Life Course Centre

The Australian Research Council Centre of Excellence for Children and Families over the Life Course
Phone +61 7 3346 7477 Email lcc@uq.edu.au lifecoursecentre.org.au

WORKING PAPER SERIES

ACCURACY OF SELF-REPORTED PRIVATE HEALTH INSURANCE COVERAGE

Ha Trong Nguyen

Telethon Kids Institute & The University of Western Australia

Huong Thu Le

Telethon Kids Institute & The University of Western Australia

Luke Connelly

The University of Queensland & The University of Bologna

Francis Mitrou

Telethon Kids Institute & The University of Western Australia

No. 2022-24

December 2022

NON-TECHNICAL SUMMARY

Studies on health insurance coverage often rely on measures self-reported by respondents and this is the case in both developed and developing countries. The presumed accuracy of survey reports of health insurance enrolment influences how these data are used for health policy evaluations, yet the accuracy of such measures outside the United States has not been thoroughly validated. This paper aims to fill that gap in the literature by presenting the first evidence on the extent and factors associated with accuracy of private health insurance (PHI) coverage reporting in an Australian context.

This paper uses linked Australian National Health Survey and administrative population tax data to explore the accuracy of self-reported private health insurance coverage in survey data. We find that 9% of individuals misreport their PHI coverage status, with 5% of true PHI holders reporting that they are uninsured and 16% of true non-insured persons self-identifying as insured. Our results show reporting errors are systematically correlated with individual and household characteristics, including age, migration status, education, marital status, employment status and income. We additionally find that most of these characteristics influence the probability of giving a false negative or a false positive report very differently. Our evidence on the determinants of errors is supportive of common reasons for misreporting. We directly investigate biases in the determinants of PHI enrolment using survey data. We find that, as compared to administrative data, survey data depict a quantitatively different picture of PHI enrolment determinants, especially those capturing age, language proficiency, labour force status or the number of children. We also show that PHI coverage misreporting is subsequently associated with misreporting of reasons for purchasing PHI, type of cover and length of cover.

This study finds that reporting accuracy of PHI coverage is quite high in a nationally representative health survey in Australia, providing some good news for studies using such survey data to document PHI coverage. Our evidence of the factors associating with PHI misreporting may provide useful insights for the constructors of surveys to consider in order to improve accuracy of responses to PHI-related questions. Our finding of a substantial relationship between PHI coverage misreporting and a range of explanatory variables indicates that reporting errors of PHI enrolment in survey data are non-classical. These non-classical errors suggest complicated biases in other studies that use self-reported PHI enrolment as an independent variable in regressions, including those evaluating effects of PHI enrolment on health care utilization and health outcomes.

ABOUT THE AUTHORS

Ha Nguyen is a Senior Research Fellow at the Telethon Kids Institute. He is also an Adjunct Senior Research Fellow at Centre for Child Health Research, the University of Western Australia, and a Research Fellow in the Life Course Centre. His general research interests have focused on applied econometrics, particularly in the fields of health economics and labour economics. He has published widely in both academic and policy outlets, with articles appearing in high-ranking economics journals, including *Journal of International Economics, Health Economics*, *American Journal of Health Economics, Labour Economics, Economics of Education Review,* and *Economic Record.* Email: ha.nguyen@telethonkids.org.au

Huong Le is a Biostatistician/Data Analyst at the Telethon Kids Institute and a Statistician and Data Officer at School of Population and Global Health, University of Western Australia. She completed her PhD in Economics from the Australian National University. Her current research interests include impact evaluation, child health and development, and intergenerational transmission of health and human capital. She has published in high-impact journals, including *Health Economics*, *American Journal of Health Economics, Review of Income and Wealth*, *Clinical Infectious Diseases* and *IZA Journal of Labor Economics.* Email: huong.le@telethonkids.org.au

Luke Connelly is Professor of Health Economics at the University of Queensland, Centre for the Business and Economics of Health and The University of Bologna, Department of Sociology and Business Law. His main interests are in health economics and insurance economics and the effects of institutions (including legal constructs) on incentives and behaviour. He has also worked in other fields of applied microeconomics, including education economics and transport economics. Email*:* l.connelly@uq.edu.au; luke.connelly@unibo.it

Francis Mitrou is Research Program Head–Population Health, at Telethon Kids Institute. He is also Associate Professor at the Centre for Child Health Research, at The University of Western Australia, and a Chief Investigator in the ARC Life Course Centre. Beginning as an economist with the Australian Bureau of Statistics before entering academia, he has expertise in national population surveys, linked administrative data, Indigenous health and wellbeing, and economic evaluation, with a focus on child and adolescent development, social disadvantage and life course trajectories. Email: Francis.Mitrou@telethonkids.org.au

Acknowledgements

This research was partly funded by the Australian Research Council Centre of Excellence for Children and Families over the Life Course #CE200100025).

ABSTRACT

Studies on health insurance coverage often rely on measures self-reported by respondents, but the accuracy of such measures has not been thoroughly validated. This paper is the first to use linked Australian National Health Survey and administrative population tax data to explore the accuracy of self-reported private health insurance (PHI) coverage in survey data. We find that 9% of individuals misreport their PHI coverage status, with 5% of true PHI holders reporting that they are uninsured and 16% of true non-insured persons self-identifying as insured. Our results show reporting errors are systematically correlated with individual and household characteristics. Our evidence on the determinants of errors is supportive of common reasons for misreporting. We directly investigate biases in the determinants of PHI enrolment using survey data. We find that, as compared to administrative data, survey data depict a quantitatively different picture of PHI enrolment determinants, especially those capturing age, language proficiency, labour force status or the number of children. We also show that PHI coverage misreporting is subsequently associated with misreporting of reasons for purchasing PHI, type of cover and length of cover.

Keywords: Health Insurance; Measurement Error; Administrative Data; Survey Misreporting; Linked Data; Australia

Suggested citation: Nguyen, H.T., Le, H.T, Connelly, L., & Mitrou, F. (2022). 'Accuracy of self-reported private health insurance coverage', Life Course Centre Working Paper Series, 2022-24. Institute for Social Science Research, The University of Queensland.

1. Introduction

Studies on health insurance coverage often rely on measures self-reported by respondents and this is the case in both developed (Propper *et al.* 2001; Frean *et al.* 2017; Bonsang & Costa-Font 2022) and developing countries (Spaan *et al.* 2012; Erlangga *et al.* 2019). The presumed accuracy of survey reports of health insurance enrolment influences how these data are used for health policy evaluations (Meyer *et al.* 2015), yet the accuracy of such measures outside the United States (US) has not been thoroughly validated (Lurie & Pearce 2021; Call *et al.* 2022). This paper aims to fill that gap in the literature by presenting the first evidence on the extent and factors associated with accuracy of private health insurance (PHI) coverage reporting in an Australian context.

The analysis in this paper relates to a large extant literature on measurement errors in survey data.[1] This body of literature has documented significant measurement errors in income (Abowd & Stinson 2013; Hurst *et al.* 2014; Bingley & Martinello 2017), employment status (Feng & Hu 2013), education (Battistin *et al.* 2014), health (Baker *et al.* 2004; Burkhauser & Cawley 2008; Cawley *et al.* 2015) and the receipt of government transfers (Meyer *et al.* 2009; Nguyen *et al.* 2021; Meyer *et al.* 2022).

Within this broadly defined literature, there is an increasing number of studies documenting measurement errors in health insurance coverage, exclusively in the context of the US and mostly limited to public health insurance in the form of Medicaid - a major public insurance program for low-income families in the US (Call *et al.* 2022). In particular, US studies have typically found underreporting of Medicaid coverage in survey data (Pascale *et al.* 2009; Call *et al.* 2013; Boudreaux *et al.* 2015; Noon *et al.* 2019; Pascale *et al.* 2019a). They have also uncovered Medicaid misreporting varies by respondent characteristics, including age, education, income and employment statuses. Several US studies have documented the (in)accuracy of PHI reporting, indicating a tendency of PHI coverage overreporting in household surveys (Cantor *et al.* 2007; Lurie & Pearce 2021). However, evidence on factors associating with misreporting of PHI coverage is relatively limited, reflecting a much smaller number of studies on the topic or limitations in data sources employed by existing studies (Pascale *et al.* 2019b; Lurie & Pearce 2021; Call *et al.* 2022).

[1] For excellent reviews of this literature, see, for instance, Bound *et al.* (2001) or Meyer *et al.* (2015).

This paper contributes to the literature by utilizing the newly available linked Australian National Health Survey and administrative population tax data to exclusively examine the accuracy in PHI reporting outside the US context. Australian literature has heavily relied on survey data to study PHI enrolment in the country (Palangkaraya & Yong 2005; Doiron *et al.* 2008; Johar *et al.* 2011; Kettlewell *et al.* 2018; Buchmueller *et al.* 2021). While there are some concerns about the accuracy of self-reported measures of PHI (Buchmueller *et al.* 2021; Liu & Zhang 2022), there is no formal validation study on this topic yet, offering an opportunity for this paper to establish itself as the first to do so. Evidence provided in this paper will be useful not only for Australian studies but also for studies from other countries which have health care systems similar to Australia's (Colombo & Tapay 2004).[2]

The exceptional data richness and sample size in this study allow us to make four other important contributions to the literature. First, we examine a much wider range of individual and family characteristics associated with misreporting of PHI coverage than was previously possible in US studies (Pascale *et al.* 2019b; Lurie & Pearce 2021; Call *et al.* 2022). This contribution is particularly beneficial since our results reveal new insights into potential reasons for PHI misreporting. Second, our data enable us to distinguish two types of PHI misreporting (i.e., false negative and false positive reporting (more on this below)). Prior US studies did not make this distinction, probably due to data constraints (Lurie & Pearce 2021; Call *et al.* 2022). Our more detailed classification of PHI misreporting coupled with the rich explanatory variable list allows us to produce new evidence that many of the characteristics that are associated with the probability of giving a false negative or a false positive report differ between these two types of misreporting. Third, for the first time in this literature, this paper directly assesses the implications of misreporting for studies using self-reported measures of PHI coverage to examine the determinants of PHI enrolment. Fourth, this paper presents novel evidence on the effect of PHI coverage misreporting on subsequent responses to other commonly asked PHI-related questions.

