATSEM, University of Canberra ACT 2601, Australia
Email annie.abello@natsem.canberra.edu.au
Fax +61 2 6201 2751
Telephone +61 2 6201 2794

This paper is posted on the following websites: http://www.iariw.org
http://www.natsem.canberra.edu.au

Contents

The construction of the current version of MediSim was supported by an Australian Research Council linkage grant (LP0219571), and by Medicines Australia, the Industry Partner to this grant. The authors would like to thank Brendan Shaw, Senior Manager, Policy and Research, Medicines Australia and members of the Pharmaceutical Economic Taskforce of Medicines Australia for their advice and input on the modeling.

1 Introduction

An important contributor to the well-being of low income families in Australia is their access to subsidised medicines through the Australian Pharmaceutical Benefits Scheme (PBS). The PBS provides Australians with affordable access to necessary and cost-effective prescription (ethical) medicines. Federal government expenditure on the PBS in 2004-05 was $Aus 5.4 billion, or 15 per cent of total Government health expenditure. There is considerable debate over the sustainability of the PBS, with the scheme historically being one of the fastest growing sectors within health.

In recent years the National Centre for Social and Economic Modelling (NATSEM) at the University of Canberra has developed a series of microsimulation models of the PBS — MediSim. MediSim was constructed to estimate current and future use and costs of PBS medicines under different policy settings, and to quantify the distributional effects of policy changes.

The first version of MediSim was completed in 1998. Since then, the model has been updated and developed in successive stages. In 2003-04, work was undertaken to update the model and improve its capability. This principally involved adding health conditions into the model's base dataset. Work to develop a facility within the model to evaluate not only the costs but also the benefits from the use of medicines was also initiated.

This paper discusses the development of MediSim's new base dataset using the Australian 2001 National Health Survey (NHS). An overview of the Australian PBS and MediSim is provided in the next section, followed by a discussion of issues considered in choosing the base data. Subsequent sections describe the various methodologies developed to modify the 2001 NHS for use as the base file for MediSim.

OVERVIEW OF THE PBS AND MEDISIM

The Australian Commonwealth Government's Pharmaceutical Benefits Scheme (PBS) aims to provide Australians with timely, reliable and affordable access to necessary and cost-effective prescription medicines.

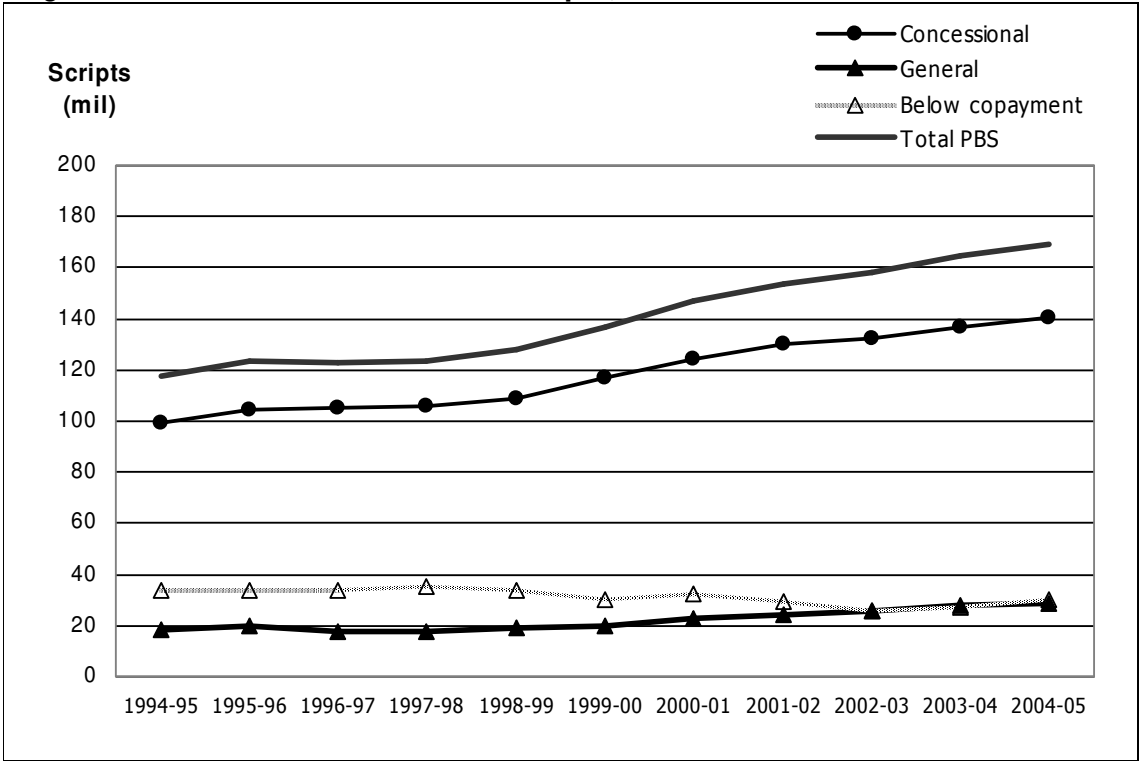
Patients are required to make a contribution to the cost of prescribed medicines listed on the PBS. Individuals and families eligible for certain federal government pensions and allowances (e.g. age pension, unemployment benefit, disability pension) are able to access PBS medicines at concessional rates. These persons are known as concession cardholders. The PBS also has 'safety net' arrangements to protect individuals and families from large overall expenses for PBS-listed medicines. The levels of patient copayments and the PBS safety net arrangements are referred to as the PBS policy settings. Patient copayments and safety net thresholds (SNTs) are revised annually in line with movements in the consumer price index (CPI) from 1 January each year.

The majority of prescribed drug sales in Australia are covered by the scheme and, on average, the government subsidises patients to the extent of 84 per cent of PBS drug costs. Currently nearly 80 per cent of total government subsidies through the PBS accrue to

concessional patients — that is, those with the specified Centrelink cards[1] and 20 per cent to general patients.

Finding ways of curbing government expenditure on the PBS, while maintaining social equity and access to 'essential' medicines, is a continual concern within federal health and financial public policy circles. Since the early 1990s government expenditure on the PBS has grown at more than 10 per cent a year – well above the growth in the health budget (6 per cent) or the economy (4 per cent in terms of gross domestic product). However, over the next couple of years, rates of growth in PBS expenditure are likely to be below the long run average growth of the PBS. The Australian Commonwealth Treasury is forecasting PBS expenditure to grow over the coming three years at an average of 5.4 percent per annum in real terms, compared to 3.3 percent for the total health budget (Commonwealth Treasury, 2005). Government's share of the costs of the PBS has steadily increased over time as PBS settings – patient copayments and safety net thresholds – which largely determine patient contributions, have increased in general only in line with inflation. In 2004-05, total PBS scripts reached 169 million and cost the government $5.4 billion. PBS-listed scripts priced below copayment reached 30 million.

Figure 1 **Volume of PBS-listed scripts, 1994-95 to 2004-05**



Data sources: Medicare Australia website; Drug Utilisation Sub-Committee (DUSC) Drug Utilisation Database, Pharmaceutical Benefits Branch, Commonwealth Department of Health and Ageing.

NATSEM models the Australian PBS using the microsimulation model MediSim. MediSim simulates the current and future use and costs of PBS medicines under existing and different policy settings (see, for example, Harding *et al* 2004, Brown *et al* 2005a and

[1] These are the Pensioner Concession Card, the Commonwealth Seniors Health Card and the Health Care Card.

Brown *et al* 2005b). It also estimates the distributional effects of policy changes on Australian families. By altering the medicines included in the model, their assigned prices and script volumes, MediSim is capable of simulating the impact of: inclusion of new drugs on to the PBS list; restriction on the drugs listed on the scheme or on the pricing of drugs; increased restrictions on drugs by indication; increased use of generics at more competitive prices; or an increased emphasis on the quality use of medicines as reflected in changes in doctor prescribing behaviour; as well as changes to copayment and safety net arrangements.

The model could be used to provide answers to relatively simple issues such as the impact of expected changes in PBS subsidised drug prices and scripts over the next 5 to 10 years on government PBS outlays, or patient out-of-pocket expenditures and related revenues to industry. It can also be used to assess more complex matters such as the likely impact of, for example, the introduction of new PBS listed drugs, the effects of demographic and socio-economic changes upon outlays, or the distributional and revenue impacts of changing the rules of the PBS (such as the introduction of differential copayment levels as operate in many European countries).

CHOICE OF BASE DATA

In 2003, NATSEM obtained joint industry and national competitive research funding to extend the capability of the model to include health outcomes and the evaluation not only of the costs but also the benefits arising from the use of medicines. A necessary step to modeling both the *costs* and *benefits* of these medicines is the inclusion of diseases and health conditions in the model's base file.

It was thought initially that the best way to add diseases into the model's dataset was to replace the model's original base file, derived from NATSEM's HES-based STINMOD/01a[2] (which in this paper we interchangeably refer to as "HES"), with either the 2001 or 1995 National Health Survey (NHS). The 1995 NHS is the best source of complete (although not up-to-date) information on illnesses and medicine usage. The 2001 NHS, on the other hand, contains the latest person-level information in Australia on long-term health conditions, drug usage for national health priority conditions[3], and health risk factors. However, the latter has a number of limitations as the base data for a microsimulation model following changes in survey design relative to previous surveys.

The 2001 NHS contains only limited information on prescribed medicine use (e.g. only usage related to priority diseases has been collected), and short-term health conditions (apart from the priority areas, the survey collected specific information on the health condition only of persons with long term conditions). Data on drug usage and the health condition/s for which the medicines are being taken, are essential in a model concerning

[2] STINMOD/01a is NATSEM's static microsimulation model of the Australian tax and transfer system, based on the 1998-99 Household Expenditure Survey (HES).

