Received: 4 January 2023 | Revised: 24 July 2023 | Accepted: 25 July 2023

DOI: 10.1002/hec.4748

RESEARCH ARTICLE

Health Economics WILEY

Accuracy of self-reported private health insurance coverage

Ha Trong Nguyen[1,2] | Huong Thu Le[1,2] | Luke Connelly[3,4] | Francis Mitrou[1,2]

[1]Telethon Kids Institute, Perth, Western Australia, Australia

[2]The University of Western Australia, Perth, Western Australia, Australia

[3]The University of Queensland, Brisbane, Queensland, Australia

[4]The University of Bologna, Bologna, Emilia-Romagna, Italy

Correspondence
Ha Trong Nguyen, GPO Box 855, Perth, WA 6872, Australia.
Email: ha.nguyen@telethonkids.org.au

Funding information
Australian Research Council Centre of Excellence for Children and Families over the Life Course, Grant/Award Number: 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 11.86% of individuals misreport their PHI coverage status, with 11.57% of true PHI holders reporting that they are uninsured and 12.37% 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 enrollment using survey data. We find that, as compared to administrative data, survey data depict a quantitatively different picture of PHI enrollment determinants, especially those capturing age, gender, language proficiency, labor force status, disability status, number of children in the household, or household income. 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

administrative data, Australia, health insurance, linked data, measurement error, survey misreporting

JEL CLASSIFICATION

C81, I13

1 | INTRODUCTION

Studies on health insurance coverage often rely on self-reported survey measures and this is the case in both high income (Bonsang & Costa-Font, 2022; Frean et al., 2017; Propper et al., 2001) and low/middle income countries (Erlangga et al., 2019; Spaan et al., 2012). In the absence of potentially more reliable administrative data sources, accurately measuring health insurance coverage in survey data is important in helping assess the socio-economic status of target populations, health insurance take-up, the distributional effects of public programs and health insurance impacts (Meyer et al., 2015). However, the accuracy of self-reported health insurance measures outside the United States (US) has not been thoroughly validated (Call et al., 2022; Lurie & Pearce, 2021). 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; Bingley & Martinello, 2017; Bollinger & Tasseva, 2023; Hurst et al., 2014; Jenkins & Rios-Avila, 2023), employment status (Feng & Hu, 2013), education (Battistin et al., 2014), health (Baker et al., 2004; Burkhauser & Cawley, 2008; Cawley et al., 2015; Johnston et al., 2009) and the receipt of government transfers (Meyer et al., 2009, 2022; Nguyen et al., 2021).

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 (Boudreaux et al., 2015; Call et al., 2013; Noon et al., 2019; Pascale et al., 2009, 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 associated 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 (Call et al., 2022; Lurie & Pearce, 2021; Pascale et al., 2019b).[2]

This paper contributes to the literature by utilizing the newly available linked Australian National Health Survey and administrative population income tax data to exclusively examine the accuracy in PHI reporting outside the US context. Australian literature has heavily used self-reported PHI measures as dependent (Bilgrami et al., 2021; Buchmueller et al., 2021; Cameron & Trivedi, 1991; Doiron et al., 2008; Ellis & Savage, 2008; Johar et al., 2011; Kettlewell et al., 2018; Palangkaraya & Yong, 2005) or independent variables (Cameron et al., 1988; Cheng, 2014; Doiron et al., 2014; Doiron & Kettlewell, 2018; Eldridge et al., 2017; Kettlewell, 2019b; Savage & Wright, 2003; Srivastava et al., 2017). While there are some concerns about the accuracy of self-reported measures of PHI coverage (Buchmueller et al., 2021; Liu & Zhang, 2022), there has been no formal validation study on this topic prior to our current study.