We show that 91% individuals correctly identify their PHI enrolment status. However, reporting errors are quite substantial as 5% of truly insured individuals self-report as

[2] Unlike the US, Australia has a compulsory universal public health insurance system (i.e., Medicare). Many other OECD countries have health care systems similar to Australia's (Colombo & Tapay 2004) but there is no validation study using data from these countries.

uninsured (i.e., the false negative rate is 5%) and 16% of truly non-insured persons self-identify as insured (i.e., the false positive rate is 16%). We find that both false positives and false negatives are correlated with a range of individual and household characteristics, including age, migration status, education, marital status, employment status and income. We additionally find that most of these characteristics influence the probability of giving a false negative or a false positive report very differently.

The results suggest that survey errors are not random, resulting in potentially important and complicated biases in multivariate analyses. We directly investigate biases in the determinants of PHI enrolment using common survey-based estimates of PHI enrolment. We find that while survey data provide a rather qualitatively accurate picture of those factors that are correlated with PHI coverage, they depict a quite quantitively different association between PHI coverage and some characteristics capturing age, language proficiency, labour force status or the number of children. Finally, we show that misreporting of PHI enrolment status in survey data is also subsequently associated with misreporting of reasons for purchasing PHI, type of cover and length of cover.

The rest of this paper is organized as follows. Section 2 describes our data and Section 3 presents main evidence on the correlates of PHI misreporting. Section 4 examines how survey error affects our understanding of PHI enrolment. In the same section, we also investigate the correlation between PHI misreporting and responses to other common PHI-related questions. Section 5 offers our conclusions and implications for future research.

2. Data

This study uses data from linked 2014-15 National Health Survey (NHS) and administrative Personal Income Tax (PIT), provided from the Australian Bureau of Statistics (ABS)'s Multi-Agency Data Integration Project (MADIP). 2014-15 NHS is a nationally representative survey conducted by the ABS during the 2014-15 financial year (i.e., between 1 July 2014 and 30 June 2015).[3] It includes 19,257 individuals, among them 14,560 are adults, in 14,723 private

[3] We do not use more recent NHSs, which are also linked to MADIP data, because they have no information on PHI (ABS 2020a). NHSs have been a popular data source to study PHI in Australia (Cameron *et al.* 1988; Cameron & Trivedi 1991; Savage & Wright 2003; Palangkaraya & Yong 2005; Doiron *et al.* 2008; Johar *et al.* 2011; Buchmueller *et al.* 2013; Kettlewell 2019b). Other survey data sources include Household, Income and Labour Dynamics in Australia (Cheng 2014; Kettlewell 2019a; Bilgrami *et al.* 2021; Buchmueller *et al.* 2021) and the 45-and-Up Study (Johar & Savage 2012; Doiron *et al.* 2014; Kettlewell *et al.* 2018). A few Australian studies have

dwellings (ABS 2017c). PHI coverage status in the 2014-15 NHS is constructed from responses to a question asking all selected persons aged 18 years and over “Apart from Medicare, [do you/does [first name]] have private health insurance?”. Thus, an adult individual is identified as being covered by PHI at the survey time if answering “Yes” to this question and not covered by PHI if “No”.

Our administrative measure of PHI coverage comes from PIT data which are provided by the Australian Tax Office (ATO) to the ABS and cover all individual income tax fillers in Australia. Because PIT data are recorded on a financial year basis, we match 2014-15 NHS with PIT recorded on the same financial year of 2014-15. In the administrative PIT data, we have information on whether an individual had PHI at any point during the 2014-15 financial year.[4]

We take the administrative PHI coverage measure to be accurate, as has been done previously in the US literature (Lurie & Pearce 2021). While administrative data may have errors (Kapteyn & Ypma 2007), personal income tax filling practices and PHI-related incentives make this unlikely in our case. Specifically, in the Australian tax filling system, information on individual income and PHI coverage is provided by a third party (e.g., employers provide income information while health insurance providers report PHI information) and this information is validated by ATO. Moreover, income tax and other PHI-related costs/benefits such as Medicare Levy Surcharge, Life Time Health Cover and premium subsidies (see, for example, AIHW (2017) for a review of these policies) are calculated using both income and PHI coverage.

18,280 individuals (95% of the original sample) in the 2014-15 NHS have been linked to the MADIP asset.[5] Among them, 10,301 individuals filled their personal income tax returns in the 2014-15 financial year and hence are observed in 2014-15 PIT data. We exclude 232 individuals aged under 18 years in the 2014-15 financial year from this sample because the question about PHI coverage was not asked for them in the 2014-15 NHS. After further

used data from PIT, the same data source as one of our data sources, to document PHI enrolment (Stavrunova & Yerokhin 2014; Kettlewell & Zhang 2021; Liu & Zhang 2022).

[4] Unfortunately, PIT data have no information about the coverage duration during this financial year.

[5] Identifiers used for linking NHS to the MADIP asset are individual name, address, birth date and gender (ABS 2020b). We experimented with additionally controlling for a variable which measures the quality of linking individuals in NHS to the MADIP asset and found it statistically insignificant in all regressions. Moreover, including this variable does not affect the estimates of other variables. As such, we do not include it in the final regressions. See Appendix Table A1 for variable description and summary statistics.

excluding 23 individuals who replied “don’t know”[6] about their PHI coverage status in 2014-15 NHS data, we have a final analytical sample of 10,013 adult individuals who appear and have valid information on PHI coverage in both datasets.

Appendix Table A2 describes factors associated with the probability that a respondent in the 2014-15 NHS is included in our final sample. As expected, because our sample focuses on tax filers and excludes those with the lowest incomes and hence are not subject to taxes, individuals included in our final sample tend to have more advantageous socio-economic backgrounds. For instance, they are more likely to have higher qualifications or better health, to be in a marital relationship, to work or to have higher income. We also observe that individuals with PHI coverage (as recorded in survey data) are more likely to be included in our sample, suggesting that the average PHI coverage rate in this sample is higher than the average rate for all Australians (AIHW 2017).

Table 1 presents unweighted (Panel A) and weighted (Panel B) sample sizes and additional statistics (Panel C) comparing PHI coverage according to the survey and administrative records for the same individuals in our sample. Unweighted statistics from survey data show that, in the 2014-15 financial year, 67% of them were covered by PHI while administrative data indicate only 64% of them were. Moreover, reporting accuracy of PHI enrolment in survey data is high with 91% of individuals displaying agreement between survey responses and administrative records. However, reporting errors are non-negligible. Particularly, 5% of individuals who self-identify as uninsured are recorded as insured in the administrative data. We denote these cases as “false negatives”, following previous studies (Bound *et al.* 2001; Meyer *et al.* 2015). By contrast, 16% of individuals who self-report as having PHI are not covered by PHI in the administrative data (hereafter denoted as “false positives”). Weighted statistics, which are derived by adjusting for survey sampling weights and reported in the last row of Table 1, depict a largely similar pattern in PHI coverage and reporting accuracy rates, suggesting that our findings are insensitive to whether we account for survey sampling weights.[7]

[6] “Don’t know” response is a potentially important issue for analysis. However, the sample size of individuals with “don’t know” responses is too small to analyze separately.

[7] We don’t adjust for survey sampling weights in regressions which control for most variables which have been used to calculate the weights (Solon *et al.* 2015). Nevertheless, the results are largely the same when we do.

Above, we found a slightly higher rate of PHI coverage in survey data than in administrative data. This finding is consistent with that in US studies which also find coverage of PHI, typically in the form of employer-sponsored health insurance, is overreported in household surveys (Cantor *et al.* 2007; Lurie & Pearce 2021). However, our findings are not in line with a common finding from US studies which typically document underreporting of Medicaid across various surveys (Pascale *et al.* 2009; Call *et al.* 2013; Noon *et al.* 2019). We further uncovered that the false positive rate is much higher than the false negative rate in Australian data. While not directly comparable, this finding is different from a commonly-reported pattern that the false negative rates are much higher than the false positive rates in a related literature on misreporting of government transfers (Meyer *et al.* 2015).

3. Factors associating with misreporting of PHI coverage

3.1. Empirical model

We turn to explore factors associating with the probability of PHI misreporting. Following previous studies (Call *et al.* 2022; Meyer *et al.* 2022), our empirical model controls for a rich list of individual and household level variables. Individual level variables include age categories, gender, Aboriginal status, migration status, self-rated English proficiency, education, marital status, general health status, mental health, disability status, previous health care utilizations, cigarette smoking status, employment status and taxable income. Household level variables consist of the number of other adults, number of children and spouse's taxable income. To control for spatial or temporal differences in reporting patterns, we also include state/territory dummies, an urban indicator, survey year and month dummies in all regressions. All explanatory variables are constructed using survey data, primarily because most of them are not available in administrative data. An exception is the respondent's and their spouse's taxable income variables, which are obtained from administrative data which are expected to contain more accurate and less missing information (Bingley & Martinello 2017).

Some variables in the above described explanatory variable list are to capture some commonly documented reasons for misreporting (Bound *et al.* 2001).[8] For instance, variables

[8] For a review, see, for example, Bound *et al.* (2001) who broadly group reasons for misreporting into three areas: cognitive process, social desirability and survey conditions. Briefly, the first area includes any factor that

representing individual cognitive process, including English proficiency, education and mental health, are to gauge the potential effects of cognitive process on misreporting (Sudman *et al.* 1996). Moreover, the inclusion of previous health service utilization that might have been associated with the use of PHI benefits is to capture their likely impact on the respondent's recalling information about their PHI coverage (Call *et al.* 2022; Meyer *et al.* 2022). Additionally, to address the differences in survey time which may affect the recall period, we control for the survey year and month dummies in all regressions (Call *et al.* 2008; Meyer & Mittag 2019).

The level of analysis is individuals because (i) PHI coverage status is recorded at an individual level in both survey and administrative data, and (ii) almost all (99%) individuals in our sample responded to the survey themselves. We examine the determinants of false negatives and the determinants of false positives separately. For the model of the determinants of false negatives, the subsample consists of those who, according to the administrative data, were covered by PHI. The sample for the false positive analysis includes those who did not have PHI in administrative data. We apply a Probit model for each regression and report average marginal effects (ME) on the chance of being a false negative or false positive reporter to facilitate the interpretation of the magnitudes.

