PRELIMINARY

Family formation and the demand for private health insurance

Vineta Salale*, Denise Doiron*, Denzil G. Fiebig* and Elizabeth Savage**
***School of Economics, UNSW**
****CHERE, UTS**
Sydney, Australia

June 2008

Abstract

This study contributes to the international literature on health insurance by focussing on a specific group of individuals namely young women. We investigate the role of family formation, in particular the effect of children, actual and desired, on the purchase of private health insurance. A unique panel data set of Australian women allows us to focus on a large cohort of young women below the age of 30. The model of insurance used, explicitly accounts for state dependence while controlling for correlation due to the presence of unobserved and time-invariant individual effects. We also estimate specifications in which attrition is modelled through the use of inverse weighted probabilities. Evidence of differential demand for insurance is found when we distinguish between young women who have started their families, those who have reached their desired number of children and those intending to have additional children. The different effects on joining and leaving cover show the importance of modelling dynamics in insurance.

1 INTRODUCTION

The Australian private health insurance system has been the subject of many studies over the last two decades; for example Cameron et al (1998), Barrett and Conlon (2003), Savage and Wright (2003), Palangkaraya and Yong (2005) and Doiron, Jones and Savage (2008). Several characteristics make this insurance system amenable to economic modelling: take-up is substantial (44% of the population insured in 2007; PHIAC 2007), insurance is not tied to employment; prices are regulated and are essentially community-rated. Furthermore, recent policy reforms have reversed a long-term downward trend in coverage causing a significant amount of churning in the data.

The contributions of this paper to the literature are twofold: firstly this is the first use of dynamic models in the study of Australian health insurance. The use of panel data allows us to specify and estimate models with both state dependence and unobserved individual-specific effects. The modelling of dynamics has been found to be important in the small but growing literature on the dynamics of private health insurance based on overseas data (Finn and Harmon, 2006; Fairlie and London, 2006). Insurance is expected to exhibit substantial inertia due to inertia in health status and to the costs of moving in and out of cover (for example waiting times, complexities in learning about available contracts, etc.) In the US context, we would also expect substantial state dependence stemming from insurance tied to employment contracts. This study also contributes to the international literature on health insurance by focussing on a specific group of individuals namely young women. Although most insurance purchasers are expected to be older based on models of risk aversion (that is believed to increase with age) and adverse selection (average risk increases with age), health insurance is also in great demand among young families around the time of pregnancy and birth of children. This aspect of health insurance has to date been mostly ignored in the literature; specifically, the treatment of this issue has been restricted to the inclusion of variables for presence and number of children in insurance models for the population as a whole. This is despite the fact that this group exhibits substantial movement in and out of cover from year to year. An analysis of the behaviour of this age group can further our understanding of the overall movements in private health insurance coverage. Unlike churning associated with other motivations, it is potentially easier to identify such behaviour when it is due to issues surrounding family formation and support.

The data set used in this paper, the Australian Longitudinal Study on Women's Health (henceforth ALSWH), allows us to focus on a large cohort of young women below the age of 30. This survey includes information on aspirations regarding children and actual numbers of children. We can thus distinguish between young women who have or have not started their families, those who have reached their desired number of children and those intending to have additional children.

Previous studies of private health insurance (henceforth PHI) have dealt with several issues related to health, health care and the intrinsic characteristics of insurance. Early interest focussed on moral hazard or the effect of coverage on the demand for health care (Manning et al, 1987 and Cameron et al, 1998). Recently, other aspects of the demand for health insurance have been researched such as the magnitude of adverse selection and the heterogeneity in risk aversion (Finkelstein and McGarry, 2003 and Doiron, Jones and Savage, 2008). Other determinants of health insurance have also attracted interest for example the degree of income elasticity (Perry and Rosen, 2001 and Propper, 1989) and at least for the US the general relationship between health insurance, employment and labour mobility (Gilleskie and Byron, 2002 and Gruber and

Madrian, 2001). Finally, recent policy reforms aimed at increasing take-up of insurance have sparked renewed interest in PHI in Australia (Palangkaraya and Yong, 2005 and Ellis and Savage, 2008).

Age and sex are found with income to be amongst the more important variables determining health insurance. The importance of demographics is not surprising since health concerns vary considerably with these traits. Being female is associated with higher coverage (Cutler and Fiebig, 2005 and Cardon and Hendel, 2001). This is usually interpreted as women having greater expected utilisation through child-bearing or greater risk aversion possibly stemming from having preferences exhibiting greater intertemporal substitution. The presence of dependent children is found to have ambiguous effects. In some studies, having children is associated with greater coverage, as is being partnered with or without dependent children (Doiron, Jones and Savage, 2008 and Barrett and Conlon, 2003). In other studies, the presence of children has insignificant or even negative effects on cover (Propper, 1989 and Hopkins and Kidd, 1996). Families with children are also more likely to respond to incentives in regards to PHI (Ellis and Savage 2008.)

The ambiguous empirical impact of variables representing presence and number of children across studies is not surprising since theoretical considerations lead to conflicting effects. When health insurance is used mostly for hospital cover as in the Australian context, the presence of children can reduce cover if the treatment of children in public hospitals is considered to be better or at least no worse than private treatment. The lack of clear benefits accompanied by the drop of disposable income and a possibly increased saving motive could yield a negative effect of children on PHI. On the other hand, risk is increased with the addition of children to the household and possibly risk aversion as well through for example increased forward-looking behaviour. There are also benefits of PHI for the treatment of children in the Australian context such as better control over the choice of doctor, more continuity of care, and less wait time for certain consultations and procedures.

What is perhaps less ambiguous is the presence of net benefits surrounding pregnancy and birth: greater choice of doctor and continuity of care; less wait time for procedures; access to private hospitals/wards (i.e. more comfort and privacy); better access to certain procedures (e.g. caesarian sections); and insurance cover for assisted reproduction technology. Hence we expect some young women to demand private health insurance in order to deal with the period of pregnancy and the birth of the child. At this point it is useful to point out that since most insurance plans include a wait time of up to a year, purchase of insurance must be made prior to the beginning of pregnancy if the purpose of the insurance is health care surrounding the pregnancy or birth of the child.

By focussing on a fairly homogeneous group, women under the age of thirty, we isolate the impacts of children from spurious effects due to the correlation of the presence of children with the age and health of families (Hopkins and Kidd, 1996). Furthermore, by using information on desired and actual number of children we can distinguish complete versus incomplete families. As discussed above, this is important in distinguishing the different purposes of health insurance. If the reason for cover is pregnancy and care around the birth of a child, families where the actual number of children equals or exceeds the desired number will tend to drop out of coverage. If the purpose of cover is to improve care for young children, then we should not observe any reduction in cover once the desired family size is attained.

Information on desired and actual fertility outcomes has not been used extensively in economic contexts. Recent exceptions include Adsera (2005) who looks at labour market effects on desired number of children using Spanish data and Yu (2006) who looks at fertility and education profiles in Australian panel data.

Although inertia is believed to be important in models of health and health-related decisions, there are still very few studies based on dynamic models of health, health care and insurance (Fairlie and London, 2006; Finn and Harmon, 2006; Propper, 2000 and Contoyannis, Jones and Rice, 2004). This is mostly due to the lack of appropriate data. In this paper, we use panel data and model dynamic effects through state dependence while controlling for correlation due to the presence of unobserved and time-invariant individual effects. Specifically, we estimate discrete dynamic models with correlated random effects such as those used in Contoyannis, Jones and Rice (2004). We also estimate specifications in which attrition is modelled through the use of inverse weighted probabilities (Wooldridge, 2002 and Contoyannis, Jones and Rice, 2004). Since we are dealing with a group of young individuals, the effects of initial conditions is expected to be relatively small compared to samples representative of the population as a whole. Nevertheless, we also estimate specifications with initial conditions modelled as in Contoyannis, Jones and Rice (2004) and Arulampalam et al (2000) based on the approach of Wooldridge (2005).