[3] The Australian health system has identified 7 national health priority areas (NHPAs), including arthritis and musculoskeletal conditions, asthma, cancer control, cardiovascular health, diabetes mellitus, injury prevention and control, and mental health. The diseases and conditions targeted under the NHPA initiative were chosen because they currently impose high social and financial costs on Australian society, and through appropriate and focused attention significant gains in the health of Australia's population can be achieved.

usage of prescribed medicines. While the national priority areas, that are included in the survey, constitute a large share of the PBS (55% of scripts and 64% of costs), the residual that constitutes non-priority areas is still large and should be included in the model.

Another major shortcoming of the 2001 NHS in regard to modeling the PBS concerns the coverage of the survey. Unlike previous national health surveys, the 2001 survey did not obtain information on all persons in the household, so the ABS recommends that analysis of the 2001 file be done at the person level only (ABS, 2003). As the PBS safety net operates at the income unit[4] level, complete information on all family or income unit members (particularly on drug usage) is needed to adequately model this critical component of the PBS system.

Given the lack of currency of the 1995 health survey, the best option was to build the new model upon a dataset based on the 2001 NHS but modified such that there was complete information on each member of a family, and with additional information on short-term health conditions and drug usage. This was achieved by statistically matching the 2001 NHS to a person-level dataset derived from NATSEM's STINMOD/01a[5], and augmenting the resulting base file with information from the 1995 NHS on short-term health conditions and drug usage.

The next sections describe the methodology used for statistical matching (Section 2), imputation of short-term conditions (Section 3) and imputation of prescribed drug usage (Section 4). Section 5 concludes.

2 Statistical Matching

Statistical matching is a procedure used to link two files or datasets where each record from one of the files is matched with a record from the second file that generally does not represent the same unit, but does represent a *similar* unit. It is a method to bring together microdata that are not available from a single data source[6].Since the records to be matched in this exercise involve sample surveys (rather than administrative data or a census), and considering the incomplete coverage of families in the 2001 NHS, the matching of records involves finding the closest *statistical match* rather than actual matching of data records of the same persons.

[4] The Australian Bureau of Statistics defines an income unit as 'one person or a group of related persons within a household, whose command over income is assumed to be shared. Income sharing is assumed to take place within married (registered or de facto) couples, and between parents and dependent children' (ABS 2001).

[5] NATSEM was given approval by the ABS to conduct the statistical matching, and as part of the approval process, a joint ABS-NATSEM Technical Working Group on Statistical Matching was formed.

[6] Rodgers (1984) gives a detailed description of statistical matching, as do Rassler (2002), Radner et al (1980), Cohen (1991), Sutherland *et al* (2002) and Moriarty and Scheuren (2001).

THE DATA TO BE MATCHED

The HES contains around 18,000 person records and contains detailed income and expenditure information. Each record has a unique identifier at the household, family, income unit and person level. This hierarchy allows identification of persons to their correct income unit, family or household. Having this detail allows the proper modeling of the PBS safety net as expenditures on PBS drugs can be summed for each income unit.

The NHS contains around 27,000 person records. The NHS records have detailed information relating to the health of each person. To link records in both data sets, we need variables that are common to both data sets and strongly relate to the modeling area, in this case health. The variables that are common to the two sets of microdata to be matched are called the "matching" variables.

ISSUES

Differences in sample size among the different surveys should not be an issue as weighted duplications of records may be created so that the sample sizes in both datasets are equal (Taylor, Gomulka and Sutherland, 2000). The main difficulty likely to be encountered is the choice of weights in the merged file. The weights in either the NHS or HES could be selected — but the risk is that the statistical distribution and parameters of the variables merged from the other survey are not maintained. If *constrained* statistical matching is used, then the goal is to try to maintain the marginal distributions of the non-matched variables by minimising the difference in the weights of the records in the merged file to those of the two original surveys (Cohen, 1991).

Prior to the actual matching, modifications to each dataset were made to make the matching variables from each dataset consistent. In the HES, we created separate data records for children, imputed concession card status (used to identify individuals and families who are eligible to access medicines at concessional rates) and estimated equivalent income deciles consistent with the 2001 NHS definition. In the NHS, we imputed a value for equivalent income decile when the response was 'not stated', and used self-assessed health status as a proxy indicator for income unit expenditure on prescriptions.

The details regarding modifications made to the data, details of the statistical matching procedure, the conditional independence assumption, definitions and differences between constrained and unconstrained matching, among other issues, are described in Technical Working Group ABS-NATSEM Collaboration on Statistical Matching (2004).

STATISTICAL MATCHING METHODOLOGY

A two-step approach is used to statistically match the NHS and HES. First, person records are grouped, mainly by their income unit characteristics, into homogeneous cells to determine the most similar records. Next, persons belonging to the same group are matched together using a distance function. The cell groups are formed to ensure a certain standard is always maintained for the statistical match. When using a small number of cell groups, the accuracy of matches can be improved by a properly formulated distance function. A similar approach of first dividing the datasets into groups of similar households, before doing any actual matching, is discussed in Sutherland *et al* (2002).

Homogeneous groups The variables used to divide person records in homogeneous groups include age (6 groups), gender (2 groups), income unit expenditure on prescriptions (4

groups initially, then collapsed to 2 groups), income unit type (4 groups initially, then collapsed to 2 groups), and concession card status (2 groups). Using these five variables resulted in 384 cell groups. It was not until income unit type and expenditure on prescriptions had been each collapsed into two groups, that it was possible to obtain populated cell groups for all combinations (96 cell groups). There were other common variables available (labour force status, number of usual residents in the household and equivalent income unit decile) but these were not used as the more grouping variables used, the less likely cells will be filled (i.*e* not empty)

Distance function The distance function is a mathematical equation that attempts to more closely match individuals from the two surveys who fall within the same cell group. It is of the Mahalanobis form, defined as

$$d_{ij}^{M} = \sqrt{\left(x_i^A - x_j^B\right)'\left(x_i^A - x_j^B\right)S_x^{-1}}$$

where S_x is the estimated covariance matrix for the X variables. This was calculated using user-defined weights corresponding to the relative importance given to each matching variable. The X variables include age, number of usual residents in the household and equivalent income unit decile. The *age* variable in the distance function has 16 possible categories, somewhat more than the 6 categories employed in the cell groups. Two approaches were considered in implementing this stage of the matching procedure: unconstrained and constrained matching.

Unconstrained matching Unconstrained matching is relatively simple and computationally easy. The approach is to match each HES record to the closest matching NHS record *with replacement*. Under this approach it is possible for the same NHS record to be matched with multiple HES records. Selecting NHS records for matching in this manner ensures that the "match quality" is high. Table 1 shows the accuracy of the match with respect to the variable age. Age was a cell group variable and this ensured a certain level of accuracy[7]. The distance function tightens the match very successfully for the unconstrained method. Consider the age group 40 to 44 in the matched dataset. 83 per cent of the records have been matched to NHS records with the correct age group.

The downside is that it can be difficult to match many of the records so, ultimately, the distribution of the non-matching variables in the matched file can be very different from their distribution in the original dataset; this "can have a deleterious effect on the validity of the results of analyzing the matched file" (Cohen 1991 p. 65). The problems with unconstrained matching become more apparent when working with data at a finer level of disaggregation.

[7] The results are based on unconstrained matching where the distance function used weights of 0.5 for age, 0.25 for equivalised income unit decile and 0.25 for the number of usual residents in the household.

Table 1 **Unconstrained matching age allocation**

Matched dataset	NHS age group															
	0 - 4	5 - 9	10 - 14	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59	60 - 64	65 - 69	70 - 74	75 plus
0 - 4	**100**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
5 - 9	.	**90**	10	.	.	.	.	.	.	.	.	.	.	.	.	.
10 - 14	.	16	**84**	.	.	.	.	.	.	.	.	.	.	.	.	.
15 - 19	.	.	.	**88**	12	.	.	.	.	.	.	.	.	.	.	.
20 - 24	.	.	.	38	**62**	.	.	.	.	.	.	.	.	.	.	.
25 - 29	.	.	.	.	.	**79**	18	4	.	.	.	.	.	.	.	.
30 - 34	.	.	.	.	.	12	**75**	14	.	.	.	.	.	.	.	.
35 - 39	.	.	.	.	.	3	12	**85**	.	.	.	.	.	.	.	.
40 - 44	.	.	.	.	.	.	.	.	**83**	13	4	0	.	.	.	.
45 - 49	.	.	.	.	.	.	.	.	15	**72**	11	1	0	.	.	.
50 - 54	.	.	.	.	.	.	.	.	3	12	**75**	9	1	.	.	.
55 - 59	.	.	.	.	.	.	.	.	1	3	13	**72**	11	.	.	.
60 - 64	.	.	.	.	.	.	.	.	0	1	3	12	**84**	.	.	.
65 - 69	.	.	.	.	.	.	.	.	.	.	.	.	.	**89**	9	2
70 - 74	.	.	.	.	.	.	.	.	.	.	.	.	.	7	**86**	7
75 plus	.	.	.	.	.	.	.	.	.	.	.	.	.	5	6	**89**

As MediSim is expected to utilize data on long-term health conditions in the NHS, we investigated the distribution of this variable. The NHS shows whether or not individuals have any of 94 possible long-term conditions. Analysis was done on each of these conditions and a comparison was made between the total number of cases for the unconstrained matching-based file and the original NHS data. For each long-term condition, the ratio of the number of conditions in the matched dataset, relative to the original dataset, was computed with a value of 1 implying that the unconstrained matched file provided a perfect representation for a given condition. For all 94 conditions, the ratio using the unconstrained matched file averaged 0.96, but ranged from 0.24 to 1.32. Such a result was not considered adequate and alternative methods needed to be developed.