3.2. Empirical results

We first investigate factors associating with the probability of being a false negative reporter.[9] The results (reported in Column 1 of Table 2) suggest that the probability of giving a false negative report on PHI coverage decreases with ages, although the marginal effects of many age groups are imprecisely estimated. Moreover, Australia-born or more educated individuals are statistically significantly (at 5% level or higher) less likely to be a false negative reporter. Similarly, individuals whose spouses have higher incomes are less likely to misreport that they are not covered by PHI. It is interesting to observe that while the parameter estimates in the

---

influences the cognitive process of responding a question, involving understanding the question, recalling information from memory and communicating the result. Social desirability relates to a tendency of respondents to provide socially desired answers which may or may not be true. Survey conditions refer to questionnaire design, survey mode and method which may affect the accuracy of survey data.

[9] Appendix Table A3 represents summary statistics by misreporting statuses, suggesting noticeable differences in various characteristics among four sub-groups. Moreover, the results from these simple pairwise comparisons largely agree with those obtained from regression-based analyses. This persistence in the results suggests that our findings are not driven by the potentially high multi-correlations among some explanatory variables.

respondent's and the spouse's income are both negative, the estimate of the spouse's income is much more pronounced in terms of statistical significance (i.e., only the estimate of the spouse's income is statistically significant) and magnitude (i.e., the estimate of the spouse's income is about double in size in an absolute value). By contrast, individuals with the marital status recorded as "separated" or individuals from households with more children are more likely to be false negatives. In the same vein, individuals with mental illness, smokers or unemployed individuals have a slightly higher likelihood of failing to report the true PHI coverage status because the estimates for their related characteristics are marginally statistically significant (at 10% level) and positive. However, other included individual or household characteristics, including gender and health related variables, do not statistically significantly predict the probability of giving a false negative report.[10]

Table 2 (Colum 2) further reveals various factors which are important in predicting the chance of having a false positive report. For instance, the negative and varied estimates on age categories indicate that the likelihood of giving a false positive report decreases with ages up to the age group of 58-62 years old, before increasing.[11] We additionally observe a greater probability of being a false positive reporter for individuals who are non-Aboriginal, married, have a bachelor or higher degree, were out of the labour force, or have poor English proficiency. Similarly, individuals from households with more adults are more likely to misreport that they were covered by PHI. By contrast, Australia-born individuals, smokers or individuals whose spouse has higher income are less likely to provide a false positive report.

The above-described results suggest noticeable differences in estimates of some variables by type of misreporting (i.e., false negatives or false positives) in terms of the direction, statistical significance or magnitude. For instance, estimates are statistically significant but have opposite signs (i.e., negative or positive) for variables describing bachelor qualification, smoker status and number of children in the false negative and false positive reporting

[10] Remaining results, reported in Appendix Table A4, show noticeable geographical differences in both types of misreporting of PHI coverage. Furthermore, the statistically significant estimates for some time variables suggest significant temporal differences in the probability of false negative reporting.

[11] This age profile of misreporting is consistent with that in a modified empirical model in which we introduce ages in a quadratic form. In particular, the results from this modified model (reported in Appendix Figure A1) show that the probability of being a false positive reporter decreases with ages up to the age of 50 years old, before increasing. Likewise, and in line with the baseline results, the probability of providing a false negative report decreases with ages up to the age of 65 years old, before increasing. We use age categories in the main analysis as this more flexible functional form of age is arguably better to detect any non-linear relationship between age and misreporting.

regressions. Moreover, estimates of variables representing age groups, Aboriginal status and out-of-labour-force status appear to be more statistically significant in the false positive reporting regression. Similarly, estimates for variables describing age categories, Aboriginal status, marital status and spouse's incomes are greater (in absolute terms) in the regression of false positives. Indeed, test statistics (reported in Column 3 of Table 2) confirm that estimates for some variables are statistically significantly (at least at 10% level) different in the false negative and false positive reporting regressions. These include variables capturing age categories (up to 48-52-year-old group), bachelor qualification, marital statuses classified as "divorced" or "separated", smoking status, out-of-labour-force status, number of other adults and number of children in the household. Furthermore, a test statistic for equality of false negative and false positive reporting equations reported at the bottom of Table 2 suggest that these two equations should be estimated separately.

Nevertheless, to improve the statistical power of empirical results and to provide a more general picture of factors associating with PHI misreporting, we present results in which we combine both types of misreporting as one outcome in Column 4 of Table 2. Specifically, we combine false negative and false positive reporting statuses as one and denote it as "any false" reporting. We then apply it as a dependent variable in a Probit regression for all individuals in our final analytic sample. The results indicate noticeable improvements in statistical power for some variables, probably because of a greater sample size. For example, the highly statistically significant estimates of all age categories show that the probability of PHI misreporting decreases with age up to 58-62 years old, before increasing. Moreover, the positive and statistically significant (at 1% level) estimate of the poor English proficiency variable suggests that individuals with poor local language skills are more likely to give an inaccurate report of their PHI coverage. Likewise, the estimate of the own income variable becomes statistically significant (at 5% level), indicating that individuals with higher income are less likely to misreport about their PHI coverage. By contrast, estimates of some variables, including those measuring whether an individual has a bachelor degree or was out of the labour force, the number of adults and the number of children, become less statistically significant. This drop in statistical significance levels for these variables is consistent with their differential estimates in the separate regressions presented above. The results also show that estimates for other variables in this auxiliary regression are largely like those in separate regressions in terms of the statistical significance and direction.

3.3. Discussion

The above analysis suggests that PHI reporting is generally less accurate among socioeconomically disadvantaged individuals, especially those who have lower qualifications, were born overseas, have poor English proficiency, or people in lower income households. At first glance, this finding appears to be contrary to that presented in US studies which usually find that socioeconomically disadvantaged individuals are more accurate in reporting their Medicaid coverage (Call *et al.* 2022). It should be noted that Medicaid is a public health insurance program for low-income individuals in the US. As such, socioeconomically disadvantaged people are more likely to be eligible for Medicaid. This study, by contrast, focuses on private health insurance and Australian studies have documented that individuals from more socioeconomically advantaged backgrounds are more likely to have PHI (Cameron & Trivedi 1991; Johar *et al.* 2011; Doiron & Kettlewell 2020). To this end, the US and Australian findings are consistent because they all suggest that the accurate reporting of health insurance coverage is higher for people with characteristics positively associated with the probability of having health insurance coverage.

Our finding when viewed with an oft-observed pattern of a relatively high stability of these characteristics and hence health insurance coverage overtime (Buchmueller *et al.* 2021) indicate an important role of the stability of health insurance coverage in reducing PHI misreporting. It is possible that the stability in health insurance coverage makes it easier for individuals who regularly have it to remember and subsequently recall that fact accurately (Sudman *et al.* 1996). To this end, our finding concords with the idea that misreporting is partly due to recall and retrieval problems.

We uncover a convex relationship between age and misreporting. This finding combined with a generally agreed concave relationship between age and cognitive skills (Deary *et al.* 2009; Hartshorne & Germine 2015) suggest some role that cognitive process may play in explaining PHI misreporting. Specifically, it is likely that, as compared to individuals in the middle of the age distribution, those at the two ends of the distribution may have lower cognitive skills and hence more difficulties in understanding the question or recalling information. We provide further evidence for the role of cognitive process by showing that individuals with mental health are more likely to misreport, probably because they have difficulties in understanding the question or recalling information.

Additionally, we find that individuals with poor English proficiency or overseas-born respondents are more likely to misreport about their PHI coverage. This result can be taken as evidence that comprehension of the question is among the causes of misreporting as these individuals may have difficulties in understanding the question. We provide further evidence of comprehension error where we find that individuals with higher qualifications are less likely to provide a false negative report. However, we also find that highly educated individuals, as represented by having a bachelor degree, are surprisingly more likely to make a false positive report. This finding, nonetheless, is consistent with social desirability being among the causes of misreporting, probably because these highly educated individuals might have found it more socially desirable to overreport their PHI coverage (Meyer *et al.* 2009).

In summary, our work documents both false negatives and false positives that are systematically correlated with individual and household characteristics. The results also suggest that many of these characteristics are associated with the probability of giving a false negative or a false positive report in very different ways. Moreover, the results show that the variables that consistently predict PHI misreporting support common reasons for misreporting, such as comprehension, recall or social desirability.

4. Additional results

4.1. The impact of PHI misreporting on estimates of PHI enrolment

Having explored the correlations of PHI coverage misreporting, we directly assess the effect of misreporting on estimates of PHI enrolment. Many studies have used survey data to study the determinants of PHI enrolment worldwide (Besley *et al.* 1999; Hullegie & Klein 2010; Nguyen & Leung 2013; Frean *et al.* 2017; Buchmueller *et al.* 2021). However, up till now, we know little about the implications of PHI misreporting on such estimates. Having true PHI coverage matched to survey data offers us the opportunity to examine whether the use of administrative data provides a different understanding of the factors associating with PHI enrolment from the survey data. To do this, we concurrently run two Probit regressions of the PHI enrolment binary variable as recorded from survey or administrative data on survey covariates. As has been done previously in Australian studies (Doiron *et al.* 2008; Johar *et al.*

2011; Buchmueller *et al.* 2013; Doiron & Kettlewell 2020), our list of covariates includes variables which are typically shown to be associated with the demand for health insurance (McGuire 2011). Essentially, we employ the same list of covariates as used in Section 3.1.

The results from this exercise, reported in Table 3, are largely consistent with previous Australian evidence. For instance, we find that individuals from more socioeconomically advantaged backgrounds are more likely to purchase PHI (Johar *et al.* 2011; Doiron & Kettlewell 2020). Specifically, our results indicate that individuals who are non-Aboriginal, were born in Australia, have better English proficiency, are in a marital relationship, work full-time or have higher incomes have a statistically significantly higher probability of purchasing PHI. We also uncover that individuals from households with fewer members or with members who earn higher income are more likely to have PHI. By contrast, and in line with prior evidence (Savage & Wright 2003; Doiron *et al.* 2008; Johar & Savage 2012), we find that smokers are much less likely to have PHI. Furthermore, estimates of health-related variables provide mixed evidence on the relationship between health and PHI coverage (Cameron & Trivedi 1991; Doiron *et al.* 2008; Buchmueller *et al.* 2013). On one hand, individuals with poorer general or mental health or individuals who had any outpatient treatment last year have a lower probability of being covered by PHI. On the other hand, individuals who had any inpatient treatment in last 12 months are statistically significantly more likely to have PHI.