2 REGULATORY ENVIRONMENT

In the OECD terminology (Colombo and Tapay 2004), Australia is classified as having duplicate, complementary and supplementary private health insurance, since some services are covered in both the public and private sectors (duplicate), coverage of copayments is allowed (complimentary), and some additional services are covered only by private insurance (supplementary). Importantly, Australia has a public system of universal coverage, Medicare, whereby anyone can be admitted as a public patient and be fully covered by Medicare, whether the person is covered by private health insurance or not. The main benefit of the private system is choice, specifically choice of doctor and avoidance of long queues for surgery. Public patients can face long waiting lists for some surgeries and must take the first available doctor. Private health insurance also gives access to treatment in private hospitals or private treatment in public hospitals. This usually means greater privacy and comfort. It is possible for individuals to self-insure and to pay for treatment in private hospitals. However, given the implicit subsidies of private health insurance through the tax system, this would only be rational among the very wealthy.

It is also possible to get private cover for ancillary services such as dental care, allied health services and complementary care. A small number of households have cover for these services only, without hospital cover (3.5% in 2004; ABS, 2006). In this paper private health insurance status indicates cover for private treatment in hospital and individuals with ancillary cover only are treated as uninsured.

With the exception of a few restricted membership funds, insurers must accept all purchasers for each policy type offered. Premiums are strictly regulated and community rating implies that insurers do not have the ability to discriminate in pricing based on sex, prior history of illness or any other risk/utilization characteristic. Only discrimination based on age is allowable (and is in fact mandated) for new enrolees post year 2000. Specifically, regulations stipulate increases in premiums of 2%, up to a maximum of 70%, for every year of age over 30 for those who joined after July 2000. Annual premiums can vary depending upon the fund, the extent of cover, the front-end deductible and the state of residence. All increases in premiums must have government approval and applications for increases are considered once each year.

The benefits of PHI related to pregnancies and children are described in the previous section of the paper. We provide in this section a few additional facts related to one source of demand namely treatment surrounding caesarean births. According to Laws et al. (2007), in Australia in 2005, 31.1% of births in hospitals were through caesarean sections. The rate was over 40.3% in private hospitals compared to 27.1% in public hospitals. Hence this type of PHI benefit can affect a substantial component of the population of women who are planning to have children.

We expect community rating to discourage low risk consumers from purchasing insurance, as contracts will be overpriced given their expected usage (Rothschild and Stiglitz, 1976). Various studies (Vaithianathan, 2004 and Gans and King, 2003) have looked at the importance of adverse selection on the Australian market. Many commentators see community rating and the resulting adverse selection as the main reason for observed steadily declining rates of coverage as healthy people dropped out because premiums did not match their risk profile. The membership level in private health insurance reached a low of close to 30% in 1998.

Beginning in 1997, the government introduced a series of incentives to increase private health insurance coverage in order to reverse the steady decline in membership and to take pressure off the public hospital system. The reforms included a 30% subsidy to insurance premiums, a tax surcharge for the high-income uninsured and Lifetime Health Cover. The last policy regulated the age-premium relationship described above. The proportion of individuals covered by private insurance increased to about 45% in late 2000 (see Salale (2006) for further details). Lifetime Health Cover appears to have been the most effective in terms of increasing aggregate insurance levels (Butler, 2001; Palangkaraya and Yong, 2005 and Ellis and Savage, 2008).

The impact of the reforms was to reduce the price of insurance for high-income groups and for individuals over the age of 30. Since we are dealing with a relatively young cohort (under 30) we do not expect a large effect of these reforms on our analysis sample. Most of the effects will be captured by the elasticity of demand with respect to income and a general specification of income is adopted. However, studies analysing the effects of the reforms also found evidence of an effect from the advertising campaign (Ellis and Savage, 2008) which suggests that young people may have been unduly induced to buy PHI despite their age. A year dummy is used to capture the possibility of advertising effects in 2000.

3 DATA

The ALSWH looks at the health and lifestyles of a representative sample of the Australian female population. The ALSWH is a 20 year longitudinal study of Australian women funded by the Australian Government Department of Health and Ageing and run by the University of Newcastle and the University of Queensland. The baseline year for the panel is 1996. The ALSWH contains three age cohorts: young - aged 18-23 in 1996; mid-age - aged 45-50 in 1996; older- aged 70-75 in 1996. Participants were randomly selected from the national Medicare database. Responses are collected through self-completion questionnaires approximately once every three years. The study design included deliberate over sampling of women living in rural and remote areas in order to try and capture the heterogeneity of health service experiences of women in these areas (Lee et al., 2005). All models in the paper include controls for the level of urbanization.

This study uses the three available waves of the young cohort - 1996 (aged 18-23), 2000 (aged 22-27) and 2003 (aged 25-30). In the initial wave, 14,779 individuals completed the questionnaire and 14247 gave consent to be part of the longitudinal study. Of these, 9,688 responded in wave 2 and 9074 completed the wave 3 questionnaire. For reasons detailed below, a balanced sample is used in this paper. The individuals who participated in all 3 waves number 7,790. After deleting respondents with missing information for the main variables (PHI coverage and aspirations regarding children) we have 7,360 individuals or 22,080 observations. This forms our analysis sample. Attrition is substantial and is mainly attributable to the high levels of mobility (changes of location, surname, etc) that characterise the younger generations (Lee et al., 2005). Estimation models that include corrections for attrition are conducted as part of the sensitivity analysis.

Table 1 describes the extent and persistence in coverage across the three waves for the analysis sample. The women have PHI in all three waves (the 'continuous cover' group 1) and those who are not covered in any of the waves (the 'never covered' group 4) together account for 63 percent of the sample. Clearly there is a large amount of persistence in these data which motivates the use of a dynamic model. Table 1 also shows that there is a substantial amount of churning across waves. For example, from wave 2 to wave 3 there are a large number of joiners (1,196 or 16%) and leavers (281 or 4%).

*****************Table 1 **************

The proportion of women who drop insurance cover at any wave, 16.8%, is smaller than the proportion who join at any wave, 26.4% (these are not mutually exclusive). This is not surprising as we expect this group to be mostly purchasing insurance since the level of coverage is fairly low for the younger aged groups. Finally, Table 1 shows that 34% of the sample has PHI cover in wave 1. This proportion is higher than that reported in the National Health Survey (NHS) 2001 for the comparable age group. Specifically, 30% of the 18-23 year olds reported being covered in the NHS. Among the unbalanced sample who consented to the longitudinal survey (14,070 observations[1]), the rate of coverage in wave 1 is 30.7% which is close to the NHS estimate. This suggests that the discrepancy in rates of PHI cover is due to our use of the balanced sample. Although the dynamic model conditions on PHI cover in wave 1 (the initial conditions), the discrepancy in rates of cover raises concerns with the representativeness of the analysis sample. A comparison of socio-economic variables with the NHS sample suggests that there is a slight over-representation of high socio-economic status individuals in our analysis sample. Models are estimated with corrections for selection from the unbalanced sample using these socio-economic indicators to reweigh the sample. More details are provided below.

A main contribution of this paper is the identification of demand for health insurance to cover future pregnancies, childbirths and care for children. This requires information on planned additional children. Information on the desired number of children is collected in every wave of the survey and is based on the question: *When you are 35, how many children would you like to have?* Since women in the sample are still under 35 years of age in wave 3, this question

[1] This excludes 177 observations with missing values for the insurance question.

provides the desired information. Details on actual children are not asked in the first 2 waves of the survey. We use the information on number of children provided in wave 3 along with their birth dates to construct children variables for the first 2 waves. This means that in order to use the information on actual children we restrict the analysis to the balanced sample. In addition, the top coding of desired number of children at 3 or more means that comparisons between actual and desired numbers cannot be made for those women with both 3 or more actual children, and 3 or more desired children. These women constitute a small fraction of the balanced sample (2% of observations). We include these observations in the analysis and add a dummy variable which equals 1 when the comparison between desired and actual children cannot be determined.