Constrained matching An alternative to unconstrained matching is constrained matching, also known as the linear programming method following the work of Barr and Turner (1978) and subsequently applied by many other authors. Constrained matching requires the use of *all* records in the two sets of microdata to be matched, and thus, it is able to preserve the marginal distributions of the non-matching variables in each of the two microdata sets. This procedure is adopted to match each HES record to the closest matching NHS record *without replacement.* As the linear programming (LP) method selects records from the NHS without replacement[8], this is expected to reduce the quality of matches compared to the unconstrained matching. A clear advantage with this methodology is that it can guarantee that marginal distributions will remain unchanged.

The linear programming approach requires the sum of HES and NHS weights to be equal. This ensures a "balanced" problem. If a weight is interpreted as the number of people a

[8] For the LP approach each record was "exploded" so that a record was repeated to the extent of its weight. The selection without replacement refers to this "exploded" data set.

record represents then an unbalanced problem leads to either people in the HES or the NHS not being matched. As this method is applied to each of the homogeneous cells, adjusting the weights of either the HES or the NHS to ensure a balanced solution will change the relative importance of each cell. The weights in this particular application were always re-aligned to the NHS population. This means that while marginal distributions will change for the HES variables, this will remain unchanged for the NHS variables.

Table 2 provides a measure of the "closeness" of the match, again with respect to the age variable. The weights that have been attached to the distance function are unchanged from those used to produce the results in Table 1. Table 2 shows that the closeness of the age match is not as robust as that of the unconstrained matching. The results are still promising with very few records in the HES being matched to NHS records where the age categories are more than 1 group apart. A similar comparison was made, looking at the closeness of the matches for income deciles and the number of usual residents in the household (Appendix A). The income results are quite poor with many HES records being matched with NHS records more than 2 categories apart, while the number of usual residents in the household shows a relatively close match.

In the unconstrained matching section the marginal distributions for the 94 long-term conditions in the NHS were discussed. The statistically matched file often over or under-reported the incidence of these conditions. The constrained matching method ensures that the incidence in the matched file is identical to that of the original NHS file.

Table 2 **Constrained matching age allocation**

Matched dataset	**NHS age group**															
	0 - 4	5 – 9	10 - 14	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59	60 - 64	65 - 69	70 - 74	75 plus
0 - 4	**100**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
5 - 9	.	**87**	13	.	.	.	.	.	.	.	.	.	.	.	.	.
10 - 14	.	11	**89**	.	.	.	.	.	.	.	.	.	.	.	.	.
15 - 19	.	.	.	**88**	12	.	.	.	.	.	.	.	.	.	.	.
20 - 24	.	.	.	25	**75**	.	.	.	.	.	.	.	.	.	.	.
25 - 29	.	.	.	.	.	**77**	20	3	.	.	.	.	.	.	.	.
30 - 34	.	.	.	.	.	14	**73**	14	.	.	.	.	.	.	.	.
35 - 39	.	.	.	.	.	1	18	**81**	.	.	.	.	.	.	.	.
40 - 44	.	.	.	.	.	.	.	.	**70**	19	8	3	.	.	.	.
45 - 49	.	.	.	.	.	.	.	.	12	**58**	22	7	1	.	.	.
50 - 54	.	.	.	.	.	.	.	.	4	12	**57**	22	6	.	.	.
55 - 59	.	.	.	.	.	.	.	.	1	3	10	**59**	27	.	.	.
60 - 64	.	.	.	.	.	.	.	.	0	1	2	15	**82**	.	.	.
65 - 69	.	.	.	.	.	.	.	.	.	.	.	.	.	**69**	23	8
70 - 74	.	.	.	.	.	.	.	.	.	.	.	.	.	9	**62**	29
75 plus	.	.	.	.	.	.	.	.	.	.	.	.	.	3	8	**89**

CONCLUSION RE THE STATISTICAL MATCHING

The purpose of the matching was to create a file structure amenable to modeling family PBS expenditure. Family PBS expenditure depends on the age, health and card status of

individuals within a family. The variables that have been used in the cell groups and the distance function attempt to account for these factors.

Two possible matching procedures have been compared: unconstrained matching, where NHS records can be matched to HES records with replacement; and constrained matching where linear programming was used to ensure that marginal distributions of at least the NHS variables remained constant. Both procedures have the ability to match relatively closely on the variables common to the two source microdata. Only the constrained matching can guarantee that marginal distributions will remain unchanged.

The work undertaken by NATSEM in collaboration with the ABS on statistical matching gave us a better understanding of the theoretical and practical issues in statistical matching, and how to evaluate the accuracy of the matched dataset. On this basis, the matched file that was estimated using constrained matching is the preferred base file for MediSim. Essentially, the person records in the original NHS were reshuffled into different families based on the HES family structure, such that information on every family member (that is essential to modeling the safety net) was available. Given that the original NHS person records have now been reconstructed into complete families, NATSEM's intention is to use only the variables from the NHS. With the exception of family structure, individual values in the NHS are preserved in the statistically matched file[9].

3 Imputation of Short Term Health Conditions

The inclusion of variables on health conditions in the model's dataset is the necessary first step to developing a facility in the model to measure health outcomes and to simulate policy changes with respect to people's health status and need for medicines. It is also essential for a microsimulation model that seeks to derive estimates of PBS drug usage for those with particular short term conditions under changing PBS rules. The 2001 NHS does not provide information on people's short-term (ST) health conditions. The 1995 NHS provides the most comprehensive data about short-term conditions, and was used to impute this onto the model population.

To impute ST conditions, the following tasks had to be undertaken: identify what short-term conditions need to be imputed; up-rate prevalence rates from 1995 to 2001; develop SAS code to impute the selected short-term health conditions onto the statistically matched dataset taking into account a range of explanatory variables; shift the focus from PBS users only to the entire Australian population; and move from a 2-weekly to an annual picture of short-term conditions.

[9] For our purposes, this only posed a problem with respect to the family income variable. As persons had been reconstructed into synthetic families, whilst retaining all original NHS variables, persons belonging to the same family had different family income values. To have a consistent family income, we averaged the income within each family. This average value was used to rank persons into income quintiles. Note however that the situation would be the same for any variable which relates to the whole family (and should thus be the same across the family), *e.g* rurality, SEIFA.

DEFINITIONS

Short-term conditions are defined as conditions that are experienced for less than 6 months or are expected to last for 6 months or less (ABS, 1996). *Prevalence rate* is the number of current cases or persons having the disease divided by the population at risk (Woodward, 2005, p.13).

WHAT CONDITIONS COUNT AS SHORT-TERM AND NEED TO BE IMPUTED

What specific health conditions are counted as short-term? Asthma, diabetes, arthritis and epilepsy are typically experienced as long-term conditions while dental problems, injuries, headaches, cough, colds, sore throat and influenza, and ear pain are typically short term in nature. In between is a wide spectrum of conditions that are characterised by some as only short-term, only long-term, or both short and long-term.

The 1995 NHS has information on the number of persons reporting specific health conditions and the proportion indicating that the condition is short-term only. We designate as short-term those conditions where the proportion of 'only short term' as opposed to 'only long-term' or 'both short and long-term' to the total number of persons is more than 5%.

Next, we eliminated those short term health conditions for which information is available on the 2001 NHS, as information was collected on the seven national health priority areas. We used the ICD9 classification of health conditions, as this classification is common to both the 1995 and 2001 health surveys. However, we aggregated some specific conditions. For example, we combined sciatica, curvature of spine, diseases of the intervertebral disc, and unspecified back problems into one category – back problems. The main purpose was to increase the probabilities as when these are too low (and considering the indivisibility of record weights), it is difficult to accurately impute such probabilities. In summary, the number of specific conditions to be imputed was narrowed down to some 50 conditions, listed in Appendix Table B3.

UPRATING SHORT-TERM CONDITIONS TO 2001 LEVELS

How did we uprate the prevalence of short term health conditions to 2001 levels, given that the latest comprehensive information we have is from the 1995 NHS? The most straightforward way was to assume that the change in the prevalence rate for short-term conditions was the same as change in the rate of prevalence for long-term conditions (noting that for most conditions, there is a great deal of overlap between long-term and short-term conditions). For each short-term condition, we applied the change in the long-term prevalence rate over the period 1995-2001, to bring up the 1995 short-term prevalence rate to 2001 levels.

The underlying assumption is that most conditions have both a short-term and long-term element, and we expect the rate of change in the long-term element to be in the same range as the rate of change in the short-term element. If there was insufficient data from the NHS, we supplemented this with information from the Australian Institute of Health and Welfare (AIHW) population hospital morbidity data over the period 1995-2001 for specific health conditions.

There were many changes in the health surveys between 1995 and 2001, including changes in the definition and classification of health conditions. The change in the number of

(specific) health conditions over the period could in part be due to differences in survey methodology. Nevertheless, whatever the reason for the changes in prevalence over the period 1995-2001, what was important was to reflect these changes in the model base file[10].

TWO-WEEK WINDOW FOR REPORTING SHORT-TERM CONDITIONS

In the 1995 health survey, short-term conditions (or recent illness) are identified through an actions-based approach. Respondents were asked whether they had taken certain types of action in the previous two weeks, and the medical condition or other reasons for those actions. A two-week reference period was adopted as a compromise between minimising respondent recall errors and ensuring sufficient observations were recorded from which reliable estimates could be produced. The data were collected over a 12-month period so any seasonality occurring for particular conditions would have been accounted for. In summary, information from the 1995 NHS can be taken to estimate the prevalence of short-term health conditions in any two-week period during that year (ABS 1996, pp 121 and 125).

Although the 2-week period is practical for recall purposes, the relative shortness of this period has implications for our purposes. While the survey is able to capture information on all survey respondents that had a long-term condition, it is only able to capture information on a fraction of the respondents that had a short-term condition in a year. For example, those experiencing short-term conditions a month back would not be counted.