It is important to have evidence from Australia since US evidence may not be readily applicable to Australia due to apparent differences in the respective health systems of each country. For instance, unlike the US, Australia has a compulsory universal public health insurance system, known as Medicare, which provides all Australians with publicly funded zero-price access to public hospitals, subsidized medical services supplied by private medical practitioners, and subsidized prescribed medicines, regardless of personal income or wealth status (Connelly et al., 2010; Duckett & Nemet, 2019).[3] Medicaid is similar to a welfare program, and as such, patterns of misreporting Medicaid coverage have been found to be similar to patterns of misreporting other types of welfare program receipt (Call et al., 2022; Celhay et al., 2022). Therefore, it would be a mistake to generalize from studies of Medicaid coverage to PHI coverage. Evidence provided in this paper will be useful not only for Australian studies but also for studies from many other countries that have health care systems similar to Australia's (Colombo & Tapay, 2004), whereby there are no validation studies using data from these countries.

The exceptional richness of the data and large sample size of 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 that are associated with the misreporting of PHI coverage than has previously been possible in US studies (Call et al., 2022; Lurie & Pearce, 2021; Pascale et al., 2019b). 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 (Call et al., 2022; Lurie & Pearce, 2021). Separately estimating determinants of false positives and negatives is necessary for error corrections and important to understand biases (Meyer et al., 2022). Moreover, our more detailed classification of PHI misreporting coupled with the rich explanatory variable list allows us to produce new evidence showing that many of the characteristics 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 the literature, this paper directly assesses the implications of misreporting for studies using self-reported measures of PHI coverage to examine the determinants of PHI enrollment. Fourth, this paper presents novel evidence on the association between misreporting of PHI coverage and subsequent responses to other commonly asked PHI-related survey questions. This evidence is important because responses to these follow-up questions often are of interest (ABS, 2017a; Buchmueller et al., 2013; Viney et al., 2006; Zhang & Prakash, 2021), but there is no evidence on how such responses depend on the accuracy of self-reported PHI coverage.

We show that 88% individuals correctly identify their PHI enrollment status. However, reporting errors are quite substantial as 11.57% of truly insured individuals self-report as uninsured (i.e., the false negative rate is 11.57%) and 12.37% of truly non-insured persons self-identify as insured (i.e., the false positive rate is 12.37%). We find that both false positives and false negatives are correlated with a range of individual and household characteristics, including age, migration status, English proficiency, education, marital status, smoking status, employment status and household 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 enrollment using common survey-based estimates of PHI enrollment. 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, gender, language proficiency, labor force status, disability status, the number of children in the household or household income. Finally, we show that misreporting of PHI enrollment 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 enrollment. In the same section, we also investigate the correlation between PHI misreporting and responses to other common PHI-related survey questions. Section 5 offers our conclusions and implications for future research.

2 | DATA AND SAMPLE

2.1 | Data

This study uses data from two sources: the 2014–15 National Health Survey (NHS) and Personal Income Tax (PIT) dataset. These two datasets are provided from the Australian Bureau of Statistics (ABS)'s Multi-Agency Data Integration Project (MADIP). We remotely accessed and analyzed the de-identified data via a virtual machine administered by the ABS. MADIP combines information on government payments, income and taxation, employment, health, education, and population demographics (including the Census) over time. Data were linked by the ABS via the Person Linkage Spine, a person-level identification key that broadly covers the resident population of Australia from 2006 onwards (ABS, 2023). The ABS links individual records deterministically, using the individual's first name, last name, address, birth date and gender as key identifiers. The 2014–15 NHS, which has been probabilistically linked to the MADIP asset (ABS, 2020b), is a nationally representative survey conducted by the ABS during the 2014–15 financial year (i.e., between July 1, 2014 and June 30, 2015).[4] It collects information from face-to-face interview from usual residents of private dwellings in Australia. Within each sampled private dwelling, individuals in scope of the survey comprise an adult and a child (if any). The 2014–15 NHS includes 19,257 individuals, among them 14,560 are adults, in 14,723 private dwellings (ABS, 2017c).

