# A study of Australian individuals' Private Health Insurance

ETC4420 Microeconometrics Assignment

## Contents



# Introduction

## Motivation of this study

Healthcare is one the most important topics in almost every country. An efficient combination of services from public and private health institutions in a country can certainly facilitate the development in that country and also enhance its residents' welfare. Given people's behavior in healthcare relies heavily on the health insurance coverage, government policy could play a significant role in this scenario. For example, policies encouraging people to buy Private Health Insurance (PHI) can also increase the private hospital usage. Hence, the study of how policies impact individuals' decision on purchasing PHI becomes more and more important. Having accurate estimations of the marginal effect of certain policies on the PHI purhcasing decision would be extremely useful for government to make right decisions in the future. As for the statement that "people's behavior in healthcare relies heavily on the health insurance coverage", it is intuitively true. But we still need to confirm its validation. What's more, we want to study to what extend does the health insurance coverage affect the individuals' beheviors. For ease of exposition, only the relationship between the utilization

of dental service and the PHI status will be studied in this paper. But the results can be generalized to other cases including private health service usage etc.

## Problem statement

Therefore, in this study, we are interested in two questions.

Firstly, what are the determinants of Australian individuals' decision on PHI purchase and how are these determinants affecting the people's choices?

Secondly, is an individual's dental service usage affected by his/her PHI status? More specifically, we want to find out the treatment effect of "having private private heal insurance" on dental service usage. We will discuss these two questions separately in the following paper.

## Literature review

Despite of the fact that a lot of studies have been conducted relating the demand for private health insurance, not much of them asked the question about how PHI is affecting the healthcare utilisation. (Srivastava, Zhao, and others 2008) In terms of dental health services, there are even fewer. Related to our second question in this study, Dawkins et al. (2004) found that making more people purchase PHI does not necessarily increase the utility of the private health services.

Although Srivastava, Zhao, and others (2008) studied the impact of private health insurance on the individuals' choice of public versus private hospital services, not directly relate to the dental care service, the methodology, RUM analysis and the recursive trivariate probit system model, used in their paper shed light on this study.

However, it is also worth noting that "having PHI which covers dental service" does not necessarily mean "paying nothing" for insured people. To illustrate this, please refer to figure 1. According to Chrisopoulos et al. (2016), "in 2013, half (50%) of people aged 5 and over had some level of private health insurance with dental cover. More people living in Major cities (53%) had insurance than in Inner regional (45%) and Outer regional areas (44%). People living in lower income households were less likely to have dental insurance than those in higher income households. Most (77%) adults with some level of health insurance that covered dental made co-contributions towards the cost of dental visits. One (1) in 10 insured adults (10%) paid all their own expenses. Of these, approximately 19% reported that this caused a large financial burden. Total expenditure on dental services (except those in hospitals) in Australia was AUD 8,706 million in 2012–13, an increase from AUD 5,945 million (adjusted for inflation) in 2002–03. The largest source of funds for this overall dental expenditure was individuals, paying directly out of pocket for 58% of total dental costs." Hence, maybe the question more interesting to researchers is the impact of "all free dental service" on "dental visits". And this needs more care than simply using PHI status.