Table 3 shows that the direction of the determinants of PHI enrolment is noticeably similar in regressions using survey or administrative data. An exception is that the marginal effects for the first two age groups (i.e., 23 to 27 and 28 to 32 years old) are negative in the regression using survey data but positive in administrative data. Indeed, the results from a Chi squared test for equality of these two coefficients from survey data and administrative data equations reported in Column 3 of Table 3 indicate that they are statistically different at 1% level. Similarly, the Chi squared test results suggest that estimates for other age groups while having the same sign are statistically significantly different (also at 1% level) between the two regressions. Likewise, according to the test results, estimates for variables measuring English proficiency, not-in-labour-force status or the number of children in the household are statistically significantly different (at least at 5% level), primarily in terms of statistical significance or magnitude, between the two regressions. The statistical significance differences among these variables are consistent with the result from a Chi squared test which is reported in the last row of Table 3 and clearly rejects equality of all estimates from survey data and administrative data equations.

Overall, the results presented in this section suggest that using survey data would provide a quite accurate (in qualitative terms) picture of factors associating with the PHI coverage. However, survey error clearly changes what we learn about PHI enrolment determinants, especially those capturing age, language proficiency, labour force status or the number of children.

4.2. Association between PHI misreporting and responses to other PHI-related questions

We next investigate the correlation between PHI misreporting and responses to other PHI-related questions. We consider responses to some commonly asked questions regarding reasons for having/not having PHI and characteristics of PHI policies, including type of membership, type of cover and length of coverage.[12] These questions are typically asked after the respondents have answered the question about their PHI coverage status (ABS 2017b; Zhang & Prakash 2021). As such, what questions are asked depends on responses to the PHI coverage question. Specifically, questions about reasons for having PHI or characteristics of PHI policies are only asked for those who self-identify as being insured. Similarly, only uninsured individuals are asked to complete the question about reasons for not having PHI. Answers to these PHI-related questions are of interest (Doiron *et al.* 2008; Buchmueller *et al.* 2013; ABS 2017a; Zhang & Prakash 2021). However, there is no evidence on how PHI misreporting affects the responses to these follow-up questions. This study thus provides the first evidence on such a relationship.

To do so, we employ a Probit regression equation in which the dependent variable is a binary one which takes the value of one if the respondents give an affirmative answer to each of the above-described PHI-related questions, and zero otherwise. In this regression, we introduce a variable describing a PHI misreporting status as an independent variable. To accommodate

[12] Specifically, reasons for having PHI are constructed from responses to a question asking an insured respondent "What are all the reasons [you are/[first name]is] covered by private health insurance?" while reasons for not having PHI are from a question asking an uninsured respondent "What are all the reasons [you are/[first name] is] not covered by private health insurance?". Type of cover is derived from responses to a question asking "Which best describes what [your/his/her] private health insurance covers?" while type of membership is from a question asking "[Are you/is [first name]] covered by family, couple, sole parent or single membership?". Length of coverage is constructed from responses to a question asking "How long [have you/has [first name]] been covered by private health insurance?".

the fact that other PHI-related questions are asked conditionally on responses to the question on PHI coverage, we necessarily redefine two variables capturing PHI misreporting statuses. Specifically, we define a dummy variable which takes the value of one if individuals self-report as having PHI in survey data but have no record of PHI coverage in administrative data (i.e., "True negative" cases),[13] and zero otherwise. This variable is only identified among those who self-identify as being insured and hence included in the regressions of reasons for having PHI or characteristics of PHI policies. Furthermore, we generate a binary variable which takes the value of one if individuals have no PHI in survey data but have PHI coverage in administrative data (i.e., "True positive" cases), and zero otherwise. As this variable is only constructed among those who self-report as being uninsured, we include it as an explanatory variable in the regressions of reasons for not having PHI. In all regressions in this section, we additionally control for a comprehensive set of individual and household level variables as described in the above sections.

The results from this experiment, reported in Table 4, show statistically significant correlations between PHI coverage misreporting, especially among those who misreport as being insured (i.e., true negative cases), and responses to other PHI-related questions. For instance, other things being equal, as compared to true PHI holders, individuals who misreport as being insured have a statistically significantly (at least at 5% level, as can be seen from Panel A) lower probability of giving some specific reasons for having PHI. These specific reasons include "Security or protection or peace of mind", "Lifetime cover or avoid age surcharge", "Choice of doctor", "Allow treatment as private patient in hospital", "Provides benefits for ancillary services or extras", "Shorter wait for treatment or concerned over public hospital" or "To gain government benefits or avoid extra Medicare levy". These individuals, by contrast, are much more likely to give some unspecific reasons for having PHI, such as "Other financial reasons" or "Other reason". Furthermore, they are much less likely to report as being covered by "Both hospital and ancillary" (versus "Hospital cover only" or "Ancillary cover only") policies or being covered by PHI for 5 years or more (see Panel C). However, we find no statistically significant association between the "true positive" misreporting status and reasons for not purchasing PHI (see Panel B). An exception is that, relative to the truly

[13] Unweighted figures from Table 1 show that 5% of individuals in the sample are identified as "true negatives" and 17% as "true positives".

uninsured persons, individuals who misreport as being uninsured are statistically significantly (at 5% level) more likely to select "Cannot afford it/too expensive" as one of main reasons for not purchasing PHI.

5. Conclusion

This study finds that reporting accuracy of PHI coverage is quite high in a nationally representative health survey in Australia, providing some good news for studies using such survey data to document PHI coverage. That said, our results also demonstrate that survey records of PHI coverage are affected by both false positive reporting error and false negative reporting error, and these reporting errors are non-random as they are systematically correlated with individual and household characteristics. Moreover, many of these characteristics are associated with the probability of giving a false negative or a false positive report in very different ways. We furthermore show that factors positively associated with PHI coverage are typically negatively correlated with the probability of misreporting. The results also show that the variables that consistently predict PHI misreporting support common reasons for misreporting, including comprehension, recall or social desirability. Our evidence of the factors associating with PHI misreporting may provide useful insights for survey designers to consider in order to improve accuracy of responses to PHI-related questions.

We also examine biases in the determinants of PHI enrolment using survey data. Our results indicate the signs of most determinants of PHI enrolment in the survey data match those in the administrative data. However, in quantitative terms, using survey data would provide a quite different picture of factors associating with the PHI enrolment, especially those capturing age, language proficiency, labour force status or the number of children. Finally, we uncover that misreporting of PHI enrolment status is also subsequently associated with misreporting of reasons for purchasing PHI, type of cover and length of cover.

Our finding of a substantial relationship between PHI coverage misreporting and a range of explanatory variables indicates that reporting errors of PHI enrolment in survey data are non-classical. These non-classical errors suggest complicated biases in other studies that use self-reported PHI enrolment as an independent variable in regressions, including those evaluating effects of PHI enrolment on health care utilization and health outcomes. To this end, further research into this form of biases, for example, by using data with more accurate measures of PHI enrolment like ours, is worthwhile. This would provide a more robust evidence base for health-related policies.

References

Abowd, J.M., Stinson, M.H., 2013. Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data. The Review of Economics and Statistics 95, 1451-1467

ABS, 2017a. Health Service Usage and Health Related Actions, Australia, 2014-15. Australian Bureau of Statistics (Ed.), Canberra

ABS, 2017b. National Health Survey 2014-15 Questionnaire. Australian Bureau of Statistics (ABS) (Ed.), Canberra

ABS, 2017c. National Health Survey: Users' Guide, 2014-15. Australian Bureau of Statistics (ABS) (Ed.), Canberra

ABS, 2020a. Integration of the 2017-18 National Health Survey and the Personal Linkage Spine. Australian Bureau of Statistics (ABS) (Ed.), Canberra

ABS, 2020b. Integration of the National Health Survey with the Multi-Agency Data Integration Project. Australian Bureau of Statistics (ABS) (Ed.), Canberra

AIHW, 2017. Private health insurance use in Australian hospitals, 2006–07 to 2015–16 Australian hospital statistics. Australian Institute of Health and Welfare (AIHW) (Ed.), Canberra

Baker, M., Stabile, M., Deri, C., 2004. What do self-reported, objective, measures of health measure? Journal of human Resources 39, 1067-1093

Battistin, E., De Nadai, M., Sianesi, B., 2014. Misreported schooling, multiple measures and returns to educational qualifications. Journal of Econometrics 181, 136-150

Besley, T., Hall, J., Preston, I., 1999. The demand for private health insurance: do waiting lists matter? Journal of Public Economics 72, 155-181

Bilgrami, A., Cutler, H., Sinha, K., Cheng, Z., 2021. The Impact of Means-Tested Premium Rebates and Tax Penalties on the Demand for Private Hospital Cover in Australia. Economic Record 97, 170-211

Bingley, P., Martinello, A., 2017. Measurement Error in Income and Schooling and the Bias of Linear Estimators. Journal of Labor Economics 35, 1117-1148

Bonsang, E., Costa-Font, J., 2022. Buying control? 'Locus of control' and the uptake of supplementary health insurance. Journal of Economic Behavior & Organization 204, 476-489

Boudreaux, M.H., Call, K.T., Turner, J., Fried, B., O'Hara, B., 2015. Measurement Error in Public Health Insurance Reporting in the American Community Survey: Evidence from Record Linkage. Health Services Research 50, 1973-1995

Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data. In: Heckman JJ & Leamer E (eds.) Handbook of econometrics. Elsevier, pp. 3705-3843.