*****************Table 2 **************

Table 2 provides sample sizes for households depending on their actual and desired numbers of children.[2] The first column of numbers refers to respondents that report a higher desired number of children than their actual number; hence they are planning to have more children in the future. The second column of figures refers to households where the desired number of children has been achieved. Since the question on aspirations refers to the desired number of children at age 35, we may be underestimating the number of households planning to have more children to the extent that women may plan to have children past the age of 35. This would result in an underestimation of the effects of unmet children aspirations as a source of demand for health insurance. Indeterminate refers to those households where the actual and desired numbers of children are both "3 or more". Of the households without children, around 10% say they do not want children (by the age of 35). In wave 2, 24% of households with children do not want more and this proportion increases to 29% in wave 3. This is consistent with the fact that the proportion of households with children increased from 20% in wave 2 to 31% in wave 3.

In order to distinguish between demand for insurance for use during pregnancy and the birth of a child and demand due to the care of children, the timing of the birth of children is key information.[3] Since the insurance status is observed at the time of the surveys, a natural distinction in the timing of childbirths is by wave. We distinguish households who experience the birth(s) of a child (multiple children) during the recent period i.e. between the current and previous survey questionnaires, and households who have children in the period before the previous survey, referred to as the early period. For each wave, households with children are in one of three groups: early period only, recent period only, and both periods. For example, with wave 3 observations, "recent period only" indicates households who had their first child born between waves 2 and 3,"early period only" indicates households that do not have additional children in wave 3 (compared to wave 2) while "both periods" indicates households in which children are born before wave 2 and between waves 2 and 3. The breakdown of household observations according to the timing of actual children is provided at the bottom of Table 2. 2,004 or 53% of 3,750 households with children, had their first child between the current and the last waves of the survey. Of these 388 or 19% want more children in the future.

*****************Table 3 **************

[2] We refer to households as units of analysis despite the fact that desired refers to the women's stated preferences and this could be different from some definitions of the preferred number of children by the household.

[3] Information on pregnancies is available in the survey and could be used to indicate timing of planned children. Unfortunately the sample sizes are small and it is not clear how women treat unborn children when answering questions on actual and desired numbers of children.

Table 3 provides sample proportions of households with PHI cover according to desired and actual children categories. The incidence of cover varies a lot across groups, from a low of 21% for households with older children who have finished their family to 41% for households who have no children but who are planning to have some. The differences in PHI cover across households based on their aspirations can be seen in the last column. For households without children, those who are planning to have children are 7.5 percentage points more likely to be insured than households without children who are not planning to have children (at least in the near future).

Several other factors are included in the model of insurance. These have been identified as important determinants of insurance demand in previous studies. The additional regressors are briefly described next and a full list with definitions is given in Appendix 1.

It is shown in many previous studies that private health insurance holders have a higher income distribution than those without cover (Perry and Rosen, 2001). This also holds for the data set under study (Salale, 2006). Insurance is generally found to be a normal good and the current government incentive structure in Australia strengthens this relationship. The ALSWH collects information on household weekly income. Income is measured in 8 categories with the top category corresponding to $A1,500 per week or an annual income of $A78,000. Dummy variables are used to capture variations in the income categories. As discussed above, we expect the income effect to vary depending on actual and desired children due to changes in equivalent household income, savings motive and risk aversion. To account for this, the insurance model is also estimated on subsets of the sample corresponding roughly to the top and bottom quintiles of the income distribution. Personal and household income categories are highly correlated and we do not include both series of dummy variables. Since health insurance is a household good, household income is used in the specifications presented below. A dummy variable is added to indicate whether the respondent lives alone.[4]

Unfortunately, income was not asked in the first wave of the survey. This does not affect most specifications of the dynamic model as the likelihood is conditional on the initial conditions and the first wave information on variables other than insurance is not used. The missing wave 1 income information does, however, affect the attrition correction for the pooled probit specifications. A dummy for missing household income is also included for missing values in waves 2 and 3 (21% of observations).

The existence of adverse selection in an insurance industry generates a positive correlation between risk and insurance choice (Chiappori and Salanie, 2000 and Cutler and Zeckhauser, 1999). However, recent studies show that this relationship may be difficult to capture due to heterogeneity in risk aversion (Finkelstein and McGarry, 2003). Doiron, Jones and Savage (2008) find a negative correlation between risk (as measured by self-assessed health) and the choice to insure in cross-sectional Australian data. They also find that the positive relationship predicted by adverse selection is recovered when measuring risk with chronic conditions. Based on risk-related behaviours, they find evidence that the counterintuitive negative correlation can be explained by the effect of heterogeneous risk aversion as measured by risk-related behaviours.

[4] Sensitivity analysis where personal income is used instead of household income yields very similar results to those presented below. These additional estimates are available from the authors.

Self assessed health is available in all waves of the survey and is used in the estimation of the insurance model. There is also information on long-term conditions but the number of women with any type of long term condition (excluding asthma) is never above 3 percent of the sample in each wave (Salale, 2006). Hence the scope for an analysis of the effects of chronic conditions on insurance among this age group is limited. In some specifications a variable measuring the number of reported long-term conditions was included. The coefficient on this variable was insignificant and did not affect results on the children variables; hence, the variable is excluded from the main model. Variables measuring risk-related behaviours (smoking, BMI, alcohol consumption) are included.[5]

The ALSWH data set also measures women's perception of access to health care services. For the purposes of this study perception of access to hospital care is an important and usually unmeasured variable. Using a bivariate analysis of the relationship between income and self-reported access, Salale (2006) shows access to hospitals may be inequitable. This follows from women with lower incomes and no hospital insurance reporting lower access levels, while women with higher incomes and hospital insurance are more likely to report the opposite. Access to care relates to either public or private care. Since private hospitals are more unevenly distributed across geographical areas, this variable is an important indicator of the usefulness of private insurance and is included in the analysis.[6]

Marital status is included with a series of dummy variables: married (26% of observations) or defacto (18% of the sample), or other (56% of observations) which includes single, separated, divorced or widowed. Variation in geographical location is captured by categorical variables representing the degree of remoteness. State dummies proved to be insignificant except for Queensland and a dummy variable indicating respondents living in this state is retained in the model. Additional explanatory variables measuring socio-economic status (education, employment status), the ownership of a health card[7] and country of birth are included. Appendix 2 provides means of all explanatory variables (other than the children indicators) by wave.

As the panel is a short, wide panel, a set of time dummies is also recommended (Wooldridge 2002) therefore a dummy for the year 2000 is included. As described above, this dummy variable will also capture advertising effects surrounding the life-time cover reform of 2000. Dummy variables are used for missing exogenous variables; in most cases these represent less than 2% of the sample. Please refer to Appendix Table 2 for more details.

[5] Categorical variables controlling for the level of exercise were included in some specifications and since the coefficients were small and insignificant the variables were later dropped from the main model.
[6] This information is only available in waves 2 and 3.
[7] There is a range of concession cards that entitle eligible Australians to higher government subsidies for some medical services and products. Eligibility is primarily linked to whether an individual or household is a recipient of qualifying government benefits such as the age or disability allowance or are in receipt of specified allowance and have sufficiently low income. War veterans may be eligible to one of the three types of Department of Veterans Affairs (DVA) cards. Every concession card lowers the copayment for prescription medicines. Many GPs and some specialists do not charge copayments for consultations with concession cardholders. DVA gold card holders do not face any copayments for health care services and DVA white cardholders have similar entitlements but only for services related to a specified war-related condition.

4 MODEL SPECIFICATION

Dynamics

Most estimates of the demand for insurance are based on cross-section data although one would expect the demand for insurance to have an important dynamic component. As summed up by Propper (1989, p791), "[…] captivity and the effect of past purchase would be promising avenues to explore". Individuals may not revisit the decision to purchase health insurance every year and there are costs associated with moving in and out of the market (e.g. the age penalty in the premiums legislated by government, and waiting times for claims). Panel or retrospective data are needed to measure these effects and until recently such data were still rare in this context. The use of dynamic models to study health insurance and indeed for most other health-related variables is complicated by the need to use nonlinear models able to accommodate discrete endogenous variables. We follow recent approaches in modelling the demand for insurance as a dynamic probit regression assuming a first-order Markov process for the dynamic effect (Contoyannis, Jones and Rice, 2004; Arulampalam, Booth and Taylor, 2000 and Erdem and Sun, 2001) such that only the most recent information is relevant in the choice decision.