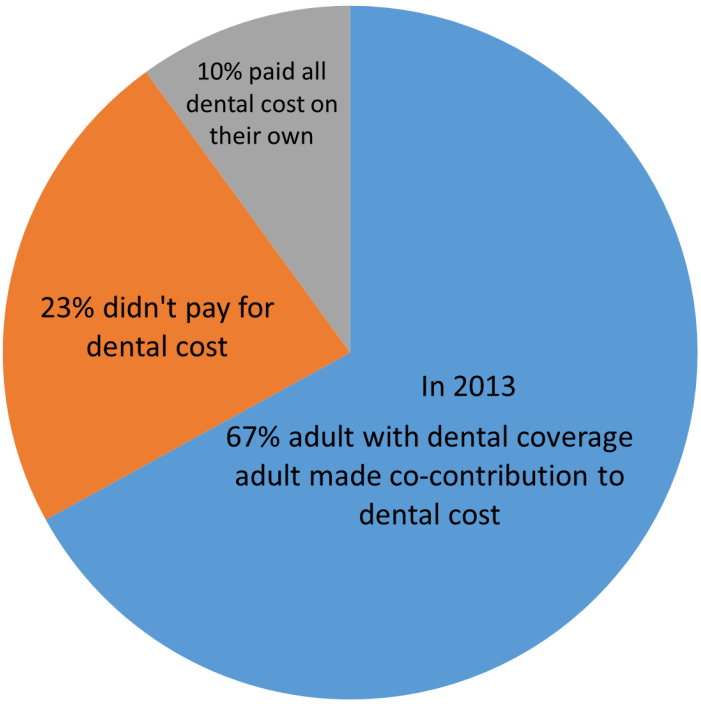


Figure 1: A description of the co-payment paid by insured people in 2013.

# Econometric Framework

## Multinomial logit

In related to our first question, there are four types of PHI status, "without PHI", "only cover hospital", "only cover ancillary" and "cover both hospital and ancillary". There are no apparent natural orders in them. And we also don't have observability issues whatsoever since our sample was randomly selected from the whole population in Australia as described in data section in this study. Therefore We use Multinomial Logit (MNL) to model it.

Suppose the utility for individual $i$ choosing choice $j$ is given by:

$$U_{ij} = x_i' \beta_j + \epsilon_{ij} \quad (i = 1, \ldots, N; \quad j = 0, 1, \ldots, J)$$

where $x_i$ only vary by $i$ and are the same across j, and $\beta_j$ differs by choice j. We observe $y_i = j$ iff.$U_{ij} > U_{ik} \quad \forall k \neq j (j, k = 0, 1, \ldots, J)$, using a Type I extreme value distribution with $F(\epsilon_{ij} = \exp[-\exp(-\epsilon_{ij})]$. Probabilities:

$$P_{ij} = P(y_i = j) = \frac{e^{x_i' \beta_j}}{\sum_{k=0}^{J} e^{x_i' \beta_j}} \quad (j = 0, 1, \ldots, J)$$

where $\beta_0$ is standardized to be zero. So

$$P_{i0} = P(y_i = 0 | x_i) = \frac{1}{1 + \sum_{k=1}^{J} e^{x_i' \beta_k}}$$

$$P_{ij} = P(y_i = j|x_i) = \frac{e^{x_i'\beta_k}}{1 + \sum_{k=1}^{J} e^{x_i'\beta_k}}$$

J=3 in this study. Since there are 4 (J+1=4) types of insurance status. The x variables used in this model are defined in data section, and can also be found in the output table.

## Endogenous binary treatment for a binary outcome variable - EBVP

For the second question in this study, the EBVP model is used. Because the main outcome variable "whether having visited a dentist in past 12 months" (y) is a binary variable. The treatment variable "whether having PHI" (T) is also a binary variable and it is endogenous because of the "self-selection" issues described in the discussion section. The model structure is given by:

$$y_i^* = x_{i1}'\beta_1 + \delta T_i + \epsilon_{i1}, \quad y_i = 1 \text{ if } y_i^* > 0, \ y_i = 0 \text{ otherwise}$$

$$T_i^* = x_{i2}'\beta_2 + \epsilon i2, \ T_i = 1 \text{ if } T_i^* > 0, \ T_i = 0 \text{ otherwise}$$

$$\text{where}(\epsilon_{i1}, \epsilon_{i2}) \sim BVN(0, 0, 1, 1, \rho), \ -1 < \rho < 1$$
$$\rho \text{ is the correlation coefficient between } \epsilon_{i1} \text{and} \epsilon_{i2}$$

We will use the Maximum likelihood estimation because it gives consistent and efficient estimates. The log-likelihood function is

$$LogL = \sum_{y=1,T=1} log\ \Phi_2[x_{i1}'\beta_1 + \delta T_i, \ x_{i2}'\beta_2, \ \rho] + \sum_{y=1,T=0} log\ \Phi_2[x_{i1}'\beta_1, \ -x_{i2}'\beta_2, \ -\rho] \quad (1)$$
$$+ \sum_{y=0,T=1} log\ \Phi_2[-x_{i1}'\beta_1 + \delta T_i, \ x_{i2}'\beta_2, \ -\rho] + \sum_{y=0,T=0} log\ \Phi_2[-x_{i1}'\beta_1, \ -x_{i2}'\beta_2, \ \rho] \quad (2)$$