The next table shows the distribution of the Australian population by term of health condition. Nine percent of the 1995 respondents reported having a short-term health condition (only) within the past 2 weeks whereas, if we were to take an annual picture, the proportion of respondents that suffered from short-term health conditions sometime during the past year would be considerably higher. Similarly, the proportion of respondents with no health condition in 1995 (whether short- or long-term) would be considerably lower than the 16% reported based on the two-week window. Comparable figures for 2001 are presented to indicate the change between the two years.

Table 3 **Distribution of persons by term of health condition, 1995 and 2001**

Year	Condition	No. of Persons	Dist.
		000	%
1995	Long term only or both recent and long-term	13,365	74
	Recent only (short-term)	1,625	9
	No health conditions (not applicable)	2,890	16
	Total population	18,061	100
2001	With a long term condition	14,737	78
	Without a long term condition	4,179	22
	Total population	18,916	100

[10] For selected long-term conditions, the ABS has made a study of the comparability of the 1995 and 2001 surveys (ABS 2003b). For some conditions, such as heart and circulatory, they have indicated that each successive survey has improved the breadth and specificity of questioning for heart and circulatory. For mental conditions, a greater level of public awareness (and acceptance) may have influenced the higher rates of disclosure of mental health conditions.

In conclusion, while the 1995 NHS does have the most comprehensive information on short-term conditions, we need to take note of the limitations of the two-week window for reporting short-term conditions, particularly when what we actually require for the modeling, are annual estimates.

IMPUTING THE PREVALENCE OF SHORT-TERM HEALTH CONDITIONS

Numerous studies have found a link between health and income (see Walker and Abello (2000) for review of relevant literature). We take concession card status as a proxy for income. Likewise, there are clear patterns in the distribution of disease by age, gender and self-assessed health status. Given the foregoing, the imputation of short-term conditions was based on concession card status, gender, age and self-assessed health status. Persons were divided into the following age groups, which were set up to represent differences in the prevalence of health conditions across these age ranges: 0-4, 5-14, 15-24, 25-39, 40-64, and 65 years and over.

Overview of imputation process - Imputing short-term conditions onto the model population involved sorting the person records in the model base file by card status, gender, age group and self-assessed health status. For each short-term health condition, person records were selected at random to 'have' the ST health condition until the uprated prevalence figures were met.

Two-monthly prevalence rates - Short-term conditions in the 1995 NHS were based on health conditions persons reported having over the previous two weeks. The prevalence of short-term (ST) health conditions on a two-weekly basis were obtained from the 1995 NHS then uprated to 2001 levels. The use of a fortnightly prevalence rate to impute short-term conditions onto the model dataset implies that we would have to impute 26 times to bring the fortnightly estimates up to annual estimates.

Another approach is to scale up or multiply by a number greater than 1.0 the prevalence rates (expressed as a percentage of the population) before imputing. However the scaled-up prevalence rates should not exceed 1.0 as a prevalence rate of 1.0 implies that all persons in the selected group will "get" the health condition. If the rate is greater than 1.0 then the target prevalence rate will not be achieved as there will not be sufficient number of persons to be imputed the disease. An inspection of the fortnightly prevalence rates for all short-term conditions showed that when multiplied by 26/6 (multiplied by 26 fortnights, and divided by 6 months), none exceeded 1.0, so a two-monthly prevalence rate was deemed to be reasonable.

To expand the two-week estimate to a two-monthly estimate, prevalence rates were multiplied by 26/6. The imputation procedure using the two-monthly prevalence rates was carried out six times, resulting in annual estimates of the prevalence of short-term conditions in the model population. Each two month imputation was independent of previous imputations. Doing the imputation this way implies that a person has one chance of getting a specific short-term health condition every two months. In actuality, some persons do experience some conditions (e.g. colds) more frequently than that. In general, however, we expect this assumption to be reasonable.

Take as an example, influenza. The two-monthly prevalence rates for this condition are shown in Table 4 by a person's age group, gender and card status; for purposes of simplicity we do not present the rates by self-assessed health status. The first figure of 0.19

indicates that over a period of two months, we expect 19% of male concession cardholders aged 0-14 to be imputed to "get" influenza while, for non-concession cardholders (the general population), the percentage is almost half this at 10%.

Using two-monthly prevalence rates, the imputation process was done six times, for each short-term condition, so we end up with annual prevalence rates. Every two months about 19% of males aged 0-14 among concessional patients would be imputed to get influenza once. Done 6 times, we expect about 114% or practically everyone in this group of males to be imputed to have influenza over a year's time, with a couple of repetitions for some as indicated by the annual prevalence exceeding 100%. Because of the imputation process used, a person could be imputed at most, to catch influenza six times in a year.

Table 4 **Two-monthly prevalence rate of influenza, 2001 (% population)**

	Concession card status			
	Concessional		General	
	Male	Female	Male	Female
0-14 years	0.19	0.14	0.10	0.10
15-34 years	0.15	0.21	0.11	0.13
35-49 years	0.10	0.20	0.16	0.17
50-64 years	0.16	0.16	0.19	0.18
65-74 years	0.14	0.12	0.15	0.13
75 years plus	0.06	0.07	0.06	0.07

Source: NATSEM estimates.

Cloning - To adequately assign short-term health conditions across the whole population, it was necessary to *clone* the NHS records i.e. create multiple records of the same person in the dataset. Each record in an ABS national survey has a weight representing the likelihood of finding persons with a similar set of characteristics in the Australian population. Records can be duplicated with each clone i.e. each new record having a proportionally smaller weight (maximum weight was set at 200). The cloning of records with smaller weights enabled us to impute short-term conditions with greater accuracy.

4 Imputing Prescribed Drug Usage

Data on the use of prescribed medicines is available in the 2001 NHS only for national health priority areas that are specified by the respondent as long-term. For short-term and non-priority health conditions, annual drug usage has to be imputed.

Imputing fortnightly drug usage involved the following steps: (a) estimating the probability of taking prescribed drugs given that one has a specific health condition; and (b) given that one takes prescribed drugs, we then model the number and type of drugs taken. Steps (a) and (b) were done separately for short-term and long-term conditions.

IMPUTING THE PROBABILITY OF TAKING PRESCRIBED DRUGS

Not all persons who have health conditions take prescribed medication or even any medication at all. For each health condition, the 1995 NHS provides information on the

proportion of persons having such conditions that take prescribed drugs[11]. This proportion varies by the type of health condition (with the proportions being very high for those with diabetes, most heart conditions, epilepsy, other hereditary diseases of the nervous system, and contraceptive management, among others). Initially, we considered the possibility of grouping conditions based on closeness in the proportion of persons taking prescribed drugs. However, it was hard to find a common pattern and the best option seemed to be to do it on a specific condition basis.

We estimated the probability of taking prescribed drugs, given that a person had a specific health condition, taking into account differences in gender, age, and card status. Card status is particularly important, as the usage of PBS drugs is much higher for concession cardholders than non-concession cardholders. Age is also particularly important, as the oldest and youngest age groups tend to have a higher proportion taking prescribed drugs.

IMPUTING THE NUMBER AND TYPES OF DRUGS

After the proportion of persons (by specific health conditions) taking prescribed drugs had been established, the next step was to identify the number and types of prescribed drugs taken for each specific health condition. The variables taken into account include type of health condition, gender, age and card status.

CONVERTING FORTNIGHTLY TO ANNUAL DRUG USAGE

There was no need to adjust the imputed drug usage for short-term conditions, due to the approach taken of scaling up the prevalence levels to a two-monthly period and implementing the imputation procedure six times. To convert the imputed fortnightly long-term drug usage into an annual figure, each person's script for a long-term condition was multiplied by 12, for each type of drug used. This presumes that a person with a chronic condition requiring the use of prescription drugs will use the drug(s) regularly throughout the year. This also assumes that scripts are issued for a 1-month's supply, which is generally the case for chronic conditions in Australia.

The actual (as against imputed) drug usage for priority conditions available on the NHS was summed up and converted to an annual figure using a factor of 12. This, together with the imputed scripts for non-priority conditions, represents the total number of scripts used annually by the total Australian population excluding Veterans[12].

[11] For most health conditions (except diabetes, heart problems, depression, psychoses, and epilepsy and a few other conditions) a large proportion of persons did not report using any prescribed medication over the previous two weeks in the 1995 NHS. One reason for this could be that not all persons who have a long-term condition currently have an acute episode of that condition, in which case they may opt to take their medication less frequently or not at all. Other reasons could be that prescribed medication is not appropriate for that condition, or that non-prescribed medicine or other types of treatment are used to manage that condition.

[12] Veterans were excluded from the model as there is a scheme that parallels the PBS, called the Repatriation Pharmaceutical Benefits Scheme (RPBS) that is separately administered for Australian war veterans and their dependants.