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 filers 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. Specifically, our administrative PHI coverage indicator takes the value of one if an individual had an appropriate level of private patient hospital cover as recorded in the PIT data at any point during the 2014–15 financial year, and zero otherwise.[5]

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?". To match with the administrative PHI coverage measure which only includes hospital cover, we assign individuals as being covered by PHI in the survey data if they (i) answer "Yes" to this question, and (ii) report that they have either "hospital cover only" or "both hospital and ancillary cover".[6] Moreover, we classify individuals as being uncovered by PHI in the survey data for the purposes of this study if they (i) respond "No" to the above question, or (ii) answer "Yes" to the same question but report they have an "ancillary cover only". This definition of the survey PHI coverage measure helps minimize any error stemming from imperfect alignment of concepts in the survey and administrative measures.[7]

We take the administrative PHI coverage measure to be accurate, as has been done previously in the US literature (Lurie & Pearce, 2021). While linked administrative variable may have errors (see, for instance, Jenkins and Rios-Avila (2023) for an excellent discussion on types of measurement errors), personal income tax filling practices and PHI-related incentives make this unlikely in our case. Specifically, in the Australian tax filling system, it is compulsory for all tax filers to complete a PHI section asking about their PHI coverage, among other details (see ATO (2015) for an example of a paper-based tax return form). Furthermore, health insurance providers are legally mandated to provide PHI information of their customers to the ATO and, as in the financial year 2014–15, this information was automatically pre-filled for 97% of tax filers who lodged their tax returns online or via their agents (ATO, 2022). The fact that the linked administrative PHI coverage measure is provided by a third party (i.e., insurance providers) and this measure is automatically linked for almost all tax filers alleviate a concern that tax filers may incorrectly report their PHI coverage status in the administrative data.

Moreover, all PHI-related costs/benefits such as Medicare Levy Surcharge, Lifetime Health Cover and premium subsidies (see, e.g., Duckett and Nemet (2019) for a review of these policies[8]) are calculated during the income tax lodgement process basing on the individual's PHI coverage status. Specifically, how much individuals pay Medicare Levy Surcharge and Lifetime Health Cover loading during the financial year depends on whether they are covered by PHI during that year. Likewise, and as expected, PHI premium subsidies can only be calculated for those with an appropriate PHI coverage. These PHI related costs/benefits thus create strong legal and financial incentives for all related parties, including tax filers, insurance providers and ATO, to get PHI coverage information right in the administrative data. Finally, all administrative measures used in this study, including taxable income and PHI coverage, have been verified by the ATO.

The mismatch in reference periods in the administrative and survey data, which has been labeled "reference period error" by Jenkins and Rios-Avila (2023), may bias our estimated error rates. Specifically, our administrative and survey data refer to different time periods, with the administrative data recording coverage at any time in the year and the survey asking about current coverage. Due to this reference period mismatch, we may incorrectly identify some false negative cases for individuals who acquire health insurance later in the year or whose plan ran out before the survey interview. However, the likely bias is probably small because only 2.5% of insured individuals in our sample report that they have been covered by PHI for less than 1 year (see statistics reported at the bottom of Table 4). While we cannot directly assess the bias from this error source, mainly due to data unavailability,[9] we include variables capturing survey administration time in all relevant regression models to account for the potential impact of reference period error on our results.

There is a concern that administrative PHI coverage reporting errors may come from the insurance provider, for instance, because some customers may have wrong information on file with their insurance provider. Two following observations lessen this concern. First, at the time of enrollment, customers of PHI providers are required to supply their unique tax file number which then can be used to link their PHI records with their tax data. Using the unique tax file numbers which are common in both PHI records and income tax data would reduce the linkage error, for instance, as compared to an alternative and popular data linkage method which uses a combination of other identifiers, such as date of birth, names or addresses (ABS, 2020b). Second, as described above, PHI-related policies create strong legal and financial incentives for all related parties, including tax filers and PHI providers, to get PHI coverage right in the administrative data.

Another potential source of administrative variable errors arises when some types of PHI (e.g., a PHI policy covered by an overseas provider[10] and hence is not eligible for PHI-related costs/benefits described above) are reported in the survey but not in the administrative data (Jenkins & Rios-Avila, 2023). Were this the case, we may over-estimate false positive rates. While this error source cannot be ruled out, the way we construct the sample (i.e., by taking only those who are linked between the 2014–15 NHS and PIT datasets) reduces the likelihood that this is a substantive problem. Specifically, we use a sample of linked tax filers who have strong financial incentives to obtain coverage from eligible PHI providers. These same individuals are thus less likely to respond that they have a PHI policy which is not listed in the administrative data.