Specifically, the demand for health insurance can be written as:

$$(1)\quad PHI_{it}^{*} = X_{it}\beta + \gamma PHI_{i,t-1} + Z_{it} \times PHI_{i,t-1}\theta + u_{it} \qquad i = 1,\ldots,N;\ t = 1,\ldots,T$$

$$PHI_{it} = \begin{cases} 1 & \text{if } PHI_{it}^{*} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

where PHI^* is a latent variable measuring the net benefits of private health insurance; PHI is the observed insurance status (insurance is purchased only when the net benefits of insurance are positive); the vector X_{it} represents observable, exogenous variables that affect insurance choice (the vector does not include a constant term); $PHI_{i,t-1}$ is the lag of the dependent variable capturing the effect of state dependence; the vector Z_{it} includes a subset of the exogenous variables that are interacted with the lagged insurance status; N indicates the number of individuals in the panel; T is the number of periods in the panel which is constant across individuals in the case of the balanced sample; u_{it} is a composite error term which is explained further below.

The use of panel data allows us to control for the presence of individual specific, unobserved (or partially observed) and time-invariant effects. In dynamic models, these can be crucial since their omission will lead to overestimation of the state dependence; if individual effects are ignored, the dynamic effects remain as the only source of correlation across time periods specified in the model (in the absence of serial correlation). We follow recent empirical papers in specifying random effects where individual-specific effects consist of a time-invariant random component independent of the explanatory variables. In some specifications, the individual-specific effects are correlated in the sense that in addition to the random component, there is a deterministic component which is a linear function of explanatory variables (or functions thereof). This is the approach used in Contoyannis, Jones and Rice (2004) and Arulampalam, Booth and Taylor (2000) and is based on work by Chamberlain (1984) and Mundlak (1978).

Specifically, the composite error u_{it} is specified as:

$$(2)\quad u_{it} = a_i + \varepsilon_{it}$$

where a_i is an individual-specific, time-constant random term drawn from a normal distribution and ε_{it} is a time-varying, individual specific random error. Under the probit specification, ε_{it} is normally distributed with a variance normalised to 1. In some specifications, following Mundlak (1978), the individual-specific effect a_i, and the exogenous variables are related in a linear manner,

(3) $a_i = \alpha_0 + \alpha_1 \bar{X}_i + \eta_i$

where $\bar{\mathbf{X}}_i$ is a vector of means of any time-varying regressors and η_i is a normally distributed random error term assumed to be independent of X.

Obtaining consistent estimates in a dynamic probit is complicated by the inclusion of lagged dependent variables as regressors. Specifically, assumptions of independence between the individual specific random component (a_i) and the explanatory variables will not hold unless the initial time period of the dynamic process is observed or the process is in long-term equilibrium with time-invariant distributional properties (Heckman, 1981). With a short panel and individuals already having made purchasing decisions in the first wave, neither of these assumptions is likely to hold. Hence we need to model or condition on initial conditions.

Following Wooldridge (2005), the distribution of the unobserved effects is modelled as conditional on the initial value of the dynamic process in addition to the means of the time-varying regressors:

(4) $a_i = \alpha_0 + \alpha_1 \bar{X}_i + \alpha_2 PHI_{i1} + \eta_i$

and $\eta_i \mid PHI_{i1}, \bar{X}_i \sim N(0, \sigma_\eta^2)$ with η_i independent of $X_{it}, \bar{X}_i$ and PHI_{i1}.

Wooldridge (2005) shows that this model can be consistently estimated with a random effects probit where $(X_{it}, PHI_{i,t-1}, Z_{it} \times PHI_{i,t-1}, PHI_{i1}, \bar{X}_i)$ is used as the set of explanatory variables. The resulting conditional maximum likelihood estimates are $\sqrt{N}$ consistent and asymptotically normal under standard regularity conditions, assuming that the moments exist and are finite (Wooldridge, 2002).

This specification allows for two types of persistence across time: state dependence and the correlation due to individual-specific time-invariant effects. Another source of correlation could be due to serial correlation in the error term ε_{it}. For example, Keane (1997) investigates the effect of auto-correlated time-varying errors in multi-nominal, multi-period probit models using marketing data. Unfortunately, testing for correlation in the time-varying errors is not feasible with the data set used in this study due to a very short panel. However, it is important to note that with further waves, testing and possible correction for auto-correlated errors would be possible.

Following Wooldridge (2005) and Contoyannis, Jones and Rice (2004) average partial effects (APE's) are calculated by integrating over the distribution of individual-specific effects. A consistent estimator of $E(PHI_t \mid X_t, PHI_{t-1}, Z_t \times PHI_{t-1}, PHI_1, \bar{X})$ for t=2 and 3 is given by:

(5) $$N^{-1}\sum_{i=1}^{N}\Phi(X_{it}\beta_{\eta} + \gamma_{\eta} PHI_{i,t-1} + Z_{it} \times PHI_{i,t-1}\theta_{\eta} + \alpha_{0\eta} + \alpha_{1\eta}\overline{X}_i + \alpha_{2\eta} PHI_{i1})$$

The η subscript indicates multiplication by $(1+\sigma_{\eta}^{2})^{-1/2}$ where the consistent estimate of σ_{η} from the random effects probit is substituted. Using the predicted probabilities, one can then derive "marginal" effects: the derivative with respect to a variable in the case of a continuous variable or shifts in the predicted probabilities in the case of categorical variables.

Selection

Inverse probability weighting (IPW) is used to account for the possibly non-random attrition in the panel discussed in the previous section (Wooldridge, 2002). Assuming the sample at wave one is random, an ideal panel would include this 'full' sample observed at all waves. Instead, observations are only included if the respondents fill out all 3 questionnaires (and answer questions on PHI and aspirations for children). Denote by s_{it} = 1, participation in the analysis sample. We assume, conditional on all observables in the first time period, that (X_{it}, PHI_{it}) are independent of the response s_{it} such that

(6) $$P(s_{it} = 1 \mid X_{it}, PHI_{it}, PHI_{i,t-1}, PHI_{i1}, S_{i1}) = P(s_{it} = 1 \mid S_{i1}), t = 2,3$$

where the vector S_{i1} includes any information collected on the full sample at the first wave. (S_{i1} can include all variables in X as well as others; in particular it can include variables believed to be correlated with PHI.)

To implement IPW, two steps are used. Firstly, a binary probit is estimated measuring the probability of an individual being in the balanced sample, and a fitted probability is calculated for each respondent. Secondly, the pooled probit is weighted by the inverse predicted probability so that a greater weighting is placed on respondents with a lower response rate. Wooldridge (2002) shows that IPW produces consistent, $\sqrt{N}$ -asymptotically normal estimators in models where the likelihood can be written as a sum of contributions across all observations. This is not the case when individual-specific random effects are present. The correction for attrition is performed on pooled probit models only. (This is also the approach followed in Contoyannis, Jones and Rice, 2004.)

In order to model the attrition process (and the correlation with insurance demand) we use a large number of indicators of mental, psychological and physical well-being all measured in the first wave of the survey. Also, all available exogenous variables listed in Appendix Table 2 plus the children variables, all measured at t=1, are included in the first stage probit. (Details are available from the authors.)

Specification of effects of desired and actual children

In order to motivate the specification of the children variables, it is useful to list the main roles of desired and actual children in the demand for health insurance. Consider a respondent who is observed with private health insurance cover in wave 3, this woman could:

- intend to become pregnant and have kids in the future; this is referred to as the effect of *aspirations* for children and we expect this to be positive;
- have had kids between waves 2 and 3, have purchased insurance for the pregnancy and childbirth and has not yet dropped cover;

- have children and want insurance for the health care of the children.