Buchmueller, T.C., Cheng, T.C., Pham, N.T.A., Staub, K.E., 2021. The effect of income-based mandates on the demand for private hospital insurance and its dynamics. Journal of Health Economics 75, 102403

Buchmueller, T.C., Fiebig, D.G., Jones, G., Savage, E., 2013. Preference heterogeneity and selection in private health insurance: The case of Australia. Journal of Health Economics 32, 757-767

Burkhauser, R.V., Cawley, J., 2008. Beyond BMI: the value of more accurate measures of fatness and obesity in social science research. Journal Of Health Economics 27, 519-529

Call, K.T., Davern, M.E., Klerman, J.A., Lynch, V., 2013. Comparing Errors in Medicaid Reporting across Surveys: Evidence to Date. Health Services Research 48, 652-664

Call, K.T., Davidson, G., Davern, M., Brown, E.R., Kincheloe, J., Nelson, J.G., 2008. Accuracy in Self-Reported Health Insurance Coverage among Medicaid Enrollees. Inquiry: The Journal of Health Care Organization, Provision, and Financing 45, 438-456

Call, K.T., Fertig, A.R., Pascale, J., 2022. Factors associated with accurate reporting of public and private health insurance type. Health Services Research 57, 930-943

Cameron, A.C., Trivedi, P.K., 1991. The role of income and health risk in the choice of health insurance: Evidence from Australia. Journal of Public Economics 45, 1-28

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. The Review of Economic Studies 55, 85-106

Cantor, J.C., Monheit, A.C., Brownlee, S., Schneider, C., 2007. The adequacy of household survey data for evaluating the nongroup health insurance market. Health Services Research 42, 1739-57

Cawley, J., Maclean, J.C., Hammer, M., Wintfeld, N., 2015. Reporting error in weight and its implications for bias in economic models. Economics & Human Biology 19, 27-44

Cheng, T.C., 2014. Measuring the effects of reducing subsidies for private insurance on public expenditure for health care. Journal of Health Economics 33, 159-179

Colombo, F., Tapay, N., 2004. Private Health Insurance in OECD Countries: The Benefits and Costs for Individuals and Health Systems. Organization for Economic Cooperation Development (OECD) Working Paper No. 15

Deary, I.J., Corley, J., Gow, A.J., Harris, S.E., Houlihan, L.M., Marioni, R.E., Penke, L., Rafnsson, S.B., Starr, J.M., 2009. Age-associated cognitive decline. British Medical Bulletin 92, 135-152

Doiron, D., Fiebig, D.G., Suziedelyte, A., 2014. Hips and hearts: The variation in incentive effects of insurance across hospital procedures. Journal of Health Economics 37, 81-97

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

Doiron, D., Kettlewell, N., 2020. Family formation and the demand for health insurance. Health Economics 29, 523-533

Erlangga, D., Suhrcke, M., Ali, S., Bloor, K., 2019. The impact of public health insurance on health care utilisation, financial protection and health status in low-and middle-income countries: A systematic review. PloS one 14, e0219731

Feng, S., Hu, Y., 2013. Misclassification errors and the underestimation of the US unemployment rate. American Economic Review 103, 1054-70

Frean, M., Gruber, J., Sommers, B.D., 2017. Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the Affordable Care Act. Journal of Health Economics 53, 72-86

Hartshorne, J.K., Germine, L.T., 2015. When Does Cognitive Functioning Peak? The Asynchronous Rise and Fall of Different Cognitive Abilities Across the Life Span. Psychological Science 26, 433-443

Hullegie, P., Klein, T.J., 2010. The effect of private health insurance on medical care utilization and self-assessed health in Germany. Health Economics 19, 1048-1062

Hurst, E., Li, G., Pugsley, B., 2014. Are household surveys like tax forms? Evidence from income underreporting of the self-employed. Review of economics and statistics 96, 19-33

Johar, M., Jones, G., Keane, M., Savage, E., Stavrunova, O., 2011. Waiting times for elective surgery and the decision to buy private health insurance. Health Economics 20, 68-86

Johar, M., Savage, E., 2012. Sources of advantageous selection: Evidence using actual health expenditure risk. Economics Letters 116, 579-582

Kapteyn, A., Ypma, J.Y., 2007. Measurement Error and Misclassification: A Comparison of Survey and Administrative Data. Journal of Labor Economics 25, 513-551

Kettlewell, N., 2019a. Risk preference dynamics around life events. Journal of Economic Behavior & Organization 162, 66-84

Kettlewell, N., 2019b. Utilization and Selection in an Ancillaries Health Insurance Market. Journal of Risk and Insurance 86, 989-1017

Kettlewell, N., Stavrunova, O., Yerokhin, O., 2018. Premium subsidies and demand for private health insurance: results from a regression discontinuity design. Applied Economics Letters 25, 96-101

Kettlewell, N., Zhang, Y., 2021. Age penalties and take-up of private health insurance. Melbourne Institute Applied Economic & Social Research Working Paper No. 28/21

Liu, J., Zhang, Y., 2022. Elderly responses to private health insurance incentives: evidence from Australia. Melbourne Institute Working Paper No. 8/22

Lurie, I.Z., Pearce, J., 2021. Health Insurance Coverage in Tax and Survey Data. American Journal of Health Economics 7, 164-184

McGuire, T.G., 2011. Chapter Five - Demand for Health Insurance. In: Mark V. Pauly TGM & Pedro PB (eds.) Handbook of Health Economics. Elsevier, pp. 317-396.

Meyer, B.D., Mittag, N., 2019. Misreporting of Government Transfers: How Important Are Survey Design and Geography? Southern Economic Journal 86, 230-253

Meyer, B.D., Mittag, N., Goerge, R.M., 2022. Errors in Survey Reporting and Imputation and Their Effects on Estimates of Food Stamp Program Participation. Journal of Human Resources 57, 1605-1644

Meyer, B.D., Mok, W.K., Sullivan, J.X., 2009. The under-reporting of transfers in household surveys: its nature and consequences. NBER Working Paper No 15181

Meyer, B.D., Mok, W.K.C., Sullivan, J.X., 2015. Household Surveys in Crisis. Journal of Economic Perspectives 29, 199-226

Nguyen, H.T., Le, H.T., Connelly, L.B., 2021. Who's declining the “free lunch”? New evidence from the uptake of public child dental benefits. Health Economics 30, 270-288

Nguyen, T.-H., Leung, S., 2013. Dynamics of health insurance enrollment in Vietnam, 2004–2006. Journal of the Asia Pacific Economy 18, 594-614

Noon, J.M., Fernandez, L.E., Porter, S.R., 2019. Response error and the Medicaid undercount in the current population survey. Health Services Research 54, 34-43

Palangkaraya, A., Yong, J., 2005. Effects of Recent Carrot-and-Stick Policy Initiatives on Private Health Insurance Coverage in Australia*. Economic Record 81, 262-272

Pascale, J., Fertig, A., Call, K., 2019a. Validation of Two Federal Health Insurance Survey Modules After Affordable Care Act Implementation. Journal of Official Statistics 35, 409-460

Pascale, J., Fertig, A.R., Call, K.T., 2019b. Assessing the accuracy of survey reports of health insurance coverage using enrollment data. Health Services Research 54, 1099-1109

Pascale, J., Roemer, M.I., Resnick, D.M., 2009. Medicaid Underreporting in the CPS: Results from a Record Check Study. Public Opinion Quarterly 73, 497-520

Propper, C., Rees, H., Green, K., 2001. The Demand for Private Medical Insurance in the UK: A Cohort Analysis. The Economic Journal 111, C180-C200

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

Solon, G., Haider, S.J., Wooldridge, J.M., 2015. What Are We Weighting For? Journal of Human Resources 50, 301-316

Spaan, E., Mathijssen, J., Tromp, N., McBain, F., ten Have, A., Baltussen, R., 2012. The impact of health insurance in Africa and Asia: a systematic review. Bull World Health Organ 90, 685-92

Stavrunova, O., Yerokhin, O., 2014. Tax incentives and the demand for private health insurance. Journal of Health Economics 34, 121-130

Sudman, S., Bradburn, N.M., Schwarz, N., 1996. Thinking about answers: The application of cognitive processes to survey methodology. Jossey-Bass, San Francisco, CA, US.

Zhang, Y., Prakash, K., 2021. Why do Australians buy private hospital insurance? Melbourne Institute Research Insight No. 06/21

Table 1: Surveyed and administrative records of private health insurance coverage status

Administrative PHI coverage status	Survey PHI coverage status					
	No		Yes		Total	
	Number of observations	Row percentage (%)	Number of observations	Row percentage (%)	Number of observations	Row percentage (%)
Panel A: Unweighted						
No	3,024	84.49	555	15.51	3,579	100.00
Yes	321	4.99	6,113	95.01	6,434	100.00
Total	3,345	33.41	6,668	66.59	10,013	100.00
Panel B: Weighted						
No	3,923,029	82.95	806,110	17.05	4,729,139	100.00
Yes	415,370	5.41	7,267,625	94.59	7,682,995	100.00
Total	4,338,399	34.95	8,073,735	65.05	12,412,134	100.00
Panel C: Additional statistics						
	PHI coverage rate (%)		Accurate reporting rate (%)	False negative rate (%)	False positive rate (%)	Any false rate (%)
	Survey data	Administrative data	Survey data	Survey data	Survey data	Survey data
Unweighted	66.59	64.26	91.25	4.99	15.51	8.75
Weighted	65.05	61.90	90.16	5.41	17.05	9.84

Notes: Sample of matched individuals aged 18 years or over, with no missing information on all included variables. "Weighted" figures are adjusted for NHS sampling weight. "False negatives" indicate cases where individuals have PHI in administrative data but have no PHI in survey data. "False positives" indicate cases where individuals have no PHI in administrative data but have PHI in survey data. "Any false" indicates either "False negatives" or "False positives".

Table 2: Factors associated with misreporting of PHI coverage

Variable	False negatives (1)	False positives (2)	Test for equality of coefficient (p value) (3)	Any false (4)
Age from 23 to 27 [a]	1.26	-13.88***	*0.00*	-5.68***
	(1.71)	(2.43)		(1.37)
Age from 28 to 32 [a]	0.07	-11.64***	*0.01*	-5.47***
	(1.72)	(2.55)		(1.40)
Age from 33 to 37 [a]	-1.15	-16.81***	*0.00*	-8.69***
	(1.76)	(2.90)		(1.51)
Age from 38 to 42 [a]	-1.46	-15.53***	*0.01*	-8.59***
	(1.80)	(2.73)		(1.51)
Age from 43 to 47 [a]	-1.85	-18.61***	*0.00*	-9.97***
	(1.78)	(2.94)		(1.55)
Age from 48 to 52 [a]	-2.47	-17.38***	*0.02*	-10.28***
	(1.79)	(3.01)		(1.58)
Age from 53 to 57 [a]	-4.58**	-13.07***	*0.59*	-10.67***
	(1.86)	(3.00)		(1.58)
Age from 58 to 62 [a]	-3.04	-16.19***	*0.08*	-11.45***
	(1.87)	(3.29)		(1.67)
Age from 63 to 67 [a]	-2.90	-11.38***	*0.39*	-10.61***
	(1.95)	(3.48)		(1.76)
Age from 68 or over [a]	-3.98*	-5.46*	*0.53*	-8.24***
	(2.04)	(3.31)		(1.72)
Male	0.94	-1.79	*0.04*	0.24
	(0.60)	(1.26)		(0.61)
Non-Aboriginal	1.07	10.28**	*0.30*	4.88*
	(2.51)	(4.98)		(2.64)
Born in Australia	-1.76***	-3.18**	*0.69*	-3.10***
	(0.63)	(1.36)		(0.64)
Poor English proficiency	1.34	3.01*	*1.00*	2.86***
	(1.07)	(1.62)		(0.87)
Diploma/Certificate [b]	-1.32**	-0.18	*0.16*	-1.16*
	(0.65)	(1.36)		(0.67)
Bachelor or higher [b]	-3.35***	4.32***	*0.00*	-1.85**
	(0.73)	(1.56)		(0.75)
Widowed [c]	1.17	1.14	*0.77*	1.43
	(1.64)	(3.54)		(1.67)
Divorced [c]	1.85*	-2.79	*0.04*	0.66
	(1.04)	(2.35)		(1.12)
Separated [c]	2.79**	-1.74	*0.06*	1.98
	(1.32)	(3.01)		(1.43)
Married [c]	0.36	4.42**	*0.16*	2.25**
	(0.86)	(1.72)		(0.89)