Finally, there is still a concern about linkage error which arises when a survey respondent is linked to the wrong individual in the administrative data (Meyer et al., 2022). Indeed, the probabilistic linkage algorithm used to merge the 2014–15 NHS to PIT data is likely to result in some linkage error (ABS, 2020b). It is unclear ex ante whether linkage error corresponds to under- or over-estimation of PHI coverage. As has been done in almost all previous studies (Jenkins & Rios-Avila, 2023), we do not explicitly model this type of measurement error, primarily due to unavailability of suitable information in the data. Fortunately, our data contain a measure of the linkage quality between NHS participants and the MADIP asset, which we include in all relevant regressions to address any concern that linkage error, however small, could be driving our results.

Overall, notwithstanding the two potentially minor issues arising from imperfect alignment of concepts in the survey and administrative measures and reference periods, the combination of income tax filling practices, PHI-related incentives and "high and good quality linkage rates" (ABS, 2020b) suggest that the administrative PHI coverage measure used in this study is sufficiently error-free, allowing us to use it as the true PHI coverage indicator. Moreover, our sampling choice and selection of explanatory variables (more on this below) will ease concerns about other remaining sources of administrative data errors.

2.2 | Sample

18,280 individuals (95% of the original sample) in the 2014–15 NHS have been linked to the MADIP asset. Among them, 10,301 individuals filled their personal income tax returns in the 2014–15 financial year and hence are observed in the 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. We further exclude 23 individuals who replied "don't know" about their PHI coverage status in 2014–15 NHS data because the sample size of individuals with "don't know" responses is too small to analyze separately. For a similar reasoning, we additionally drop 83 individuals who responded "insured but type of cover not known" to the PHI cover type question. After dropping 45 more observations with missing information on included variables (more on this below),[11] we have a final analytical sample of 9919 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 low-income individuals who are not subject to personal income tax, the individuals included in our final sample tend to have more advantageous socio-economic backgrounds than those who are excluded from the sample. For instance, included individuals 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 the survey data) are more likely to be included in our sample, suggesting that the PHI coverage rate in this sample is higher than the average rate for all Australians. Because individuals with more advantageous socio-economic backgrounds are over-represented in our sample, the results from this study may not be generalized to the whole population. However, the results are particularly relevant as individuals of this demographic are typically the target population of public policy aimed at increasing PHI coverage to augment publicly funded healthcare (AIHW, 2017).

2.3 | Descriptive statistics

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, 61% of them were covered by PHI while administrative data indicate only 64% of them were. Moreover, reporting accuracy of PHI enrollment in survey data is high with 88% of individuals displaying agreement between survey responses and administrative records. However, reporting errors are non-negligible. Particularly, 11.57% 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, 12.37% 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.[12]

Above, we found a slightly higher rate of PHI coverage in administrative data than in survey data. This finding is consistent with a common finding from US studies which typically document underreporting of Medicaid across various surveys (Call et al., 2013; Noon et al., 2019; Pascale et al., 2009). However, our findings are not in line with those 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). We further uncovered that the false positive rate is somewhat 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 includes a rich list of individual and household level explanatory 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, and employment status. Household level variables consist of the number of other adults, number of children and taxable income (and its square to capture a potential non-linear relationship).[13] To account for spatial or temporal differences in reporting patterns, we also include state/territory dummies, an urban indicator, survey month-year dummies in all regressions. Finally, as explained above, we include a variable describing the quality of linking individuals in NHS to the MADIP asset to account for a potential relationship between data linkage error and PHI reporting errors.