In addition, the presence of children is expected to affect the demand for insurance through an income effect (a reduction in purchasing power and/or an increased savings motive) and a shift in preferences towards risk. Finally, there could be learning effects in the sense that the desire for insurance around pregnancy and childbirth could be different in the case of the first child relative to latter children.

A comparison of the demand for PHI by households who have achieved their desired number of children with households who desire additional children provides information on the effect of aspirations for children. This is done separately for households with and without children, and for those with children, families with recent children are separated from those who have not had children since the last survey.

As described above, the effects of the *presence* of actual children can be complex and the direction of the overall effect is ambiguous. We compute differences in predicted cover between households who had children recently (i.e. between the current and the previous wave) and those who did not. These effects are computed separately for households who have achieved the desired size versus those who plan to have more children and for households who had children before the last survey questionnaire (i.e. households with children in the early period).

If insurance is wanted mostly for the period of pregnancy and the birth of a child rather than for the care of young children, then households with recent children should have a higher incidence of cover than households with older children. We refer to this as the *timing* effect and it is calculated as the difference in cover among households with children in the recent period only and families who had children in the earlier period only. This difference is computed separately for those households who plan to have more children in the future and those who don't.

In order to capture an income effect, we look at families who do not want more children (there is no demand for future pregnancies and births) who did not have children in the recent period (there is no residual cover from recent childbirths) and take differences in predicted probabilities of cover between households with children and those without children. This is referred to as an *income* effect and we expect it to be negative if the impact from the reduction in disposable income (and possibly greater savings motive) is greater than any effect from increased risk aversion.

Finally, in order to isolate the effect of learning the value of insurance through previous childbirths, we look at households who desire more children and compare those who had children in the earlier period from those who do not have children. Focusing on those with children in the earlier period excludes residual cover from recent childbirths. This total impact will combine learning and income effects so we subtract the income effect to get the *learning* effect.

5 RESULTS

Specifications
The insurance model represented by equation (1) is estimated with pooled probits, random effects probits and inverse probability weighted (IPW) probits on observations from waves 2 and 3. Wave 1 information is used in specifying the initial conditions for PHI, the lagged PHI status

for wave 2 observations and in estimating the probabilities of attrition for the IPW model. Various parametrizations of the regressors are used. The main model on which most of the results presented below are based, includes the following explanatory variables: actual and desired children variables described above, lagged PHI cover, initial PHI cover, marital status, household income, single income household indicator, left home since last wave indicator, age, self-assessed health, health card holder indicator, Queensland indicator, rural and remote area indicators, education, risk-related variables (smoking, alcohol, body mass index), country of birth, study and work status, access to health services indicators. Other variables were also included in earlier models but were not important quantitatively or statistically and were dropped from the analysis.[8] The use of personal income instead of household income did not affect any of the results presented below. Given the coarseness of the information on income, household and personal income categories are too closely correlated to both be used in the model.

Table 4 presents results for the pooled probits (the main model), random effects probits and IPW models. Included are the predicted probabilities of insurance cover averaged over the sample for the various household types based on the desired and actual children categories. Only the predicted probabilities for the pooled probits are presented due to the similarity of the results across specifications. Also presented are the average marginal effects (the shifts in the predicted probabilities) for the various types of households averaged across the sample. For the random effects model, the marginal effects are averaged over the distribution of individual-specific unobserved effects (i.e. average partial effects). Standard errors from a bootstrap estimator with 200 iterations are provided for the marginal effects. (Standard errors on the predicted probabilities are all relatively small and are not given to save on space.)

The random effects specification does not include any correlated effects. Specifications of the random effects probits with the mean household income as explanatory variable for the random effects yield virtually identical results to those shown in the table. Random effects specifications yield very similar results to the pooled probit model although formal tests lead to rejections of the pooled models in favour of the random effects specification. The IPW specification also yields similar results. Given the relative simplicity and speed of estimation of the pooled probits, many of the results presented below are based on the pooled probit specifications.

*****************Table 4 **************

Table 5 presents similar results for specifications which include interactions of the children variables with the lagged insurance cover. This allows us to distinguish between effects of desired and actual children on joining and dropping insurance plans. Table 6 presents results of models estimated on subsets of the sample based on the household income category.

*****************Table 5 **************

*****************Table 6 **************

Effects of actual and desired children variables

Results are similar across the three specifications and the discussion will focus on the numerical results for the pooled probit. As expected the aspiration effects are uniformly positive. Women

[8] These include: a full set of state dummies, exercise indicators, the number of long-term conditions, and lagged health status. The results presented below were only affected at the third digit by the inclusion of these variables.

who desire additional children in the future are more likely to have insurance cover than comparable women who have finished their family. Wanting more children in the future raises your probability of insurance cover by around 3 percentage points for those without recent children and by closer to 5 percentage points for those who had children born between the last 2 waves.

The presence of children (having children between the last 2 waves) has different effects depending on whether these recent children were the first children in the family or not. The effects are negative when these recent children were the first children and positive when not. We checked to see if this difference was capturing differences in the family size. Specifically, the total number of children was added as a control and the models were re-estimated. The impact on the marginal effects presented in Table 4 was at the third digit only. Hence it is the recentness of the children that is causing the difference.

Timing effects are negative; i.e. having recent (rather than older) children means that households are more likely to have cover. These effects are large, 7.4 percentage points for those households who desire additional children in the future and 5.6 percentage points for comparable women who have finished their family. This points to households wanting PHI for pregnancy and birth of the child rather that care of young children. Note that this is not due to persistence in cover since the model includes lagged insurance status.

As expected, income effects of children are negative and quite large, estimated to be 9.5 percentage points. Learning effects are positive which is also as expected, but this estimate is very small.

Other variables
Table 7 presents a selection of other results. Average marginal effects are provided but only for key variables and those with significant effects. *p*-values for the underlying coefficient estimates are also provided. These are calculated using standard errors that are robust to correlation across waves by individuals.

Important factors determining the demand for insurance include marital status, perceived access to hospitals, location and country of birth. In all specifications, significance tests on the lagged dependent variable yield strong rejections of the static specification in favour of the dynamic model. The marginal effects for self-assessed health are small and there is no evidence of a gradient. For this age group, there is little evidence of either adverse or favourable selection.

******************Table 7** **************

6 CONCLUSION

In this paper, we use panel data and model dynamic effects through state dependence while controlling for correlation due to the presence of unobserved and time-invariant individual effects. We also estimate specifications in which attrition is modelled through the use of inverse weighted probabilities. Since we are dealing with a group of young individuals, the effects of initial conditions is expected to be relatively small compared to samples representative of the population as a whole. Nevertheless, we also estimate specifications with initial conditions modelled.

We find evidence of differential demand for private health insurance by young women based on actual and desired numbers of children. Women with and without children who desire more children are more likely to purchase insurance. Effects are quantitatively important. Also the effect is stronger for those with children. The different effects on joining and leaving cover show the importance of modelling dynamics in insurance.

REFERENCES

Arulampalam., W., Booth., A., Taylor., M., (2000) *Unemployment Persistence*. Oxford Economic Papers, 52 (1), 24-50

Australian Bureau of Statistics [ABS] (2006) *National Health Survey: Summary of results.* Canberra, Australia

Barrett, G., Conlon, R., (2003) *Adverse selection and the contraction in the market for private health insurance*: 1989-1995. The Economic Record, 79 (246), 279-296

Bell, S., (2002) *Data book for the 2000 Phase 2 survey of the young cohort (22-27 years) Australian Longitudinal Study on Women's Health.* The Research Centre for Gender and Health, University of Newcastle, NSW Australia. January, 2002

Butler, J.R.G., (2001) *Policy change and private health insurance: Did the cheapest policy do the trick?* NCEPH Working Paper Number 44. The Australian National University.