Results of interest: treatment effect of T on y:

$$TE_i = \Phi(x_{i1}'\beta_1 + \delta) - \Phi(x_{i1}'\beta_1)$$

# Data

Our data source is the 2004-05 Australian National Health Survey (NHS) which was conducted using a representative sample of the non-institutionalised residential population. The

sample design ensured that individuals were randomly selected within each state and territory (*Information Paper, National Health Survey, Curf 2004-05* 2006). The data we use in this study is a randomly selected subset of 22,000 individuals from this data source.

To study the factors may having impact on PHI purchasing, we consider different potential areas, personal demographic, health condition, income and education. Relevant variables in our dataset was defined in table below. After checking the cross tabulation between type of PHI and age band, this study focus on people who are older than 14 years because lack of information for people aged 14 years and below. What's more, from the cross tabulation with percentage between age band and types of PHI, we found that there are some consecutive age band having very similar distribution against types of PHI. Hence, we double checked the cross tabulation between original age band with dental visits. It turned out that we will not lose much information by doing re-grouping to the age band according to its behavior. Hence, we re-grouped the age band into six groups as given in the "definition of variables" table below and the first group "age 0-14" will be omitted from all regressions in this study.

Intuitively thinking, the dental visit has to be correlated with age and life style factors such as smoking and drinking behavior. Because these factors are more likely to cause dental issues. What's more, other demographic may also have impact on the dental visits behavior, so we will also include the ones used in the first question. In summary, for the main outcome equation, we will include all variables used in the first question, and also include smoking habit and drinking behavior. As to the treatment equation, we want to include some IV variables to address the endogeneity problem for the treatment. Idealy, the instrument will be strongly correlated with the treatment eg. whether having PHI or not; but not correlated with the error terms in the main outcome equation, eg. uncorrelated with dental visits. However, given the limitation of the data we have, the only potential IV is the "number of times consulted general pratitioner". Since people often visit GP are supposed to have a higher demand for insurance, but more GP visits does not lead to more dental visits. Since the general health condition is to some extend independent with the dental health. Maybe it is the reason for the dental care service to be so independent from other departments in the medical system. So in this research, we will use this variable as an IV in the treatment equation.

Table 1: Definition of variables

| Area | Variables |
|---|---|
| Type of private health insurance cover | phi=0 for people without PHI, phi=1 for hospital cover only, phi=2 for ancillary cover only, phi=3 for both hospital and ancillary cover. |
| Whether having visited dentist in the past 12 months | dvisit=1 for people have visited dentist in last year, 0 otherwise. |

| Area | Variables |
|---|---|
| Age | age14 for age group 0-14 years, 0 otherwise (omitted from regression); age1519 for age group 15-19 years, 0 otherwise (ref category); age2039 for age group 20-39 years, 0 otherwise; age4064 for age group 40-64 years, 0 otherwise; age6579 for age group 65-79 years, 0 otherwise; age80 for age group 80 years and over, 0 otherwise. |
| Culture | Australia=1 for people born in Australia, 0 otherwise; English=1 for people born in English-speaking coutry, 0 otherwise; nonEng=1 for others, 0 otherwise (ref category). |
| Other demographic variables | male=1 for male, and 0 otherwise; married=1 if married or de facto, and 0 otherwise. |
| Long term medical conditions | condnoc=0 for people has no long term condition, 1 for having 1 long term condition, 2 for having 2 long term condition, 3 for having 3 long term condition, 4 for having 4 long term condition, 5 for having more than 5 long term condition. |
| General health assessment | excelh=1 for excellent health, 0 otherwise; verygood=1 for verygood health, 0 otherwise; good=1 for good health, 0 otherwise; fair=1 for fair health, 0 otherwise; poor=1 for poor health, 0 otherwise (ref category). |
| Education | degree=1 for a tertiary degree, 0 otherwise; dipcert=1 for having diploma or certificate, 0 otherwise; less12yr=1 for no qualification, 0 otherwise. |
| Income | income3=1 if weekly personal income falls in 1st to 3rd income decile, 0 otherwise (ref category); income4=1 if weekly personal income falls in the 4th income decile, 0 otherwise; income5=1 if weekly personal income falls in the 5th income decile, 0 otherwise; income6=1 if weekly personal income falls in the 6th income decile, 0 otherwise; income7=1 if weekly personal income falls in the 7th income decile, 0 otherwise; income8=1 if weekly personal income falls in the 8th income decile, 0 otherwise; income9=1 if weekly personal income falls in the 9th income decile, 0 otherwise; income10=1 if weekly personal income falls in the 10th income decile, 0 otherwise. |
| Main occupation | workft=1 if employed full-time, 0 otherwise; workpt=1 if employed part-time, 0 otherwise; unemp=1 if unemployed, 0 otherwise; nlf=1 if not in the labour force, 0 otherwise (ref category). |