PBS SCRIPTS: ALIGNING IMPUTED DRUG USAGE TO ADMINISTRATIVE DATA

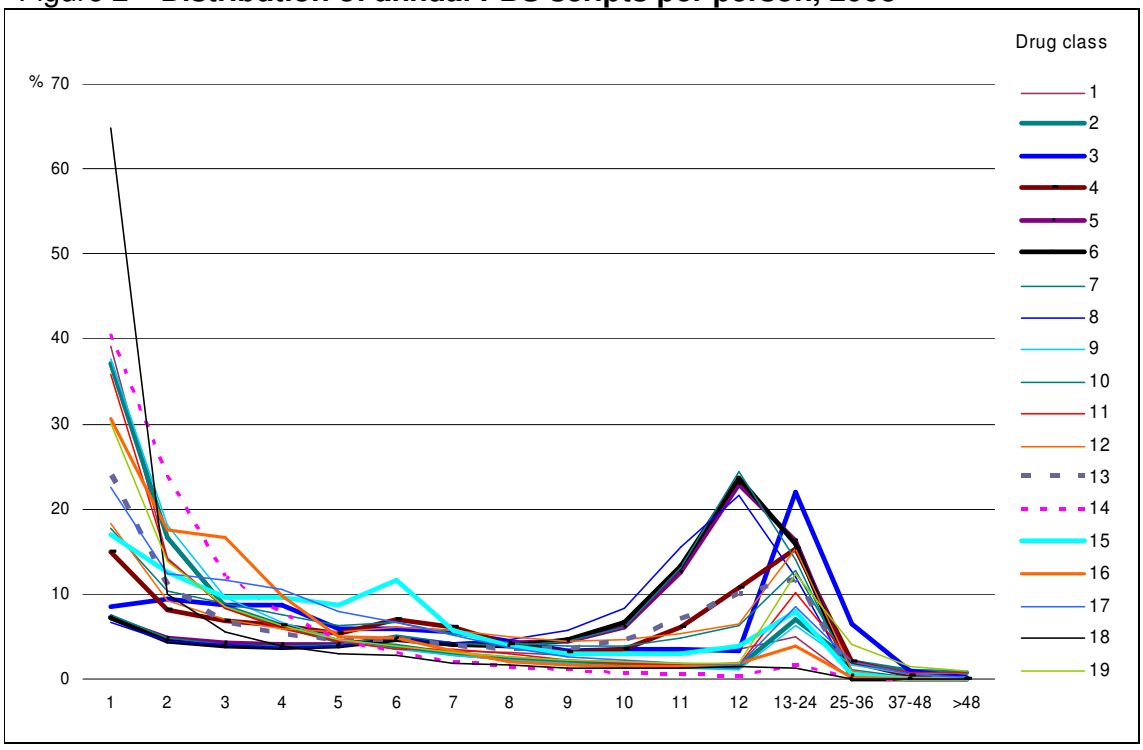
After annual estimates of total prescribed drug usage had been imputed, we then shifted the focus to PBS prescriptions. In Australia, total prescribed drug usage includes three types of medicines:

- scripts for drugs with a cost to government under the PBS (known elsewhere as 'benefit' drugs and called *Group 1* scripts in MediSim);
- scripts for PBS-listed prescribed medicines not attracting a government subsidy — that is, scripts with a price below the PBS copayment level (below copayment drugs - *Group 2* scripts); and
- scripts for prescribed drugs not listed under the PBS (private medicines - *Group 3 Scripts*).

The focus of the model is on PBS benefit or Group 1 drugs only so the alignment procedure selects and then aligns PBS scripts to administrative numbers.

The total annual scripts (imputed and actual) were summed up for each of the 19 drug classes in MediSim (see Appendix Table B1 for drug classes used in MediSim). For benchmarking purposes, we obtained administrative data on the number of PBS scripts per year per person by concession card status, gender, five-year age group and drug class for the year 2003[13] (see figure 2). We also had administrative data on the number of persons using PBS drugs, by the same categories.

Figure 2 **Distribution of annual PBS scripts per person, 2003**



[13] 2003 is the first calendar year for which Medicare Australia had data on the distribution of scripts per person by gender, age group and drug classification for both concession cardholders and non-cardholders. Previous to this, data were only collected on concession cardholders that had not yet reached the safety net.

Data source: Medicare Australia
Note: See Appendix Table B1 for listing of drug classes.

For each of the 19 drug classes, Figure 2 shows the proportion of PBS beneficiaries using only 1 script per year, 2 scripts per year, and so forth. Particular drug classes, such as drug class 18 (direct acting antivirals) and drug class 14 (antibiotics) are used only once a year by most persons, whereas other drugs taken for chronic conditions have higher annual scripts, with a peak at 12 scripts and 24 scripts per year.

Total script usage may be viewed as the product of number of persons using a particular drug type, and number of scripts used per drug type. To validate the data on drug usage, we examined this on two fronts. First, we looked at the number of persons taking drugs, and second, we looked at the number of scripts per person. Steps were taken to more closely align the imputed drug usage to administrative numbers.

Number of persons taking drugs The total number of scripts in the model base file looked reasonable. However, when looking at the number of persons using drugs and their average number of scripts, there were large discrepancies between what we had in the model, at that stage, and comparable administrative numbers from Medicare Australia[14]. In particular, administrative data showed many more persons using PBS medicines than we had estimated in the model, for most (14 out of 19) drug classes, particularly for the very young and the very old. The exceptions were these five drug classes: anti-inflammatories, vasodilators beta blockers, calcium channel blockers, anxiolytics and hypnotics and direct acting antivirals.

This discrepancy may be attributed in part to the non-inclusion of institutionalised persons in the NHS data (with the NHS survey only covering people living in private dwellings)[15]. Supporting this assumption about the institutionalised, the "lack of persons" was true for most drug classes, even those that cater primarily to priority conditions, for which actual drug usage was available in the 2001 NHS. This includes asthma medications, diabetes medications, heart condition related drugs, mental health related therapies and cancer drugs. Thus, we revised the methodology to increase the proportions of persons taking medication for the 14 drug classes, until they matched administrative numbers. These individuals were taken from the group of persons having health conditions that required the use of those particular drugs.

Finally, for each drug class, we identified persons in the base file using the drug, listed their scripts classified by their card status-gender-age group, and designated (at random) some as PBS scripts until we had sufficient number of PBS users to match administrative numbers for each drug class-card status-gender-age group. Note that since persons may take multiple drugs, persons in the base file may have some of their scripts designated as PBS scripts, and others designated as non-PBS scripts.

Number of scripts per person Next we looked at the number of annual scripts per person in the model base file and compared this with administrative data from Medicare Australia. The administrative data shows, for each drug class, card status, gender and age group, how

[14] Medicare Australia was previously known as the Health Insurance Commission (HIC).
[15] The institutionalised include those in hospitals and in homes for the aged and disabled. While we have no data on their drug usage, we expect this to be much higher than the drug usage of persons not living in those institutions.

many persons used one script, two scripts, and so forth on an annual basis. To make this clearer we present an extract of the data on the distribution of scripts in Table 5. The table shows the distribution of PBS scripts of males with concession card status, using anti-inflammatories (drug class 1) in 2003. For example, the first cell in the table shows that 231 or the majority of young male children aged between 0-4 years used only one script in 2003.

The alignment procedure was as follows: for each of the 19 drug classes, by card status, gender and age group, scripts in the model base file were ranked from highest to lowest. This is compared with administrative data on the distribution of scripts. The administrative data on the distribution of scripts per person by drug class was used to revise the initial level of PBS scripts. In part, this revision of scripts may be viewed as a refinement to the crude method taken of giving each person with a long-term condition 12 scripts to convert the two-weekly drug usage into an annual figure.

Table 5 **Example on distribution of PBS scripts per person (anti-inflammatories, males with concessional card status) by age group, 2003**

No. of scripts	Number of persons by age group					
	0-4	5-14	15-24	25-34	35-64	65+
1	231	3,426	21,342	40,650	86,569	85,152
2	32	335	3,092	10,005	35,847	41,216
3	6	119	860	4,194	20,611	27,669
4	9	70	415	2,344	15,746	23,305
5	6	37	198	1,368	10,463	16,614
6	2	24	130	908	8,504	14,748
7	2	19	98	706	7,417	13,763
8	1	10	34	452	6,172	12,110
9	1	11	37	420	7,357	15,798
10	1	7	99	867	12,238	22,213
11	0	7	88	597	6,746	12,779
12	0	19	54	489	5,960	11,315
13-24	0	12	51	413	7,075	14,677
25-36	0	0	2	48	363	330
37 or more	0	0	0	4	45	26
Total	**291**	**4,096**	**26,500**	**63,465**	**231,113**	**311,715**

Data source: Medicare Australia

CONCLUDING NOTE ON IMPUTATION OF DRUG USAGE

Figures comparing the distribution of drug usage in the model base file with administrative data on actual PBS scripts are provided in Appendix tables C2 to C4. Because of the alignment method adopted, the designated PBS scripts in the model very closely approximate the distribution of (actual) administrative data on total scripts by drug class and the distribution of scripts per person, by drug class, concession card status, gender and age group. This represents a substantial advance in the data on drug usage. In the previous version of the model, while the aggregates per drug class were accurate, we were less certain about whether *scripts per person* were reasonable. Similarly, the distribution of PBS users in the model closely approximates the distribution of PBS users based on administrative data. Note that in order to more closely align scripts at the person level with

administrative data, the model's records had to be cloned further, with each record having a maximum weight of 100 or less.

5 Summary and Conclusion

A number of steps were taken to overcome the 2001 NHS survey limitations as the main base file for MediSim. First, we statistically matched the NHS with another ABS national survey, to create synthetic families and get a complete record for every individual within each family, as family level information is needed to model PBS safety nets. The statistical matching allowed the retention of the health information available on NHS 2001, whilst borrowing the family structure from another survey. Next, we imputed short-term health conditions based on detailed information in the previous (1995) NHS and converted the two-weekly prevalence rates derived from the 1995 survey's two-week recall period to annual figures. Finally, we imputed annual drug usage for short-term and non-priority long-term health conditions. These initial estimates were then aligned to administrative data on PBS benefit drugs.

The application of statistical matching methods and use of complementary data sets significantly improved the usefulness of the 2001 NHS as a base dataset for MediSim, and enabled improved modeling of the PBS safety net. These enhancements to the national health survey have improved the capability of MediSim as a microsimulation tool for policy-makers, the pharmaceutical industry, and health researchers and consumers.