Notes: Results (in average marginal effects) are from a Probit regression. Coefficient estimates, standard errors and sample means are multiplied by 100 for aesthetic purposes. Test statistics (p value) are from a Chi squared (χ^2) test for equality of coefficient in false negative and false positive reporting equations are reported in Column 3. [a], [b] and

[c] denotes “age from 18 to 22 years”, “having year 12 or below qualification”, and “never married” as the base group, respectively. Other control variables include state/territory, survey month and year dummies. Robust standard errors are in parentheses. The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Table 2: Factors associated with misreporting of PHI coverage (continued)

Variable	False negatives (1)	False positives (2)	Test for equality of coefficient (p value) (3)	Any false (4)
Poor health	0.02	-2.21	*0.43*	-0.53
	(0.94)	(1.90)		(0.95)
Mental distress	1.53*	0.56	*0.30*	1.57*
	(0.92)	(1.86)		(0.95)
Disable	0.11	-1.62	*0.33*	-0.82
	(0.60)	(1.37)		(0.65)
Inpatient treatment	-1.24	2.07	*0.08*	-0.80
	(0.85)	(1.98)		(0.91)
Outpatient treatment	-0.75	-3.63	*0.57*	-0.93
	(1.10)	(2.28)		(1.11)
Smoker	1.43*	-5.76***	*0.00*	-0.94
	(0.78)	(1.60)		(0.81)
Part-time employed [d]	0.16	-2.86*	*0.17*	-1.17
	(0.78)	(1.57)		(0.84)
Unemployed [d]	3.37*	-1.11	*0.10*	1.38
	(1.85)	(3.34)		(1.83)
Not in the labour force [d]	-0.90	5.22***	*0.02*	1.78*
	(1.08)	(1.99)		(1.07)
Number of adults in household	-0.27	1.68**	*0.05*	0.25
	(0.42)	(0.69)		(0.37)
Number of children in household	0.86***	-1.28*	*0.00*	0.25
	(0.31)	(0.67)		(0.33)
Own annual income ($100,000)	-1.48	-1.59	*0.64*	-3.76**
	(1.14)	(2.70)		(1.55)
Spouse's annual income ($100,000)	-2.93***	-7.27***	*0.83*	-6.89***
	(0.67)	(2.74)		(0.96)
Observations	6,434	3,579		10,013
Sample mean	4.99	15.51		8.75
Test for equality of two equations (p value)			*0.00*	

Notes: Results (in average marginal effects) are from a Probit regression. Coefficient estimates, standard errors and sample means are multiplied by 100 for aesthetic purposes. Test statistics (p value) are from a Chi squared test for equality of coefficient in false negative and false positive reporting equations are reported in Column 3. [d] denotes "full-time employed" as the base group. Other control variables include urban, state/territory, survey month and year dummies. Robust standard errors are in parentheses. The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Table 3: Determinants of private health insurance coverage from survey and administrative data

	Survey data	Administrative data	Test for equality of coefficient (p value)
Variable	(1)	(2)	(3)
Age from 23 to 27 [(a)]	-8.03***	2.01	*0.00*
	(3.04)	(3.10)	
Age from 28 to 32 [(a)]	-0.58	10.47***	*0.00*
	(3.00)	(3.05)	
Age from 33 to 37 [(a)]	6.41**	19.48***	*0.00*
	(3.05)	(3.09)	
Age from 38 to 42 [(a)]	2.89	14.42***	*0.00*
	(3.04)	(3.09)	
Age from 43 to 47 [(a)]	2.93	16.54***	*0.00*
	(3.07)	(3.12)	
Age from 48 to 52 [(a)]	6.05*	19.06***	*0.00*
	(3.12)	(3.16)	
Age from 53 to 57 [(a)]	12.29***	23.64***	*0.00*
	(3.16)	(3.18)	
Age from 58 to 62 [(a)]	14.44***	28.46***	*0.00*
	(3.28)	(3.31)	
Age from 63 to 67 [(a)]	19.02***	31.68***	*0.00*
	(3.53)	(3.56)	
Age from 68 or over [(a)]	18.04***	25.47***	*0.01*
	(3.64)	(3.65)	
Male	-4.13***	-3.74***	*0.55*
	(1.03)	(1.04)	
Non-Aboriginal	11.41***	11.80***	*0.88*
	(3.40)	(3.48)	
Born in Australia	9.63***	11.61***	*0.02*
	(1.16)	(1.16)	
Poor English proficiency	-6.47***	-11.74***	*0.00*
	(1.93)	(1.98)	
Diploma/Certificate [(b)]	4.80***	5.58***	*0.35*
	(1.17)	(1.19)	
Bachelor or higher [(b)]	15.08***	15.02***	*0.83*
	(1.38)	(1.37)	
Widowed [(c)]	3.55	3.95	*0.87*
	(3.13)	(3.11)	
Divorced [(c)]	-2.29	-0.63	*0.13*
	(1.81)	(1.83)	
Separated [(c)]	-3.11	-1.01	*0.14*
	(2.33)	(2.36)	
Married [(c)]	7.80***	8.03***	*0.87*
	(1.40)	(1.40)	

Notes: Results (in average marginal effects) are from a Probit regression. Coefficient estimates, standard errors and sample means are multiplied by 100 for aesthetic purposes. Test statistics (p value) are from a Chi squared (χ^2) test for equality of coefficient from survey data and administrative data equations are reported in Column 3. [(a)], [(b)] and [(c)] denotes “age from 18 to 22 years”, “having year 12 or below qualification”, and “never married” as the base group, respectively. Other control variables include urban, state/territory, survey month and year dummies. Robust standard errors are in parentheses. The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Table 3: Determinants of private health insurance coverage from survey and administrative data (continued)

Variable	Survey data (1)	Administrative data (2)	Test for equality of coefficient (p value) (3)
Poor health	-5.35***	-5.43***	*0.98*
	(1.69)	(1.70)	
Mental distress	-3.10*	-2.87*	*0.82*
	(1.69)	(1.71)	
Disable	0.85	2.10*	*0.10*
	(1.12)	(1.14)	
Inpatient treatment	6.98***	7.17***	*0.91*
	(1.59)	(1.60)	
Outpatient treatment	-4.32**	-4.72**	*0.76*
	(1.90)	(1.91)	
Smoker	-13.00***	-13.20***	*0.93*
	(1.30)	(1.31)	
Part-time employed [(d)]	-5.34***	-5.16***	*0.78*
	(1.31)	(1.30)	
Unemployed [(d)]	-5.99*	-8.61**	*0.34*
	(3.49)	(3.55)	
Not in the labour force [(d)]	-1.18	-5.95***	*0.00*
	(1.91)	(1.91)	
Number of adults in household	-1.92**	-1.89**	*0.93*
	(0.97)	(0.95)	
Number of children in household	-2.65***	-1.77***	*0.02*
	(0.56)	(0.57)	
Weekly income of respondent ($1000)	6.95***	6.68***	*0.54*
	(0.96)	(0.90)	
Weekly income of other HH members ($1000)	6.70***	6.31***	*0.39*
	(0.93)	(0.88)	
Observations	8,102	8,102	
Sample mean	66.09	64.23	
Test for equality of two equations (p value)			*0.00*

Notes: Results (in average marginal effects) are from a Probit regression. Coefficient estimates, standard errors and sample means are multiplied by 100 for aesthetic purposes. Test statistics (p value) are from a Chi squared (χ^2) test for equality of coefficient from survey data and administrative data equations are reported in Column 3. [(d)] denotes "full-time employed" as the base group. Other control variables include urban, state/territory, survey month and year dummies. Robust standard errors are in parentheses. The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Table 4: Association between PHI misreporting and responses to other PHI-related questions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Reasons for having PHI (Sample of 6,668 individuals with PHI reported in NHS)												
	Security or protection or peace of mind	Lifetime cover or avoid age surcharge	Choice of doctor	Allow treatment as private patient in hospital	Provides benefits for ancillary services or extras	Shorter wait for treatment or concerned over public hospital	Always had it or parents pay it or condition of job	To gain government benefits or avoid extra Medicare levy	Other financial reasons	Has condition that requires treatment	Elderly or getting older or likely to need treatment	Other reason
True negatives	-7.26***	-9.48***	-5.93***	-10.44***	-5.45**	-10.51***	-3.73*	-13.86***	2.40***	-1.54	0.28	5.80***
	(2.05)	(2.23)	(2.26)	(2.34)	(2.35)	(2.36)	(1.99)	(2.43)	(0.86)	(1.42)	(1.55)	(0.68)
Sample mean	68.46	22.38	33.34	47.42	44.86	42.73	27.65	25.22	4.47	9.60	14.59	4.45
Panel B: Reasons for not having PHI (Sample of 3,345 individuals without PHI reported in NHS)												
	Cannot afford it/too expensive	High risk category	Lack of value for money/not worth it	Medicare cover sufficient	Do not need medical care/in good health/have no dependents	Will not pay Medicare levy and PHI premium	Disillusionment about having to pay out of pocket costs/gap fees	Prepared to pay cost of private treatment from own resources	Pensioner/Veteran's Affairs/health concession card	Not high priority/previously included in parents cover	Other	
True positives	6.29**	1.13	1.22	-4.48*	-0.29	-1.10	-0.11	-2.08	-1.00	1.60	0.31	
	(2.83)	(0.00)	(2.38)	(2.70)	(1.95)	(1.26)	(0.00)	(1.51)	(1.29)	(1.59)	(1.48)	
Sample mean	57.73	0.89	21.02	29.45	12.94	4.60	10.40	7.03	4.87	8.82	7.83	
Panel C: Characteristics of PHI policy reported in NHS (Sample of 6,668 individuals with PHI reported in NHS)												
	Type of cover				Type of membership				Length of coverage			
	Hospital cover only	Ancillary cover only	Both hospital and ancillary cover	Insured but type of cover not known	Family	Couple	Sole parent	Single	Less than 1 year	1 year to less than 2 years	2 years to less than 5 years	5 years or more
True negatives	4.46***	4.61***	-11.62***	0.81**	1.41	2.09	-0.04	-2.89*	1.88***	2.08***	1.72	-7.20***
	(1.15)	(1.02)	(1.57)	(0.35)	(1.74)	(1.51)	(0.68)	(1.51)	(0.57)	(0.74)	(1.26)	(1.43)
Sample mean	9.37	7.62	81.76	1.24	46.27	19.89	2.49	31.34	2.52	4.24	9.76	83.47