Cameron, A.C., Trivedi, P.K., Milne, F., Piggott, J., (1988) *A Microeconometric Model of the Demand for Health Care and Health Insurance in Australia*. Review of Economic Studies, 55 (1), 85-106

Cardon, J., Hendel, I., (2001) *Asymmetric information in health insurance: evidence from the National Medical Expenditure Survey*. RAND Journal of Economics, 32 (3), 408-427

Chamberlain, G., (1984) *Panel Data*. Z. Griliches and M. Intriligator (eds) Handbook of Econometrics, Amsterdam, North-Holland, 1247-1318

Chiappori, P.A, Salanie, B., (2000) *Testing for Asymmetric Information in Insurance Markets.* The Journal of Political Economy, 108 (1), 56-78

Colombo, F., Tapay, N. (2004) *Private health insurance in OECD countries: The benefits and costs for individuals and health systems*. OECD towards high-performing health systems (Chap. 6) Paris OECD.

Contoyannis, P., Jones, A., Rice, N., (2004) *The Dynamics of Health in the British Household Panel Survey*. Journal of Applied Econometrics, 19, 473-503

Cutler, D., Zeckhauser, R., (1999) *The Anatomy of Health Insurance*. NBER working paper 7176

Cutler, H., Fiebig, D.G., (2005) *The demand for private health insurance allowing for substitution*. Working paper, 16 September, 2005

Ellis R., Savage E. (2008) *Run for Cover Now or Later? The impact of premiums, threats and deadlines on supplementary private health insurance in Australia*. International Journal of Health Care Finance and Economics (accepted February, 2008).

Erdem, T., Sun, B., (2001) *Testing for choice dynamics in panel data*. Journal of Business and Economic Statistics, 19, 142-152

Doiron D., Jones G., Savage E. (2008) *Healthy, Wealthy and Insured? The Role of Self-Assessed Health in the Demand for Private Health Insurance*. Health Economics 17**:** 317-334.

Fairlie, R. and London, R. (2006) *An Analysis of the Dynamics of Health Insurance Coverage and Implications for Employer-Mandated Insurance*, Employment Policies Institute: Washington, D.C.

Finkelstein, A., McGarry, K., (2003) *Private Information and its Effect on Market Equilibrium: New Evidence from Long-Term Care Insurance.* NBER Working Paper 9957, September.

Finn, C. and Harmon, C. (2006) *A dynamic model of demand for private health insurance in Ireland*, IZA Discussion Paper No 2472, November.

Gans, J., King, S., (2003) *Anti-insurance: Analysing the Health Insurance System in Australia*. The Economic Record, 79 (247), 473-486

Gilleskie, D. and Byron, L. (2002) *The Impact of Employer-Provided Health Insurance on Transitions*. Journal of Human Resources 37(1):129-162.

Greene, W., (2003) *Econometric Analysis*, 5th Ed. Pearson Education, New Jersey, USA

Gruber, J. and Madrian, B. (1997) *Employment separation and health coverage*. Journal of Public Economics 66(3), 349-382.

Heckman, J.J., (1981) *The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process*, in C.F. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA, 114-78

Hopkins, S., Kidd, M.P., (1996) *The determinants of the demand for private health insurance under Medicare*. Applied economics, 28, 1623-1632

Keane, M., (1997) *Modelling Heterogeneity and State Dependence in Consumer Choice Behaviour*. Journal of Business and Economic Statistics, 15 (3), 310-327

Laws, P.J., Abeywardana, S., Walker, J., Sullivan, E.A., (2007) *Australia's mothers and babies 2005*. Perinatal Statistics Series no. 20 Cat. no. PER 40. Sydney: AIHW National Perinatal Statistics Unit.

Lee, C., Dobson, A., Brown, W., Bryson, L., Byles, J., Warner-Smith, P., Young, A., (2005). *Cohort Profile: The Australian Longitudinal Study on Women's Health*. International Journal of Epidemiology, 34, 987-991.

Manning, W., J. Newhouse, N. Duan, E. Keeler, A. Leibowitz, and M. Marquis. (1987) *Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment*. American Economic Review, 77, 251-277.

Mundlak, Y., (1978) *On the Pooling of Time Series and Cross Section Data*. Econometrica, 46, 69-85

Palangkaraya, A., Yong, J. (2005) *Effects of Recent Carrot-and-Stick Policy Initiatives on Private Health Insurance Coverage in Australia*. The Economic Record, 81(254) September, 262-72.

Perry CW, Rosen HS (2001) *The Self-Eemployed are Less Likely to Have Health Insurance Than Wage Earners. So What? NBER Working Paper 8316*. National Bureau of Economic Research: Cambridge.

Private Health Insurance Administration Council (PHIAC) (2007) *Quarterly statistics*. Available at http://www.phiac.gov.au/statistics/membershipcoverage/table1.htm

Propper, C., (1989) *An econometric analysis of the demand for private health insurance in England and Wales*. Applied Economics, 21, 777-792

Propper, C. (2000) *Demand for private health care in the UK*. Journal of Health Economics, 19, 855-876.

Rothschild, M., and Stiglitz, J., (1976) *Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information*. Quarterly Journal of Economics 90, 629-49

Salale, V., (2006) *Modelling Dynamic Choice: Private Health Insurance in Australia*. CHERE Research Report 24, CHERE, Sydney

Savage, E., and Wright, D., (2003) *Moral Hazard and adverse selection in Australian private hospitals: 1989-1990*. Journal of Health Economics 22, 331-359

Vaithianathan, R., (2004) *A Critique of the Private Health Insurance Regulations*. Australian Economic Review, 37 (3), 257-270

Wooldridge, J., (2002) *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, Cambridge Massachusetts

Wooldridge, J., (2003) *Introductory Econometrics: A Modern Approach, Second Edition.* Thomson South-Western, United States of America.

Wooldridge, J., (2005) *Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity*. Journal of Applied Econometrics, 20, 39-54

Yu, P. (2006) *Higher education, the bane of fertility? An investigation with the HILDA survey*. CEPR Discussion Paper#512, ANU.

Appendix 1: Variable definitions

Variable Name	Definition
Y2000	= 1 if year is 2000, 0 otherwise
Household income	
HINC1	= 1 if household has no income, 0 otherwise
HINC2	= 1 if household has $1-$199pw, 0 otherwise
HINC3	= 1 if household has $120-$299pw, 0 otherwise
HINC4	= 1 if household has $300-$499pw, 0 otherwise
HINC5	= 1 if household has $500-$699pw, 0 otherwise
HINC6	= 1 if household has $700-$999pw, 0 otherwise
HINC7	= 1 if household has $1000-$1499pw, 0 otherwise
HINC8	= 1 if household has $1500+ pw, 0 otherwise
HINCMISS	= 1 if Missing/Don't want to answer, 0 otherwise
HINCALONE	= 1 if respondent lives alone, 0 otherwise
AGE	= Age of respondent (years)
LVHOME	= 1 if respondent left family home in past 12 months, 0 otherwise
Employment Status	
STUDY	= 1 if studies but does not work, 0 otherwise
WORK	= 1 if works but does not study, 0 otherwise
WORKSTUDY	= 1 if works and studies, 0 otherwise
NOWORKSTUDY	= 1 if doesn't work or study, 0 otherwise
WORKMISS	= 1 if work/study status is missing, 0 otherwise
Qualifications	
TERTIARY	= 1 if completed tertiary qualifications, 0 otherwise
DIPLOMA	= 1 if completed a diploma, 0 otherwise
OTHERQUAL	= 1 if completed other qualifications, 0 otherwise
ONLYSCHOOL	= 1 if completed only primary or secondary schooling, 0 otherwise
QUALMISS	= 1 if qualification is missing, 0 otherwise
State	
QLD	= 1 if lives in Queensland, 0 otherwise
Area	
URBAN	= 1 if lives in a major urban area, 0 otherwise
RURAL	= 1 if lives in a rural area, 0 otherwise
REMOTE	= 1 if lives in remote area, 0 otherwise
AREAMISS	= 1 if area of residence is missing, 0 otherwise
Marital Status	
MARRIED	= 1 if married, 0 otherwise
DEFACTO	= 1 if in defacto relationship, 0 otherwise
OTHER	= 1 if single, separated, divorced or widowed, 0 otherwise
MARISTATMISS	= 1 if marital status is missing, 0 otherwise
HEALTHCARD	= 1 if respondent has health care card, 0 otherwise
Self-Assessed Health	
SAHSEX	= 1 if reports excellent health, 0 otherwise
SAHSVGOOD	= 1 if reports very good health, 0 otherwise
SAHSGOOD	= 1 if reports good health, 0 otherwise
SAHSFAIR	= 1 if reports fair health, 0 otherwise
SAHSPOOR	= 1 if reports poor health, 0 otherwise
SAHSMISS	= 1 if health status status is missing, 0 otherwise
Risk Indicators	
SMOKE	= 1 if smoke daily, 0 otherwise
ALCRISK	= 1 if alcohol consumption is risky or high by NHMRC guidelines, 0 otherwise
BMI	= body mass index (kilograms/(metres squared))
BMIMISS	= 1 if BMI is missing, 0 otherwise
Country of Birth	
AUS	= 1 if born in Australia, 0 otherwise
EUROPE	= 1 if born in Europe, 0 otherwise
ASIA	= 1 if born in Asia, 0 otherwise
OTHER	= 1 if born in other region, 0 otherwise
COBMISS	= 1 if country of birth is missing, 0 otherwise
Access to Hospitals	
ACCESS1	= 1 if rate access to a hospital to be 'Excellent' or 'Very Good', 0 otherwise
ACCESS2	= 1 if rate access to a hospital to be 'Good', 0 otherwise
ACCESS3	= 1 if rate access to a hospital to be 'Fair' or 'Poor', 0 otherwise
ACCESSMISS	= 1 if opinion on access to hospital is missing, 0 otherwise