Appendix

A Statistical matching tables

Table A1 **Quality of the match: number of usual residents variable**

	NHS	HES number of usual residents					
		1	**2**	**3**	**4**	**5**	**6**
Unconstrained	**1**	93	7	.	.	.	.
	2	7	90	3	0	0	0
	3	3	9	80	7	1	1
	4	1	3	7	86	2	1
	5	1	2	3	13	79	2
	6	3	1	6	11	14	65
Constrained	**1**	81	14	3	1	0	0
	2	4	82	10	3	1	0
	3	2	18	61	18	1	0
	4	1	7	12	72	7	1
	5	1	1	2	16	67	14
	6	2	1	2	3	20	72

Table A2 **Quality of the match: equivalent income decile variable**

	HES equivalent income unit decile									
NHS	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Unconstrained										
1	76	20	4	0	0	.	.	.	.	.
2	13	75	9	3	0	0	.	.	.	.
3	6	6	73	9	4	1	0	.	.	.
4	0	1	5	72	16	5	1	0	0	0
5	0	1	3	11	70	12	3	0	0	0
6	.	0	1	2	7	82	5	2	1	0
7	0	0	.	0	1	8	85	3	2	0
8	.	.	.	.	0	0	4	87	5	4
9	.	.	.	.	0	0	0	3	88	9
10	.	.	.	.	.	0	.	0	3	97
Constrained										
1	35	10	7	12	14	12	3	2	1	3
2	12	31	17	16	9	7	3	1	2	3
3	5	11	28	15	16	11	2	2	3	6
4	3	7	2	23	15	12	9	6	8	15
5	4	6	3	4	23	13	11	12	11	13
6	3	6	4	4	7	31	16	12	9	9
7	2	7	4	6	5	8	42	13	6	6
8	2	5	4	5	4	5	9	47	13	7
9	1	3	5	8	6	5	8	7	46	12
10	1	1	4	8	7	3	5	5	12	54

B Classifications in MediSim

Table B1 **Drug classes in MediSim**

1	Anti-inflammatories
2	Asthma medications
3	Diabetes medications
4	Vasodilators & beta blockers
5	ACE inhibitors
6	Angiotensin IIs
7	Calcium channel blockers
8	Cholesterol & triglyceride reducers
9	Analgesic medications
10	Antipsychotics
11	Anxiolytics & hypnotics
12	Antidepressants
13	Stomach medications
14	Antibiotics
15	Antineoplastics
16	Genitourinary
17	Anti-epileptics
18	Direct acting antivirals
19	All other medications

Table B2 **Health conditions in MediSim**

1	NHPA Arthritis	Rheumatoid arthritis
		Osteoarthritis
		Arthritis NEC
2	NHPA Asthma	Asthma
3	NHPA Diabetes	Diabetes Mellitus - Type 1
		Diabetes mellitus - Type 2
		Diabetes unspecified
4	NHPA Heart or circulatory condition	Atherosclerosis
		Fluid problems NOS
		Varicose veins
		Haemorrhoids
		Other diseases of circulatory system
		Hypertension
		Heart disease
		Stroke
		Ill-defined symptomatic heart condition
		Cerebrovascular disease
5	NHPA Mental	Nerves tension nervousness
		Other mental disorders
		Depression
		Psychoses
		Emotional problems NEC
		Body image & eating disorders
		Alcohol and drug dependence
		Mental retardation specific delays in development
6	NHPA Cancer	Skin cancer
		Breast cancer
		Neoplasms NEC
7	NHPA Injury poisoning	Complications surgical NEC
		Fractures
		Dislocations, sprains and strains
		Internal injuries
		Open wounds
		Bruising and crushing
		Entry of foreign bodies
		Burns and scalds
		Poisoning other than by food
		Other injuries
		Injuries type not stated
8	Musculoskeletal system	Back problems
		Other diseases musculoskeletal
		Osteoporosis
		Rheumatism
		Absence of limbs or parts
		Musculoskeletal deformities
9	Other endocrine	Thyroid disease

	Gout
	Obesity
	Other endocrine and immune diseases
	High blood sugar
	High cholesterol
10 Respiratory system	Bronchitis/ Emphysema
	Sinusitis
	Cough or sore throat
	Other diseases of respiratory system
	Common cold
	Hayfever
	Influenza
11 Eye & adnexa	Blindness not corrected glasses
	Other diseases of eye and adnexa
	Visual disturbances
	Cataracts
	Glaucoma
	Hypermetropia/Far-sighted
	Myopia / Short-sighted
	Presbyopia
12 Ear & mastoid	Otitis media
	Deafness (complete/ partial)
	Ear pain
	Other diseases ear & mastoid proc
13 Nervous system	Epilepsy
	Other diseases nervous system
	Migraine
	Paralysis
	Other hereditary diseases nervous system
14 Infectious diseases	Herpes
	Tinea
	Other infectious diseases
15 Digestive system	Diarrhoea enteritis
	Ulcer
	Hernia
	Constipation
	Dental problems
	Other diseases digestive system
16 Genito-urinary	Kidney diseases
	Other diseases urinary system
	Other diseases genital system
	Disorders of menstruation
17 Skin subcutaneous tissue	Skin rash NOS
	Eczema dermatitis
	Acne
	Other diseases skin and subcutaneous tissue
	Psoriasis
18 Diseases of the blood	Diseases of blood
19 Complications childbirth	Complications of pregnancy etc

20	Other Signs & symptoms	Allergy NEC Insomnia Pyrexia Localised swelling Difficulty breathing Chest pain Abdominal pain Heartburn Dizziness Headache due to stress Headache unspecified or trivial Virus Other symptoms ill-defined conditions
21	Preventive measure	Immunisation
22	Congenital conditions	Congenital anomalies
23	Disability nec	Speech impediment NEC Blackouts loss of cons NEC Missing organs NEC

Note: NHPA = National Health Priority Area

Table B3 **Short term health conditions imputed in MediSim**

Musculoskeletal system	Back problems
	Other diseases musculoskeletal
Other endocrine system	Thyroid disease
	Gout
	Other endocrine and immune diseases
Respiratory system	Bronchitis/ Emphysema
	Cough or sore throat
	Other diseases of respiratory system
	Common cold
	Influenza
Eye & adnexa	Other diseases of eye and adnexa
	Visual disturbances
Ear & mastoid	Otitis media
	Ear pain
	Other diseases ear & mastoid proc
Nervous system	Other diseases nervous system
	Migraine
Infectious diseases	Herpes
	Tinea
	Other infectious diseases
Digestive system	Diarrhoea enteritis
	Constipation
	Dental problems
	Other diseases digestive system
Genito-urinary system	Kidney diseases
	Other diseases urinary system
	Other diseases genital system
	Disorders of menstruation
Skin and subcutaneous tissue	Skin rash NOS
	Eczema dermatitis
	Acne
	Other diseases skin and subcutaneous tissue
	Psoriasis
Diseases of the blood	Diseases of blood
Complications of childbirth	Complications of pregnancy etc
Other signs & symptoms	Allergy NEC
	Insomnia
	Pyrexia
	Localised swelling
	Difficulty breathing
	Chest pain
	Abdominal pain
	Heartburn
	Dizziness
	Headache due to stress
	Headache unspecified or trivial
	Virus
	Other symptoms ill-defined conditions

C Drug usage tables

Table C1 **Distribution of drug usage per specific health condition**

Health group		Health condition	Drug class																			
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	All
1	NHPA Arthritis	Rheumatoid arthritis	48	.	.	0	.	.	0	.	10	.	1	1	.	.	.	.	.	.	39	100
		Osteoarthritis	63	.	0	1	0	.	0	0	19	.	1	1	.	.	.	.	.	.	14	100
		Arthritis NEC	63	1	1	0	.	.	1	0	16	.	0	1	.	.	.	.	.	.	18	100
2	NHPA Asthma	Asthma	.	98	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	100
3	NHPA Diabetes	Diabetes Mellitus - Type 1	.	.	.	97	0	0	.	.	1	.	.	0	.	.	.	.	.	.	1	100
		Diabetes mellitus - Type 2	.	.	.	93	0	2	1	1	0	.	.	0	.	.	.	.	.	.	3	100
		Diabetes unspecified	1	.	.	92	0	1	.	1	0	.	.	1	.	.	.	.	.	.	3	100
4	NHPA Heart circ	Atherosclerosis	5	.	.	16	.	1	33	6	11	.	.	2	.	.	.	.	.	.	26	100
		Fluid problems NOS	1	.	.	0	1	.	.	.	0	.	.	.	.	.	.	.	.	.	98	100
		Varicose veins	9	.	.	.	1	.	.	.	14	.	4	.	.	.	.	.	.	.	72	100
		Haemorrhoids	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	96	100
		Other dis circulatory system	1	0	.	10	2	4	28	0	47	0	.	.	.	.	.	.	.	.	8	100
		Hypertension	0	0	0	20	27	0	25	0	2	0	0	0	.	.	.	.	.	.	25	100
		Heart disease	0	0	0	34	9	0	22	1	7	0	0	0	.	.	.	.	.	.	26	100
		Stroke	2	.	.	10	2	13	18	1	31	.	.	.	.	.	.	.	.	.	23	100
		Ill-def symp heart cond	1	.	0	25	9	0	18	1	5	0	0	0	.	.	.	.	.	.	40	100
		Cerebrovasc dis	9	.	.	.	.	29	.	.	11	.	.	.	.	.	.	.	.	.	51	100
5	NHPA Mental	Nerves tension nervousness	.	.	.	.	.	.	.	.	.	11	43	39	.	.	.	.	.	.	8	100
		Other mental disorders	.	.	.	.	.	.	.	.	.	9	5	74	.	.	.	.	.	.	11	100
		Depression	.	.	.	.	.	.	.	.	.	12	9	69	.	.	.	.	.	.	10	100
		Psychoses	.	.	.	.	.	.	.	.	.	60	11	15	.	.	.	.	.	.	14	100
		Emotional problems NEC	.	.	.	.	.	.	.	.	.	34	14	44	.	.	.	.	.	.	8	100
		Body image & eating disorders	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	100
		Alcohol and drug dependence	.	.	.	.	.	.	.	.	59	.	.	6	.	.	.	.	.	.	35	100
		Mental ret delays devt	.	.	.	.	.	.	.	.	.	15	.	79	.	.	.	.	.	.	6	100
6	NHPA Cancer	Skin cancer	.	.	.	.	.	.	.	.	7	.	.	.	.	.	92.5	.	.	.	.	100