| Area | Variables |
|---|---|
| Smoking habit | nonsmoke=1 for never smoke, 0 otherwise (ref category); exsmoke=1 for once regularly smoke but not any more, 0 otherwise; currentsmk=1 for current smoke but not regular, 0 otherwise; regularsmk=1 for currently regular smoker, 0 otherwise. |
| Alcohol risk | low=1 for low risk (ref category), 0 otherwise; medium=1 for medium risk, 0 otherwise; high=1 for high risk, 0 otherwise. |
| Times visited GP | gpvisit=1 for visited GP once, 2 for twice, 3 for 3 or more times, 0 for no visits. |

# Results

Because the coefficients obtained in MNL has no meaningful interpretation, neither does its sign. Hence we do not report its estimation in this paper. The table 2 is the average marginal effect obtained from MNL.

Some interesting findings that worthy noting here are, "poor health" does have an positive effect on "purchasing PHI" since the sign of all other health conditions are negative.

Average Predicted Probability are computed as shown in the table below. As we can see from this table, the average predicted probability is very close to the actual proportion of each type of insurance status in this data.

Hit and loss function was also calculated. We can see the hit and loss table indicates the model does not have a strong prediction ability.

For the second question, the estimation output of EBVP can be found in the table below. As we have learned that it is very difficult to calculate the marginal effects for this kind of model. Because the non-linear recursive structure.

From this output, we can see some interesting findings. $\rho$ is negative which is inconsistent with our expectation, and also it is not significant. The reason for this is likely due to the lack of good IV as we have discussed above. We have tried to remove the IV (number of times visited GP) in the regression, and it indeed has some influence on the estimation of $\rho$ by increasing it. Hence, we believe that good IV can really help to address this problem.

# Discussion

Firstly we need to acknowledge the limitation of this study in that the data we have is not very big. What's more, many features that could have contributed to this study are not included in this subset. For instance, if we have information on individuals' other insurance purchasing