Appendix 2: Sample means and standard deviations

Variable Name	Total (N=22080)		Wave 1 (N=7360)		Wave 2 (N=7360)		Wave 3 (N=7360)	
	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Y2000	0.333				1.000		0.000	
Household income:								
HINC1	0.003				0.004		0.004	
HINC2	0.003				0.006		0.003	
HINC3	0.015				0.029		0.017	
HINC4	0.039				0.071		0.047	
HINC5	0.074				0.123		0.098	
HINC6	0.105				0.159		0.155	
HINC7	0.134				0.202		0.201	
*HINC8**	0.155				0.184		0.280	
HINCMISS	0.139				0.222		0.195	
HINCALONE	0.036				0.051		0.058	
AGE	23.891	3.132	20.380	1.500	24.153	1.462	27.139	1.451
LVHOME	0.091		0.177		0.074		0.021	
Employment Status:								
STUDY	0.139		0.338		0.039		0.041	
WORK	0.371		0.406		0.151		0.555	
WORKSTUDY	0.265		0.133		0.454		0.208	
*NOWORKSTUDY**	0.211		0.107		0.353		0.172	
WORKMISS	0.014		0.016		0.003		0.024	
Qualifications:								
TERTIARY	0.327		0.132		0.402		0.447	
DIPLOMA	0.194		0.154		0.207		0.221	
OTHERQUAL	0.028		0.025		0.029		0.031	
*ONLYSCHOOL**	0.431		0.685		0.328		0.280	
QUALMISS	0.019		0.004		0.032		0.021	
STATE OF QLD	0.213		0.213		0.212		0.214	
Area:								
*URBAN**	0.561		0.548		0.541		0.595	
RURAL	0.392		0.412		0.417		0.347	
REMOTE	0.038		0.038		0.038		0.038	
AREAMISS	0.009		0.002		0.004		0.020	
Marital Status:								
MARRIED	0.256		0.091		0.255		0.422	
DEFACTO	0.177		0.125		0.210		0.196	
*OTHER**	0.563		0.780		0.530		0.378	
MARISTATMISS	0.004		0.005		0.004		0.003	
HEALTHCARD	0.121				0.198		0.164	
Self-Assessed Health:								
*SAHSEX**	0.133		0.131		0.133		0.136	
SAHSVGOOD	0.410		0.407		0.398		0.426	
SAHSGOOD	0.349		0.347		0.352		0.347	
SAHSFAIR	0.094		0.100		0.103		0.079	
SAHSPOOR	0.011		0.010		0.012		0.012	
SAHSMISS	0.003		0.006		0.001		0.001	
Risk Indicators:								
SMOKE	0.167		0.164		0.180		0.155	
ALCRISK	0.039		0.050		0.034		0.033	
BMI	20.279	9.470	19.481	8.844	20.583	9.286	20.773	10.183
BMIMISS	0.082		0.106		0.074		0.067	
Country of Birth:								
*AUS**	0.927		0.927		0.927		0.927	
EUROPE	0.009		0.009		0.009		0.009	
ASIA	0.017		0.017		0.017		0.017	
OTHER	0.041		0.041		0.041		0.041	
COBMISS	0.006		0.006		0.006		0.006	
Access to Hospitals:								
*ACCESS1**	0.353				0.514		0.547	
ACCESS2	0.191				0.292		0.282	
ACCESS3	0.076				0.121		0.108	
ACESSMISS	0.045				0.073		0.064	

*Indicates the omitted category in regressions. Standard deviations are provided for continuous variables only.

TABLES

Table 1: Private health insurance cover, analysis sample

Group	Wave 1: 1996	Wave 2: 2000	Wave 3: 2003	No of women	%
1 'Continuous cover'	Y	Y	Y	1398	19.0
2 'Recent leaver'	Y	Y	N	166	2.3
3 'Leaver'	Y	N	N	599	8.1
4 'Never covered'	N	N	N	3252	44.2
5 'New joiner'	N	N	Y	839	11.4
6 'Joiner'	N	Y	Y	634	8.6
7 'Churn out'	N	Y	N	115	1.6
8 'Churn in'	Y	N	Y	357	4.9
			Total	**7360**	**100.0**
% with insurance cover	34%	31%	44%	36.5%	

Notes: Y indicates has cover and N no cover.

Table 2: Sample sizes by household type, analysis sample.

Actual Children	Desired Children			
	Desired > Actual	**Desired ≤ Actual**	**Indeterminate**	**Total (%)**
	Wave 2			
No children	5380	521	0	5901 (80)
(%)	(91)	(9)		(100)
With children	968	354	137	1459 (20)
(%)	(66)	(24)	(9)	(100)
Total	6348	875	137	7360 (100)
	Wave 3			
No children	4505	564	0	5069 (69)
(%)	(89)	(11)		(100)
With children	1335	658	298	2291 (31)
(%)	(58)	(29)	(13)	(100)
Total	5840	1222	298	7360 (100)
Timing of children	**Households with children, waves 2 & 3**			
Early period only	550	252	13	815 (22)
Recent period only	1506	388	110	2004 (53)
Both periods	247	372	312	931 (25)
Total	2303	1012	435	3750 (100)

Notes: Indeterminate indicates that desired and actual numbers of children are equal to or greater than 3. The early period refers to time prior to the previous wave while the recent period refers to time between the previous and the current wave. For example, with wave 3 observations, "Early period only" indicates households that do not have additional children in wave 3 compared to wave 2 while "Both periods" indicates households in which children are born before wave 2 and between waves 2 and 3.

Table 3: Observed private health insurance cover by household type, waves 2 and 3 (number of observations=14720).

Household type according to actual children:		PHI Sample Frequencies D > A	PHI Sample Frequencies D ≤ A	Differences
No Children	(1)	0.406	0.331	0.075
With Children:				
Early period only	(2)	0.236	0.206	0.030
Recent period only	(3)	0.397	0.278	0.119
Both periods	(4)	0.373	0.301	0.072

Notes: Frequencies for the indeterminate group are not shown. The early period refers to time prior to the previous wave while the recent period refers to time between the previous and the current wave. For example, with wave 3 observations, "Early period only" indicates households that do not have additional children in wave 3 compared to wave 2 while "Both periods" indicates households in which children were born before wave 2 and between waves 2 and 3. D indicates desired while A indicates actual number of children.