Notes: "True negatives" indicate cases where individuals have PHI in survey data but have no PHI in administrative data. "True positives" indicate cases where individuals have no PHI in survey data but have PHI in administrative data. Results (in average marginal effects) are from a Probit regression. Marginal impact estimates, standard errors and sample means are multiplied by 100 for aesthetic purposes. All regressions control for variables as described in Table 3. Robust standard errors are in parentheses. The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Online Appendix

for refereeing purposes and to be published online

Appendix Table A1: Variable description and summary statistics

Variable	Description	Mean	SD	Min	Max	N
Age	Age at the survey time (years)	45.69	15.40	18.00	96.00	10,013
Male	Dummy variable: = 1 if male and = 0 if otherwise	0.48	0.50	0.00	1.00	10,013
Non-Aboriginal	Dummy variable: = 1 if non-Aboriginal and = 0 if otherwise	0.98	0.13	0.00	1.00	10,013
Born in Australia	Dummy variable: = 1 if born in Australia and = 0 if otherwise	0.71	0.45	0.00	1.00	10,013
Poor English proficiency	Self-rated proficiency in spoken English: = 1 if "very well", 2 "well", 3 "not well", and 4 "not at all"	1.06	0.28	1.00	4.00	10,013
Year 12 or lower	Dummy variable: = 1 if completed qualification is Year 12 or lower and = 0 if otherwise	0.32	0.47	0.00	1.00	10,013
Diploma/Certificate	Dummy variable: = 1 if completed qualification is diploma or certificate and = 0 if otherwise	0.35	0.48	0.00	1.00	10,013
Bachelor or higher	Dummy variable: = 1 if completed qualification is bachelor or higher and = 0 if otherwise	0.33	0.47	0.00	1.00	10,013
Never married	Dummy variable: = 1 if registered marital status is never married and = 0 if otherwise	0.30	0.46	0.00	1.00	10,013
Widowed	Dummy variable: = 1 if registered marital status is widowed and = 0 if otherwise	0.03	0.18	0.00	1.00	10,013
Divorced	Dummy variable: = 1 if registered marital status is divorced and = 0 if otherwise	0.11	0.31	0.00	1.00	10,013
Separated	Dummy variable: = 1 if registered marital status is separated and = 0 if otherwise	0.05	0.21	0.00	1.00	10,013
Married	Dummy variable: = 1 if registered marital status is married and = 0 if otherwise	0.52	0.50	0.00	1.00	10,013
Poor health	Dummy variable: = 1 if self-assessed health is rated as "fair" or "poor" and = 0 if otherwise	0.10	0.31	0.00	1.00	10,013
Mental distress	Dummy variable: = 1 if Kessler 10 score is categorised as high or very high distress and = 0 if otherwise	0.10	0.29	0.00	1.00	10,013
Disable	Dummy variable: = 1 if currently has a disability and = 0 if otherwise	0.32	0.47	0.00	1.00	10,013
Inpatient treatment	Dummy variable: = 1 if had any inpatient hospital treatment in last 12 months and = 0 if otherwise	0.12	0.33	0.00	1.00	10,013
Outpatient treatment	Dummy variable: = 1 if had any outpatient clinic hospital treatment in last 12 months and = 0 if otherwise	0.08	0.27	0.00	1.00	10,013
Smoker	Dummy variable: = 1 if currently smokes cigarette and = 0 if otherwise	0.15	0.36	0.00	1.00	10,013
Full-time employed	Dummy variable: = 1 if current employment status is full-time employed and = 0 if otherwise	0.56	0.50	0.00	1.00	10,013
Part-time employed	Dummy variable: = 1 if current employment status is part-time employed and = 0 if otherwise	0.25	0.44	0.00	1.00	10,013
Unemployed	Dummy variable: = 1 if current employment status is unemployed and = 0 if otherwise	0.02	0.15	0.00	1.00	10,013
Not in the labour force	Dummy variable: = 1 if current employment status is not in the labour force and = 0 if otherwise	0.16	0.37	0.00	1.00	10,013
Number of adults in household	Number of other individuals aged 18 or older in household	1.94	0.79	1.00	7.00	10,013
Number of children in household	Number of children aged 0-17 years in household	0.67	1.03	0.00	8.00	10,013
Own annual income	Own taxable income ($100,000 per financial year) – from PIT data	0.61	0.72	-2.32	32.77	10,013
Spouse's annual income	Spouse's taxable income ($100,000 per financial year) – from PIT data	0.36	0.64	0.00	10.00	10,013
Weekly income of respondent	Cash income of respondent (weekly, $1000)	0.75	3.65	-0.37	19.23	8,780
Weekly income of other HH members	Cash income of other household members (weekly, $1000)	1.58	3.97	-19.23	20.00	8,102

Notes: Sample of 10,013 individuals aged 18 or older, without missing information all above variables.

Appendix Table A2: Determinants of the probability of being matched between 2014-15 NHS and 2014-15 PIT

Variable	Sample 1	Sample 2
	(1)	(2)
Age from 23 to 27 [(a)]	3.41*	4.32**
	(1.99)	(1.99)
Age from 28 to 32 [(a)]	4.41**	5.36***
	(1.98)	(1.97)
Age from 33 to 37 [(a)]	2.10	3.27*
	(1.99)	(1.97)
Age from 38 to 42 [(a)]	4.64**	6.22***
	(2.02)	(2.00)
Age from 43 to 47 [(a)]	5.77***	6.98***
	(2.06)	(2.03)
Age from 48 to 52 [(a)]	4.36**	5.89***
	(2.00)	(1.98)
Age from 53 to 57 [(a)]	1.83	3.24*
	(1.97)	(1.97)
Age from 58 to 62 [(a)]	3.15	5.15***
	(1.97)	(1.97)
Age from 63 to 67 [(a)]	3.15	4.34**
	(2.03)	(2.06)
Age from 68 or over [(a)]	-7.55***	-2.76
	(1.96)	(1.99)
Male	0.99	1.23*
	(0.66)	(0.66)
Non-indigenous status	1.94	3.18*
	(1.90)	(1.88)
Born in Australia	1.90**	2.38***
	(0.74)	(0.75)
Poor English proficiency	-4.93***	-5.76***
	(0.82)	(0.85)
Diploma/Certificate [(b)]	5.23***	4.80***
	(0.71)	(0.71)
Bachelor or higher [(b)]	4.89***	4.34***
	(0.88)	(0.88)
Widowed [(c)]	-2.03	-1.07
	(1.32)	(1.34)
Divorced [(c)]	0.68	0.73
	(1.11)	(1.11)
Separated [(c)]	2.55	2.43
	(1.57)	(1.59)
Married [(c)]	3.24***	4.18***
	(0.97)	(0.97)

Notes: Results (in average marginal effects) are from a Probit regression with a binary dependent variable which takes the value of one if the individual appears in both 2014-15 NHS and 2014-15 PIT data, and zero if otherwise. “Sample 1” includes individuals aged 18 years or over and having no missing information on all included variables. “Sample 2” includes individuals in “Sample 1” and excludes individuals without PHI information in PIT data. Estimated marginal impacts, standard errors and sample means are multiplied by 100 for aesthetic purposes. [(a)], [(b)] and [(c)] denotes “age from 18 to 22 years”, “having year 12 or below qualification”, and “never married” as the base group, respectively. Other control variables include state/territory, survey month and year dummies. Robust standard errors are in parentheses. The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Appendix Table A2: Determinants of the probability of being matched between 2014-15 NHS and 2014-15 PIT (continued)

Variable	Sample 1	Sample 2
	(1)	(2)
Poor health	-5.13***	-5.45***
	(0.80)	(0.82)
Mental distress	0.11	0.24
	(0.91)	(0.93)
Disable	-2.49***	-2.27***
	(0.67)	(0.68)
Inpatient treatment	0.36	0.36
	(0.87)	(0.89)
Outpatient treatment	1.15	0.45
	(1.00)	(1.02)
Smoker	-1.69**	-2.68***
	(0.83)	(0.82)
Part-time employed [(d)]	0.78	0.17
	(1.07)	(1.02)
Unemployed [(d)]	-6.23***	-9.97***
	(1.79)	(1.83)
Not in the labour force [(d)]	-18.42***	-24.22***
	(0.92)	(0.89)
Have PHI [(e)]	8.96***	9.20***
	(0.65)	(0.65)
Unknown PHI status [(e)]	6.68	3.62
	(9.62)	(9.55)
Number of adults in household	0.60	0.18
	(0.55)	(0.56)
Number of children in household	2.56***	2.01***
	(0.44)	(0.45)
Weekly income of respondent ($1000)	1.05***	1.37***
	(0.35)	(0.36)
Weekly income of other household members ($1000)	1.04***	1.27***
	(0.34)	(0.35)
Rural areas	-0.72	-0.27
	(1.07)	(1.08)
Observations	11,733	10,585
Sample mean	78.91	76.63

Notes: Results (in average marginal effects) are from a Probit regression with a binary dependent variable which takes the value of one if the individual appears in both 2014-15 NHS and 2014-15 PIT data, and zero if otherwise. "Sample 1" includes individuals aged 18 years or over and having no missing information on all included variables. "Sample 2" includes individuals in "Sample 1" and excludes individuals without PHI information in PIT data. Estimated marginal impacts, standard errors and sample means are multiplied by 100 for aesthetic purposes. [(d)] and [(e)] denotes "full-time employed" and "have no PHI" as the base group, respectively. Other control variables include state/territory, urban dummy, survey month and year dummies. Robust standard errors are in parentheses. The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Appendix Table A3: Summary statistics by misreporting statuses