Health group	Health condition	Drug class																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	All
	Breast cancer									0						100					100
	Neoplasms NEC									21		5	1			73.6					100
7 NHPA Injur pois	Complic surgical NEC	1	1		. 2	0		1		. 13		. 3	4		59					16	100
	Fractures	18	1		.	.	.		.	. 53		. 5	0		17					5	100
	Disloc sprains and strains	50		.	.	.	.	1		. 30		. 7	1		8					2	100
	Internal injuries	11		.	.	.	.		.	. 37		. 0	0		41					11	100
	Open wounds	0		.	.	.	.		.	. 12		. 0	0		70					19	100
	Bruising and crushing	23		.	.	.	.		.	. 43		.	.	.	27					7	100
	Entry of foreign bodies	0		.	.	.	.		.	. 5		.	.	.	75					20	100
	Burns and scalds	0		.	.	.	.		.	. 14		.	.	.	68					18	100
	Poisoning other than by food	0		.	.	.	.		.	. 0		.	.	.	79					21	100
	Other injuries	0	2		.	.	.	1	3	10		. 0		.	66					18	100
	Injuries type not stated	100		.	.	.	.		.	. 0		.	.	.	0					0	100
8 Muscoskel sys	Sciatica	46		.	.	.			.	. 38		. 8	0							8	100
	Dis of the interver disc	29		.	.	.			.	. 51		. 6	4							11	100
	Back problems (unspecified)	35		.	. 0				.	. 46	0	8	3							7	100
	Other dis musculoskel	32			. 1			1		. 19	0	4	3							40	100
	Curvature of spine	32			.	.			.	. 53		. 0	15							0	100
	Osteoporosis	10			.	.			.	. 6		. 1	0							83	100
	Rheumatism	48		.	.	. 6			. 3	11		. 0	0							32	100
	Absence of limbs or parts	0		.	.	.	.		.	. 19		. 0	0							81	100
	Musculoskeletal deformities	39		.	.	.	.		.	. 30		. 3	0							28	100
9 Other endocrine	Thyroid disease																			100	100
	Gout	15	0		. 0	0		0		. 2		. 0		.						82	100
	Obesity		.	.	.	.	.		. 2	3		.	.	.						95	100
	Oth endocrine and imm	0	0	0	0	0		0	1	0	0	0	0							98	100
	High blood sugar		.	. 91		.	.		.	.	.	.	.	.						9	100
	High cholesterol		.	. 0	0	1	1	0	97	1		.	.	.						1	100
10 Respirat sys	Bronchitis/ Emphysema		. 66		1		.	0		. 0		. 0	0		17				0	16	100

Health group	Health condition	Drug class																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	All
	Sinusitis	0	52	.	.	.	.	.	.	3	0	.	.	.	22	.	.	.	1	21	100
	Cough or sore throat	0	19	.	.	.	.	.	.	3	.	.	.	.	39	.	.	.	1	37	100
	Other dis respiratory system	0	24	.	.	1	.	1	.	2	0	0	.	.	36	.	.	.	1	34	100
	Common cold	.	12	.	.	.	.	.	.	8	.	.	.	.	41	.	.	.	1	39	100
	Hayfever	.	49	.	.	.	.	.	.	1	.	0	0	.	25	.	.	.	1	24	100
	Influenza	.	8	.	.	.	.	.	.	4	.	.	.	.	45	.	.	.	1	42	100
11 Visual dist	Blindness not corr glasses	20	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	79	100
	Other dis eye and adnexa	.	2	.	3	.	.	0	.	2	.	.	.	.	60	.	.	.	1	33	100
	Visual disturbances	.	.	.	2	.	.	.	.	.	.	13	.	.	.	.	.	.	.	85	100
	Cataracts	.	.	.	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	92	100
	Glaucoma	.	.	.	27	.	.	1	.	0	.	.	.	.	.	.	.	.	.	72	100
	Hypermetropia/Far-sighted	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	100
	Myopia / Short-sighted	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	100
	Presbyopia	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	100
12 Ear & mastoid	Otitis media	.	1	.	.	.	.	.	.	1	.	0	.	.	62	.	.	.	1	34	100
	Deafness (complete/ partial)	.	.	.	.	.	.	48	.	.	.	.	.	.	34	.	.	.	.	18	100
	Ear pain	3	2	.	.	.	.	.	.	15	.	.	.	.	51	.	.	.	1	28	100
	Oth dis ear & mastoid proc	.	1	.	1	.	.	12	.	1	1	7	1	.	50	.	.	.	1	27	100
13 Nerv sys	Epilepsy	0	.	.	0	.	.	.	.	.	0	5	1	.	.	.	.	93	.	.	100
	Other dis nervous system	16	.	.	.	.	.	.	.	19	1	1	15	.	.	.	.	.	.	48	100
	Migraine	3	.	.	9	.	.	2	.	30	0	1	4	.	.	.	.	.	.	50	100
	Paralysis	.	.	.	.	.	.	.	.	14	12	10	.	.	.	.	.	.	.	64	100
	Oth hered dis nerv sys	1	.	.	1	1	.	2	.	7	.	7	5	.	.	.	.	.	.	76	100
14 Infectious dis	Herpes	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	98	.	100
	Tinea	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	100
	Other infectious dis	1	3	.	.	.	.	.	.	2	.	.	2	.	.	.	.	.	92	??	100
15 Digestive sys	Diarrhoea enteritis	.	.	.	.	.	.	.	1	2	.	0	0	75	.	.	.	.	2	19	100
	Ulcer	0	0	.	0	.	.	0	.	0	.	.	0	40	49	.	.	.	.	10	100

Health group	Health condition	Drug class																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	All
	Hernia	.	.	.	0	.	.	0	.	2	.	.	1	.	.	.	.	.	.	96	100
	Constipation	.	.	.	1	3	.	.	.	.	.	.	1	77	.	.	.	.	.	19	100
	Dental problems	2	2	.	.	.	.	.	.	24	.	1	.	.	.	.	.	.	.	72	100
	Other dis digestive system	1	.	.	0	0	.	1	0	4	0	1	1	.	.	.	.	.	.	92	100
16	Genito-urinary	Kidney diseases	1	0	.	3	2	.	1	1	3	.	.	1	.	.	.	89	.	.	.	100
	Other dis urinary system	.	.	1	0	1	.	0	.	.	.	1	7	.	.	.	76	.	4	10	100
	Other dis genital system	3	1	.	0	.	.	0	.	5	.	1	1	.	.	.	85	.	4	.	100
	Disorders of menstruation	6	.	.	0	.	.	.	.	2	.	0	1	.	.	.	88	.	4	.	100
17	Skin subcut tiss	Skin rash NOS	.	.	.	.	.	.	.	.	0	.	.	.	.	.	.	.	.	.	100	100
	Eczema dermatitis	0	1	.	.	.	.	.	.	.	0	0	0	.	.	.	.	.	.	99	100
	Acne	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	100
	Oth dis skin and subcut tiss	0	.	.	.	.	.	0	.	2	.	0	0	.	.	.	.	.	8	89	100
	Psoriasis	0	1	.	.	.	.	.	0	0	.	1	.	.	.	.	.	.	.	98	100
18	Dis of the blood	Diseases of blood	2	.	.	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	95	100
19	Complic chldbrth	Complic of pregnancy etc	.	13	.	.	.	.	.	.	.	.	.	.	.	69	.	.	.	.	18	100
20	Oth symptoms	Allergy NEC	1	33	.	.	.	.	.	.	1	.	1	1	.	.	.	.	.	.	64	100
	Insomnia	0	.	.	0	.	.	.	.	3	1	75	16	.	.	.	.	.	.	4	100
	Pyrexia	4	.	.	.	.	.	.	.	21	.	.	.	.	25	.	25	.	25	.	100
	Localised swelling	27	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	73	100
	Difficulty breathing	.	51	.	2	11	15	5	1	.	.	1	.	.	.	.	.	.	.	15	100
	Chest pain	9	8	.	23	.	15	12	.	17	.	.	2	.	.	.	.	.	.	15	100
	Abdominal pain	7	.	.	1	.	.	1	.	14	.	.	.	77	.	.	.	.	.	.	100
	Heartburn	2	.	.	.	.	.	.	.	.	.	.	.	98	.	.	.	.	.	.	100
	Dizziness	.	.	.	.	.	.	10	.	10	2	.	4	.	.	.	.	.	.	75	100
	Headache due to stress	.	.	.	.	.	.	.	.	48	.	23	.	.	.	.	.	.	.	28	100
	Headache unspec or trivial	4	0	.	1	.	.	0	.	86	0	0	2	.	.	.	.	.	.	5	100
	Virus	1	4	.	.	0	.	.	.	12	.	.	.	.	.	.	.	.	83	.	100
	Oth symp ill-defined cond	2	3	.	3	0	.	0	1	9	1	5	5	.	.	.	.	.	.	71	100
21	Preventive	Checkup/ examination	8	.	5	.	2	.	.	.	.	.	.	4	.	.	.	.	.	.	81	100

Health group	Health condition	Drug class																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	All
	Contraceptive management	.	.	.	.	0	.	.	.	.	.	1	.	.	.	.	99	.	.	.	100
	Counselling	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	100
	Immunisation	.	.	.	.	8	.	.	.	.	.	.	.	.	.	.	.	.	.	92	100
22 Congenital	Congenital anomalies	21	.	.	.	2	.	.	.	10	.	.	.	.	.	.	.	.	.	67	100
23 Disability	Speech impediment NEC	.	29	.	.	.	.	.	.	.	.	71	.	.	.	.	.	.	.	0	100
	Blackouts loss of cons NEC	.	.	.	.	.	.	.	.	.	.	42	.	.	.	.	.	.	.	58	100
	Missing organs NEC	.	49	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	51	100

Due to rounding totals don't always add up to 100.
Source: NATSEM estimates using 1995 NHS.