Table 4: Effects of children variables on probabilities of private health insurance – various models. (standard errors in parentheses)

	Average predicted probabilities of PHI - Pooled Probits		Average marginal effects of aspirations for children (Col. 2 – col. 3) – various models:		
	D > A	D ≤ A	Pooled Probits	Random Effects	IPW
No Children (1)	0.395	0.368	0.027 (0.001)	0.029 (0.001)	0.030 (0.001)
With Children:					
Early period only (2)	0.302	0.273	0.029 (0.001)	0.033 (0.002)	0.045 (0.001)
Recent period only (3)	0.376	0.329	0.047 (0.001)	0.051 (0.002)	0.040 (0.001)
Both periods (4)	0.355	0.309	0.046 (0.001)	0.047 (0.002)	0.043 (0.001)

Average marginal effects of children – various models:						
	Pooled probits		Random effects		IPW	
	D > A	D ≤ A	D > A	D ≤ A	D > A	D ≤ A
Presence:						
No children in early per. (Row (3) - row (1))	-0.019 (0.001)	-0.039 (0.001)	-0.022 (0.001)	-0.044 (0.003)	-0.022 (0.001)	-0.037 (0.001)
With children in early per. (Row (4) - row (2))	0.053 (0.001)	0.036 (0.001)	0.055 (0.002)	0.040 (0.002)	0.050 (0.001)	0.048 (0.001)
Timing:						
Row (2) – row (3)	-0.074 (0.001)	-0.056 (0.002)	-0.074 (0.001)	-0.056 (0.003)	-0.063 (0.001)	-0.058 (0.002)
Income effects:						
Row (2) – row (1)		-0.095 (0.001)		-0.100 (0.002)		-0.095 (0.001)
Learning effects:						
Row (2) – row (1) – income effect	0.002 (0.002)		0.004 (0.002)		0.010 (0.002)	

Notes: The number of observations in each model is 14720. Regressions also include: lagged insurance cover, insurance cover in the first period, household income (8 categories), age an indicator for those who left home in the previous year, self assessed health (5 categories), marital status (3 categories), an indicator for those living in Queensland, an indicator for those who have a health card, level of urbanization (3 categories), an indicator for smokers, drinkers, body mass index, country of birth (4 categories), adummy for the year 2000, work/study status (4 categories), perceived access to hospitals (3 categories), education (4 categories) and 9 indicators for missing values for income, urbanization, marital status, education, access, work/study, body mass index, self-assessed health, and country of birth. Standard errors are computed from 200 bootstrapped samples clustered by individual identifiers. In the case of random effects probits, the marginal effects are average partial effects; i.e. an average is taken over the distribution of individual-specific and unobserved random effects.

Table 5: Effects of children variables on probabilities of buying or dropping private health insurance – pooled probits (standard errors in parentheses)

	Average predicted probability of buying PHI		Average marg. effects of aspirations	Average predicted probability of dropping PHI		Average marg. effects of aspirations
	D > A	D ≤ A	Diff.	D > A	D ≤ A	Diff.
No Children (1)	0.251	0.198	0.053 (0.001)	0.316	0.295	0.020 (0.001)
With Children:						
Early period only (2)	0.138	0.101	0.037 (0.001)	0.327	0.253	0.073 (0.003)
Recent period only (3)	0.217	0.182	0.035 (0.002)	0.296	0.390	-0.094 (0.003)
Both periods (4)	0.142	0.139	0.003 (0.002)	0.049	0.300	-0.250 (0.006)
Average marginal effects of children :						
	D > A	D ≤ A		D > A	D ≤ A	
Presence:						
No children in early per. (Row (3) - row (1))	-0.033 (0.001)	-0.015 (0.002)		-0.020 (0.001)	0.095 (0.004)	
With children in early per. (Row (4) - row (2))	0.004 (0.002)	0.038 (0.001)		-0.277 (0.003)	0.046 (0.004)	
Timing:						
Row (3) – row (2)	-0.079 (0.001)	-0.082 (0.002)		0.031 (0.003)	-0.137 (0.004)	
Income effects:						
Row (2) – row (1)		-0.097 (0.001)			-0.042 (0.003)	
Learning effects:						
Row (2) – row (1) – income effect	-0.016 (0.002)			0.053 (0.004)		

Notes: The number of observations in each model is 14720. Variables include all those listed in the notes to Table 4 and interactions between the children indicators and the lagged insurance cover. Standard errors are computed from 200 bootstrapped samples clustered by individual identifiers.

Table 6: Effects of children variables on probabilities of having private health insurance by household income group – pooled probits (standard errors in parentheses)

	High Income (top 23%)			Low Income (bottom 19%)		
	Avg. predicted probs. of PHI		Avg. marg. effects of aspirations	Avg. predicted probs. of PHI		Avg. marg. effects of aspirations
	D > A	D ≤ A	Diff.	D > A	D ≤ A	Diff.
No Children (1)	0.567	0.522	0.045 (0.001)	0.248	0.247	0.001 (0.001)
With Children:						
Early period only (2)	0.585	0.585	-0.000 (0.002)	0.212	0.148	0.064 (0.001)
Recent period only (3)	0.593	0.565	0.028 (0.003)	0.233	0.196	0.037 (0.001)
Both periods (4)	0.693	0.447	0.247 (0.002)	0.147	0.214	-0.066 (0.001)
Average marginal effects of children :						
	D > A	D ≤ A		D > A	D ≤ A	
Presence:						
No children in early per. (Row (3) - row (1))	0.026 (0.001)	0.043 (0.002)		-0.016 (0.001)	-0.051 (0.001)	
With children in early per. (Row (4) - row (2))	0.109 (0.002)	-0.139 (0.002)		-0.065 (0.001)	0.066 (0.001)	
Timing:						
Row (3) – row (2)	-0.008 (0.002)	0.020 (0.003)		-0.021 (0.001)	-0.048 (0.001)	
Income effects:						
Row (2) – row (1)		0.063 (0.002)			-0.099 (0.001)	
Learning effects:						
Row (2) – row (1) – income effect	-0.045 (0.002)			0.063 (0.002)		

Notes: Separate pooled probits are estimated on the subset of households based on their income groups. The low income group includes 2953 observations while the high income group consists of 3410 observations. Standard errors are computed from 200 bootstrapped samples clustered by individual identifiers.

Table 7: Average marginal effects of selected other variables on the probability of insurance cover – pooled probits

	Average predicted probability	Average marginal effects	(p-value)
Lagged insurance			
Not insured in t-1	0.222		
Insured in t-1	0.685	0.464	(0.000)
Wave 1 insurance			
Not insured in wave 1	0.363		
Insured in wave 1	0.400	0.028	(0.000)
Age (in years)		0.017	(0.000)
Marital status			
Single/separated	0.349		
Defacto	0.333	-0.017	(0.083)
Married	0.442	0.093	(0.000)
Perceived access to hospitals			
Access level 1 (high)	0.413		
Access level 2 (medium)	0.350	-0.064	(0.000)
Access level 3 (low)	0.316	-0.098	(0.000)
Self-assessed health			
Excellent	0.389		
Very Good	0.372	-0.017	(0.090)
Good	0.379	-0.010	(0.339)
Fair	0.380	-0.009	(0.526)
Poor	0.392	0.003	(0.935)
Smoking			
Not a smoker	0.385		
Smoker	0.334	-0.508	(0.000)
Health card			
Does not have a health card	0.382		
Has a health card	0.354	-0.028	(0.005)
Country of birth			
Australia	0.377		
English speaking	0.352	-0.025	(0.107)
Other European	0.401	0.023	(0.486)
Asia	0.443	0.066	(0.010)
Level of urbanization			
Urban	0.395		
Rural	0.344	-0.051	(0.000)
Remote	0.451	0.055	(0.002)
Year 2000			
Not in year 2000	0.406		
Year 2000	0.347	-0.059	(0.000)

Notes: p-values based on the standard errors for the underlying coefficients. These are robust to correlation for individuals across waves.