Table 2: Average Marginal Effects obtained from MNL for the first question

| Variables | Y=0 | Y=1 | Y=2 | Y=3 |
|---|---|---|---|---|
| | dy/dx (sd) | dy/dx (sd) | dy/dx (sd) | dy/dx (sd) |
| age2039 | 0.291 ( 0.015 ) | -0.027 ( 0.01 ) | -0.01 ( 0.006 ) | -0.254 ( 0.015 ) |
| age4064 | 0.147 ( 0.015 ) | 0.016 ( 0.01 ) | -0.01 ( 0.006 ) | -0.153 ( 0.015 ) |
| age6579 | 0.102 ( 0.017 ) | 0.047 ( 0.011 ) | -0.022 ( 0.008 ) | -0.128 ( 0.017 ) |
| age80 | 0.171 ( 0.023 ) | 0.075 ( 0.012 ) | -0.085 ( 0.019 ) | -0.162 ( 0.023 ) |
| Australia | -0.076 ( 0.01 ) | -0.004 ( 0.006 ) | -0.006 ( 0.004 ) | 0.086 ( 0.01 ) |
| English | -0.021 ( 0.014 ) | -0.021 ( 0.008 ) | 0.008 ( 0.005 ) | 0.034 ( 0.014 ) |
| male | 0.066 ( 0.007 ) | -0.011 ( 0.004 ) | -0.007 ( 0.003 ) | -0.048 ( 0.007 ) |
| married | -0.157 ( 0.007 ) | 0.022 ( 0.004 ) | 0.007 ( 0.003 ) | 0.129 ( 0.007 ) |
| condnoc | -0.02 ( 0.002 ) | 0 ( 0.001 ) | 0.001 ( 0.001 ) | 0.019 ( 0.002 ) |
| excelh | -0.221 ( 0.019 ) | 0.042 ( 0.012 ) | 0.014 ( 0.009 ) | 0.166 ( 0.02 ) |
| verygood | -0.176 ( 0.018 ) | 0.031 ( 0.011 ) | 0.013 ( 0.009 ) | 0.132 ( 0.019 ) |
| good | -0.133 ( 0.018 ) | 0.024 ( 0.011 ) | 0.017 ( 0.009 ) | 0.092 ( 0.019 ) |
| fair | -0.062 ( 0.019 ) | 0.018 ( 0.012 ) | 0.017 ( 0.009 ) | 0.027 ( 0.02 ) |
| degree | -0.171 ( 0.011 ) | 0.026 ( 0.006 ) | -0.003 ( 0.005 ) | 0.148 ( 0.01 ) |
| dipcert | -0.031 ( 0.008 ) | -0.003 ( 0.005 ) | 0.005 ( 0.003 ) | 0.03 ( 0.008 ) |
| income4 | 0.046 ( 0.013 ) | 0.001 ( 0.007 ) | -0.005 ( 0.006 ) | -0.042 ( 0.013 ) |
| income5 | -0.008 ( 0.013 ) | -0.009 ( 0.008 ) | 0.003 ( 0.005 ) | 0.013 ( 0.013 ) |
| income6 | -0.01 ( 0.014 ) | -0.001 ( 0.009 ) | 0.001 ( 0.006 ) | 0.01 ( 0.014 ) |
| income7 | -0.03 ( 0.016 ) | -0.003 ( 0.01 ) | 0.008 ( 0.006 ) | 0.026 ( 0.016 ) |
| income8 | -0.1 ( 0.015 ) | 0.001 ( 0.009 ) | -0.003 ( 0.006 ) | 0.102 ( 0.015 ) |
| income9 | -0.175 ( 0.015 ) | 0.016 ( 0.008 ) | -0.005 ( 0.006 ) | 0.164 ( 0.014 ) |
| income10 | -0.313 ( 0.018 ) | 0.035 ( 0.009 ) | -0.016 ( 0.008 ) | 0.294 ( 0.016 ) |
| workft | -0.079 ( 0.012 ) | 0.003 ( 0.007 ) | 0.011 ( 0.005 ) | 0.066 ( 0.012 ) |
| workpt | -0.089 ( 0.011 ) | 0.001 ( 0.007 ) | 0.008 ( 0.004 ) | 0.08 ( 0.011 ) |
| unemp | 0.129 ( 0.025 ) | -0.049 ( 0.021 ) | 0.007 ( 0.009 ) | -0.087 ( 0.027 ) |

Table 3: Average Predicted Probabilities from MNL for 1st question.

| pp | obs | mean | sd | min | max |
|---|---|---|---|---|---|
| PHI=0 | 18,229 | 0.500 | 0.197 | 0.029 | 0.953 |
| PHI=1 | 18,229 | 0.081 | 0.035 | 0.005 | 0.273 |
| PHI=2 | 18,229 | 0.038 | 0.016 | 0.002 | 0.118 |
| PHI=3 | 18,229 | 0.381 | 0.176 | 0.030 | 0.859 |

Table 4: Hit and Loss from MNL for 1st question.

| PHI | PHI hat=0 | PHI hat=3 | Total |
|---|---|---|---|
| 0 | 7,164 | 1,860 | 9,024 |
| 1 | 807 | 651 | 1,458 |
| 2 | 440 | 248 | 688 |
| 3 | 3,143 | 3,758 | 6,901 |

status, we could use it as an IV in our second regression to address the endogeneity problem. Because highly "risk-averse" people are more likely to buy other types of insurance such as life insurance. And this type of insurance has no direct effect on dental visits whatsoever. In addition, related to what we have discussed in introduction, "having coverage on dental service" often does not equal to "free dental service". Based on different insurance policy, people still need to pay a lot of money to get dental care. This is another pitfall for us to fully study the impact of PHI on dental visits. To properly study this point, we may need to do some research on some functional form between having PHI and the co-payment made for dental care.