	False negatives			False positives		
	Yes	No	Difference (Yes - No)	Yes	No	Difference (Yes - No)
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Age	43.56	47.76	-4.21***	42.78	42.24	0.53
Male	0.52	0.47	0.05*	0.49	0.51	-0.02
Non-Aboriginal	0.99	0.99	0.00	0.99	0.97	0.02***
Born in Australia	0.69	0.74	-0.05**	0.59	0.70	-0.11***
Poor English proficiency	1.07	1.04	0.03***	1.17	1.09	0.08***
Year 12 or lower	0.36	0.26	0.10***	0.41	0.40	0.00
Diploma/Certificate	0.36	0.33	0.03	0.30	0.39	-0.08***
Bachelor or higher	0.28	0.40	-0.13***	0.29	0.21	0.08***
Never married	0.29	0.24	0.05**	0.40	0.39	0.01
Widowed	0.03	0.04	0.00	0.05	0.03	0.02***
Divorced	0.12	0.10	0.02	0.07	0.13	-0.06***
Separated	0.07	0.04	0.03***	0.03	0.06	-0.03***
Married	0.49	0.58	-0.10***	0.45	0.39	0.06***
Poor health	0.10	0.09	0.01	0.11	0.14	-0.02
Mental distress	0.12	0.07	0.05***	0.11	0.13	-0.02
Disable	0.31	0.33	-0.01	0.29	0.32	-0.03
Inpatient treatment	0.10	0.13	-0.03*	0.11	0.10	0.01
Outpatient treatment	0.06	0.08	-0.02	0.07	0.09	-0.02
Smoker	0.16	0.10	0.06***	0.13	0.25	-0.13***
Full-time employed	0.60	0.58	0.02	0.46	0.54	-0.07***
Part-time employed	0.24	0.24	0.00	0.24	0.29	-0.05**
Unemployed	0.03	0.01	0.02***	0.03	0.04	-0.01
Not in the labour force	0.13	0.17	-0.04*	0.27	0.14	0.13***
Number of adults in household	1.92	1.94	-0.03	2.12	1.90	0.22***
Number of children in household	0.83	0.65	0.18***	0.52	0.72	-0.2***
Own annual income ($100,000)	0.58	0.73	-0.15***	0.37	0.43	-0.06***
Spouse's annual income ($100,000)	0.28	0.47	-0.19***	0.13	0.18	-0.05***
Observations	321	6113		555	3024	

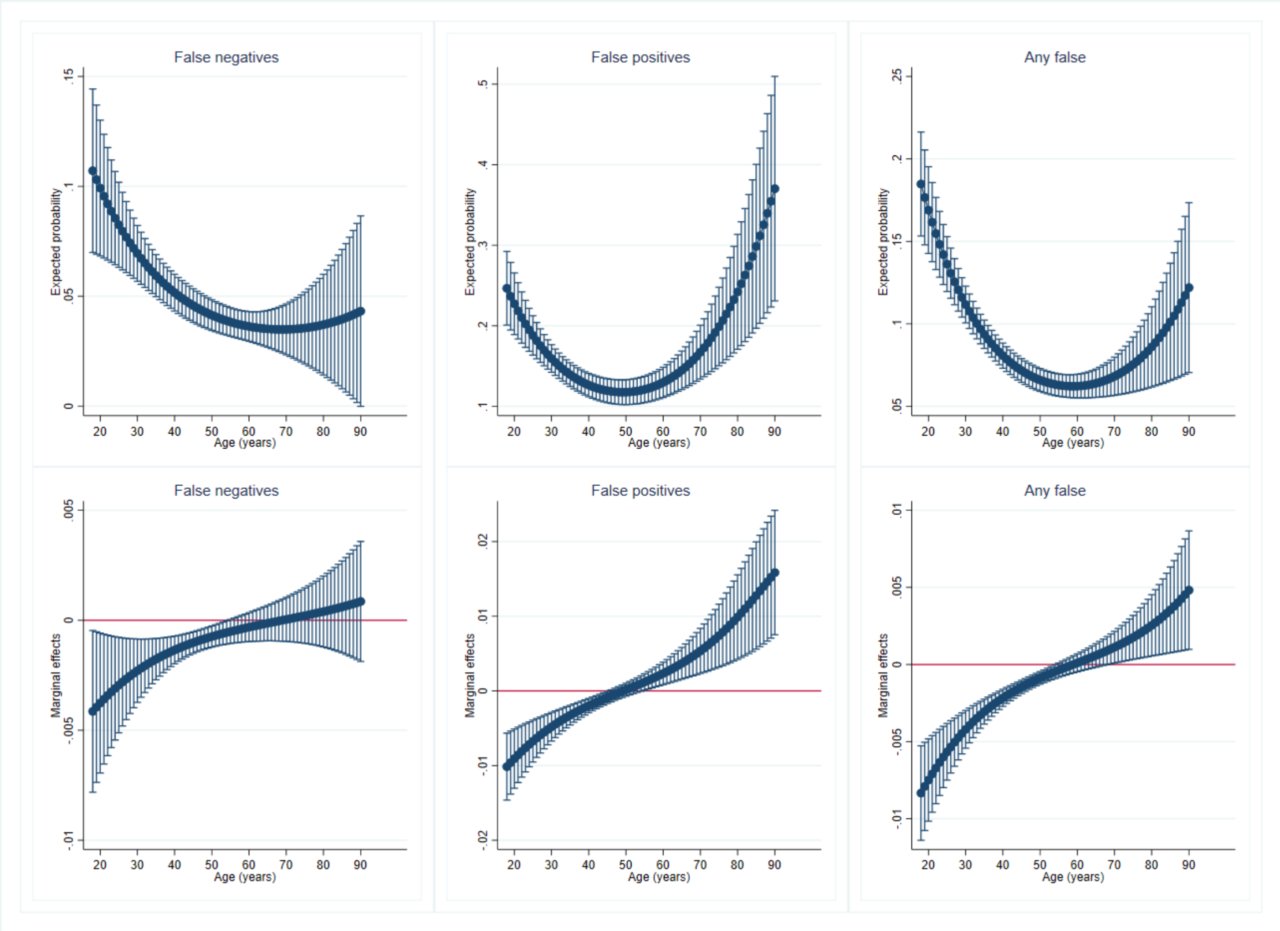
Notes: "False negatives" indicate cases where individuals have PHI in administrative data but have no PHI in survey data. "False positives" indicate cases where individuals have no PHI in administrative data but have PHI in survey data. Figures are unweighted sample mean. Statistics are calculated using the sample of the regression of false negatives (first 3 columns) or false positives (last 3 columns) on the list of all included variables. Tests are performed on the significance of the difference between the sample mean for "Yes" and "No" sub-group. The symbol * denotes statistical significance at 10% level, ** at 5% level, and *** at 1% level.

Appendix Table A4: Determinants of misreporting PHI coverage – Remaining results

Variable	False negatives (1)	False positives (2)	Test for equality of coefficient (p value) (3)	Any false (4)
Rural areas	1.24	-1.82	0.12	0.73
	(0.96)	(2.05)		(1.01)
Victoria [e]	-5.75***	-1.05	*0.00*	-3.31***
	(1.01)	(1.90)		(0.92)
Queensland [e]	-3.99***	-5.77***	*0.27*	-4.60***
	(0.93)	(2.05)		(0.99)
South Australia [e]	-3.29***	2.50	*0.00*	-1.49
	(0.93)	(2.06)		(0.97)
Western Australia [e]	-3.53***	5.55**	*0.00*	-1.84*
	(0.90)	(2.23)		(0.99)
Tasmania [e]	-7.02***	-6.18**	*0.01*	-7.09***
	(1.32)	(2.49)		(1.25)
Northern Territory [e]	-2.66**	2.15	*0.04*	-0.62
	(1.21)	(2.84)		(1.31)
Australian Capital Territory [e]	0.52	-3.98	*0.11*	-0.51
	(0.86)	(2.60)		(1.06)
February [f]	-1.87	3.27	*0.08*	0.69
	(1.33)	(2.93)		(1.37)
March [f]	0.64	2.60	*0.79*	2.30
	(1.33)	(3.07)		(1.43)
April [f]	-2.10	-0.40	*0.33*	-1.04
	(1.43)	(3.10)		(1.46)
May [f]	0.72	0.21	*0.73*	1.08
	(1.22)	(2.92)		(1.34)
June [f]	-1.66	-3.85	*0.98*	-1.90
	(1.46)	(3.14)		(1.51)
July [f]	-36.76***	-7.71	*0.00*	-7.82
	(2.50)	(10.35)		(6.90)
August [f]	-38.52***	-4.32	*0.00*	-7.92
	(2.88)	(10.82)		(7.06)
September [f]	-36.47***	-0.90	*0.00*	-5.40
	(2.84)	(10.83)		(7.07)
October [f]	-38.05***	-8.16	*0.00*	-8.73
	(2.91)	(10.84)		(7.08)
November [f]	-37.91***	-6.56	*0.00*	-8.19
	(2.86)	(10.81)		(7.06)
December [f]	-39.37***	-3.86	*0.00*	-8.35
	(2.96)	(10.83)		(7.08)
2015	-37.93***	-6.22	*0.00*	-8.63
	(2.52)	(10.36)		(6.91)
Observations	6,434	3,579		10,013
Sample mean	4.99	15.51		8.75
Test for equality of two equations (p value)			*0.00*	

Notes: Results (in average marginal effects) are from a Probit regression. Coefficient estimates, standard errors and sample means are multiplied by 100 for aesthetic purposes. Test statistics (p value) are from a Chi squared test for equality of coefficient in false negative and false positive reporting equations are reported in Column 3. [e] and [f] denotes "New South Wales" and "January" as the base group, respectively. Robust standard errors are in parentheses. The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Appendix Figure A1: Age profile of PHI misreporting – Results from a quadratic functional form of age



Notes: Results (in expected probability (the first panel) and average marginal effects (second panel)) and their 95% confidence intervals are from a Probit regression. Control variables include individual, household and local level variables, as described in the text. An exception is that age categories are now replaced by a quadratic functional form of age.