Drug class descriptions

1 Anti-inflammatories
2 Asthma medications
3 Diabetes
4 Vasodilators
5 ACE inhibitors
6 Angiotensins
7 Calcium channel blockers
8 Cholesterol & trig
9 Analgesics
10 Mental: antipsychotics
11 Mental: anxiolytics & hypnotics
12 Mental: antidepressants
13 Stomach medications
14 Antibiotics
15 Cancer: antineoplastics
16 Genitourinary
17 Anti-epileptics
18 Direct acting antivirals
19 All other medications

Table C2 **Distribution of persons by card status and age: actual vs. model estimates** (percent)

	Medicare Australia 2003		Model 2003-4		Percentage point difference	
	General	Concessional	General	Concessional	General	Concessional
All	**100**	**100**	**100**	**100**		
0-4	1	5	2	6	0	0
5-9	2	5	2	6	0	0
10-14	2	5	3	5	0	-1
15-19	4	6	4	7	0	-1
20-24	4	5	4	5	0	1
25-29	6	4	6	4	0	0
30-34	8	5	7	4	1	0
35-39	8	5	9	5	0	0
40-44	11	5	13	5	-2	0
45-49	13	5	15	4	-2	1
50-54	15	4	15	4	0	0
55-59	14	5	11	6	4	0
60-64	8	7	6	7	2	0
65-69	2	9	2	10	1	-1
70-74	1	9	1	9	0	0
75+	1	16	1	14	0	2

Table C3 **Scripts per person by card status, gender and age group: actual vs. model estimates**

		Medicare Australia		Model		% Difference	
Gender	Age group	Concessional	General	Concessional	General	Concessional	General
	All	**21.5**	**9.6**	**21.9**	**8.7**	**-2**	**10**
Male	0-4	4.2	2.5	4.4	2.6	-5	-2
	5-14	4.1	2.7	4.4	2.8	-7	-4
	15-24	5.1	3.8	5.4	3.4	-6	12
	25-39	10.1	5.6	10.0	4.9	1	13
	40-64	22.1	13.0	22.9	11.1	-3	17
	65 +	35.6	26.2	35.4	18.6	1	41
Female	0-4	3.7	2.5	4.0	2.5	-8	-1
	5-14	3.5	2.6	3.9	2.6	-10	-3
	15-24	5.8	3.6	5.7	3.3	2	7
	25-39	10.0	5.2	9.5	4.7	5	12
	40-64	24.6	10.8	24.0	10.1	3	6
	65 +	41.7	19.5	41.7	19.8	0	-1

Table C4 **Average no. of scripts by drug class: actual vs. model estimates**

	Model		Medicare Australia	
	Concessional	General	Concessional	General
Total scripts/person	21.50	9.62	21.85	8.74
No. of drugs classes/person	3.41	1.85	3.45	1.67
Ave script/person/drug	**6.30**	**5.20**	**6.34**	**5.23**
1 Anti-inflammatories	4.55	3.13	4.55	3.13
2 Asthma medications	6.54	3.32	6.54	3.32
3 Diabetes	10.67	5.64	10.67	5.62
4 Vasodilators	8.31	4.62	8.31	4.60
5 ACE inhibitors	9.46	8.18	9.46	8.18
6 Angiotensins	9.40	8.45	9.39	8.45
7 Calcium channel blockers	9.28	7.68	9.28	7.67
8 Cholesterol & trig. reducers	9.40	8.03	9.39	8.03
9 Analgesics	5.02	5.26	5.02	5.24
10 Antipsychotics	6.92	5.78	6.92	5.71
11 Anxiolytics & hypnotics	6.05	3.07	6.05	3.05
12 Antidepressants	7.49	6.43	7.49	6.43
13 Stomach medications	7.01	5.15	7.01	5.15
14 Antibiotics	2.87	2.32	2.87	2.32
15 Cancer: antineoplastics	5.71	5.79	5.71	5.78
16 Genitourinary	3.92	2.71	3.92	2.71
17 Anti-epileptics	6.08	5.20	6.08	5.19
18 Direct acting antivirals	2.26	2.67	2.26	2.67
19 All other medications	7.77	4.42	7.77	4.42

References

Abello, A, L Brown, A Walker and L Thurecht 2003, *An Economic Forecasting Microsimulation Model of the Australian Pharmaceutical Benefits Scheme*, Technical Paper no. 30, NATSEM, University of Canberra.

Australian Bureau of Statistics 2001, *Income Distribution, Australia*, Catalogue No. 6523.0, Canberra.

_____ 2003a, *National Health Survey: User's Guide, 2001*, Catalogue No. 4363.0.55.001, Canberra.

_____ 2003b, Occasional Paper: Long-term Health Conditions – A Guide to Time Series Comparability from the National Health Survey, Australia (ABS Catalogue no. 4816.0.55.001).

_____ 1996, *National Health Survey User's Guide, Australia, 1995*, Catalogue no. 4363.0.

Barr, R.S. and Turner, JS, 1978. "A new, linear programming approach to microdata file merging", 1978 Compendium of Tax Research, Office of the Treasury, Washington, D.C.

Brown, L, A Abello, B Phillips and A Harding 2004, Moving Towards an Improved Micro-Simulation Model of the Australian Pharmaceutical Benefits Scheme, *Australian Economic Review*, vol 37:1, pp. 41-61.

Brown, L, A. Abello and A Harding 2005a, "Distributional Impacts of the Pharmaceutical Benefit Scheme's Safety Net", Presented at the Australia Health Economist's Conference 29-30th of September 2005, Auckland.

_____2005b, "Distributional Impacts of the PBS in 2005-06: An Investment in Australian Families", The 3rd Annual Conference on The Future of the Pharmaceutical Benefits Scheme, 24-25 August 2005, Sydney

Cohen, M. 1991, 'Statistical matching and microsimulation models' in Citro, C.F. and Hanushek, E.A. (eds), *The Uses of Microsimulation Modelling*, *Review and Recommendations*, Vol 1, National Academy Press, Washington

Commonwealth Department of the Treasury, 2005, Budget Strategy and Outlook 2005-06, Budget Paper No. 1, Treasury, Commonwealth of Australia, May.

Harding, A, A Abello, L Brown and B Phillips, 2004, The Distributional Impact of Government Outlays on the Australian Pharmaceutical Benefits Scheme in 2001-02, *Economic Record*, 80 (Special Issue) S83-S96.

Moriarty, C., and Scheuren, F., 2001, 'Statistical Matching: Pitfalls of Current Procedures', in *Proceedings of the Annual Meeting of the American Statistical Association*, August 5-9, 2001.

Radner, D. B., Allen, R., Gonzalez, M. E., Jabine, T. B., and Muller, H. J. 1980, "Report on Exact and Statistical Matching Techniques," Statistical Policy Working Paper 5, U.S. Government Printing Office.

Rassler, Susanne, 2002. Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches, Springer-Verlag.

Rodgers, W., 1984, An Evaluation of Statistical Matching, Journal of Business and Economic Statistics, Vol. 2, No. 1, January 1984.

Sutherland, H, R Taylor and J Gomulka 2002, Combining household income and expenditure data in policy simulations, *Review of Income and Wealth,* vol. 48:4, pp. 517-536.

Taylor, R., Gomulka, J. and Sutherland, H. 2000, *Creating Order out of Chaos? Identifying Homogeneous Groups of Households Across Multiple Datasets*, Paper presented at the 26th General Conference of the International Association for Research in Income and Wealth, Cracow, Poland, 27 August to 2 September.

Technical Working Group ABS-NATSEM Collaboration on Statistical Matching 2004, Statistical Matching of the HES and NHS: An Exploration of Issues in the Use of Unconstrained and Constrained Approaches in Creating a Basefile for a Microsimulation Model of the Pharmaceutical Benefits Scheme, *ABS Methodology Advisory Committee Paper*, June 2004.

Walker, A 2000, Distributional impact of higher patient contributions to Australia's Pharmaceutical Benefits Scheme, *Australian Health Review*, vol. 23:2, pp. 32–46.

Walker, A and Abello, A 2000, *Changes in the Health Status of Low Income Groups in Australia, 1977-78 to 1995*, Discussion Paper no. 52, National Centre for Social and Economic Modelling, University of Canberra

——, R Percival and S Fischer 1998, *A Microsimulation Model of Australia's Pharmaceutical Benefits Scheme,* Technical Paper no. 15, National Centre for Social and Economic Modelling, University of Canberra.

——, —— and A Harding 2000, 'The impact of demographic and other changes on expenditure on pharmaceutical benefits in 2020 in Australia', in Mitton, L, Sutherland, H and Weeks, M (eds), *Microsimulation Modelling for Policy Analysis: Challenges and Innovations*, Cambridge University Press, pp. 149–71.

Woodward, M 2005, *Epidemiology: Study Design and Data Analysis*, Chapman & Hall/CRC Press, pp. 12-13.

Yoshizoe, Y and M Araki 1999, Use of statistical matching for household surveys in Japan. In *52nd Session of the International Statistical Institute*, Helsinki, Finland