A good feature of this data set is that it was randomly selected from the Australian population, both insured and uninsured people are observed. Therefore, there is no observability issue in this case. However, there still be endogeniety problem in the model because people may "self-select" themselves into purchasing PHI. To be more specific, the unobservable factors affecting PHI self-selection may ne correlated with the factors making a person seeing a dentist. For instance, for people who are highly risk averse, the chance for them to purchase PHI is higher than other people, and it is also likely that they visit dentists more frequently to prevent possible dental disease. Since the "risk averse" characristic is unobservable, it will be left in the error terms. Thus regressor T (whether having PHI) will be positively correlated with the error term in the outcome Y (dentist visits) equation. And if this is the case, the effect of having PHI will be overestimated.

# References


Chrisopoulos, Sergio, JE Harford, A Ellershaw, and others. 2016. *Oral Health and Dental Care in Australia: Key Facts and Figures 2015.* Australian Institute of Health; Welfare.

Dawkins, Peter, Elizabeth Webster, Sandra Hopkins, Jongsay Yong, and A Palangkaraya. 2004. "Recent Private Health Insurance Policies in Australia: Health Resource Utilization, Distributive Implications and Policy Options." *A Report Prepared for the Department of Premier and Cabinet, the Government of Victoria.*

*Information Paper, National Health Survey, Curf 2004-05.* 2006. Australia Bureau of Statistics.

Srivastava, Preety, Xueyan Zhao, and others. 2008. *Impact of Private Health Insurance on*

Table 5: Cross tabulation between age and PHI types

| Years | PHI=0 | PHI=1 | PHI=2 | PHI=3 | Total |
|---|---|---|---|---|---|
| 15-19 | 741 | 91 | 67 | 601 | 1,500 |
| 20-24 | 742 | 50 | 39 | 304 | 1,135 |
| 25-29 | 813 | 45 | 61 | 361 | 1,280 |
| 30-34 | 921 | 86 | 59 | 589 | 1,655 |
| 35-39 | 880 | 124 | 72 | 667 | 1,743 |
| 40-44 | 813 | 138 | 73 | 741 | 1,765 |
| 45-49 | 716 | 136 | 79 | 689 | 1,620 |
| 50-54 | 547 | 127 | 60 | 712 | 1,446 |
| 55-59 | 537 | 138 | 56 | 677 | 1,408 |
| 60-64 | 494 | 121 | 45 | 514 | 1,174 |
| 65-69 | 440 | 97 | 26 | 359 | 922 |
| 70-74 | 454 | 90 | 31 | 290 | 865 |
| 75-79 | 414 | 101 | 16 | 192 | 723 |
| 80-84 | 301 | 82 | 3 | 128 | 514 |
| 85- | 211 | 32 | 1 | 77 | 321 |

Table 6: Cross tabulation between age and dental visits

| Years | Dental visits=0 | Dental visits=1 | Total |
|---|---|---|---|
| 0-4 | 871 | 24 | 895 |
| 0-4 | 1,208 | 109 | 1,317 |
| 5-9 | 1,396 | 163 | 1,559 |
| 15-19 | 1,435 | 114 | 1,549 |
| 20-24 | 1,103 | 60 | 1,163 |
| 25-29 | 1,230 | 60 | 1,290 |
| 30-34 | 1,583 | 80 | 1,663 |
| 35-39 | 1,672 | 81 | 1,753 |
| 40-44 | 1,661 | 111 | 1,772 |
| 45-49 | 1,538 | 90 | 1,628 |
| 50-54 | 1,362 | 87 | 1,449 |
| 55-59 | 1,317 | 95 | 1,412 |
| 60-64 | 1,096 | 90 | 1,186 |
| 65-69 | 869 | 57 | 926 |
| 70-74 | 814 | 53 | 867 |
| 75-79 | 681 | 46 | 727 |
| 80-84 | 499 | 20 | 519 |
| 85- | 311 | 14 | 325 |

*the Choice of Public Versus Private Hospital Services*. University of York. Centre for Health Economics.