All explanatory variables are constructed using survey data, primarily because most of them are not available in administrative data. An exception is the household taxable income variable, which is obtained from administrative data which are expected to contain more accurate and less missing information (Meyer & Mittag, 2019b, 2021). For other self-reported variables, there

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	3109	87.63	439	12.37	3548	100.00
Yes	737	11.57	5634	88.43	6371	100.00
Total	3846	38.77	6073	61.23	9919	100.00
Panel B: Weighted						
No	4,029,831	85.26	654,429	13.85	4,726,522	100.00
Yes	848,114	11.06	6,733,597	87.81	7,668,003	100.00
Total	4,877,946	39.36	7,388,026	59.61	12,394,526	100.00

	Unweighted	Weighted
Panel C: Additional statistics		
PHI coverage rate (%)		
Survey data	61.23	59.61
Administrative data	64.23	61.87
False negative rate (%)	11.57	11.06
False positive rate (%)	12.37	13.85
Any false rate (%)	11.86	12.12

Note: 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".

Abbreviation: PHI, private health insurance.

are concerns about their accuracy (Bound et al., 2001; Meyer et al., 2015). These concerns are alleviated for some variables used in this study for two main reasons. First, variables such as age, gender and residential location are obtained from administrative data because they are used as identifiers to link the 2014–15 NHS to PIT data (ABS, 2020b) and we use a sample of linked 2014–15 NHS-PIT data. Second, a recent validation study shows that some health related variables in the 2014–15 NHS align reasonably well with administrative health related indicators (ABS, 2020c).

Some variables in the above described explanatory variable list are to capture some commonly documented reasons for misreporting (Bound et al., 2001; Celhay et al., 2022).[14] For instance, variables 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 administrative time which may affect the recall period, we include the survey month-year dummies (e.g., August 2014, September 2014 or June 2015, with July 2014 being set as the baseline group) in all regressions (Call et al., 2008; Meyer & Mittag, 2019a).

The level of analysis is individuals because (i) PHI coverage status is recorded at an individual level in both survey and administrative data,[15] 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.[16] The results (reported in Column 1 of Table 2) suggest that individuals who were born in Australia, have a bachelor or higher degree, had an inpatient treatment in the previous year or were out of the labor force are statistically significantly (at 5% level or higher) less likely to be a false negative reporter. Similarly, individuals from higher income households are less likely to misreport that they are not covered by PHI. It is interesting to observe that while the parameter estimate of the income variable is negative, the estimate of the income squared variable is positive and statistically significant at the 1% level, suggesting a non-linear relationship between income and the probability of giving a false negative report. By contrast, individuals with poor English proficiency, individuals with the marital status recorded as “separated” or “divorced”, smokers, unemployed individuals or individuals from households with more children are more likely to be false negatives because the estimates for their related characteristics are positive and statistically significant (at 5% level or higher). 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.[17]

Table 2 (Column 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 43–47 years old, before increasing.[18] We additionally observe a greater probability of being a false positive reporter for individuals who have poor English proficiency, have a bachelor or higher degree, were out of the labor force, or live in households with more adults. By contrast, individuals who were born in Australia, are divorced, smoke, worked part-time, reside in higher income households 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, divorced status, smoker status and out-of-labor-force status in the false negative and false positive reporting regressions. Moreover, estimates for variables describing separated-marital status, previous inpatient treatment status, unemployment status, and number of children are more statistically significant in the regression of false negatives. By contrast, estimates of variables representing age groups, part-time-employment status and number of adults in the household appear to be more statistically significant in the false positive reporting regression. 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 58-62-year-old group), bachelor qualification, marital statuses classified as “divorced” or “separated”, inpatient treatment status, smoking status, all employment statuses, 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.

TABLE 2 Factors associated with misreporting of private health insurance (PHI) coverage.

Variable	False negatives (1)	False positives (2)	Test for equality of coefficient in false negative and false positive equations (*p* value) (3)	Any false (4)
Age from 23 to 27[a]	8.43***	−9.71***	*0.00*	−2.36
	(2.78)	(2.17)		(1.73)
Age from 28 to 32[a]	5.38*	−8.80***	*0.00*	−2.67
	(2.77)	(2.30)		(1.75)
Age from 33 to 37[a]	1.47	−10.80***	*0.00*	−5.92***
	(2.84)	(2.55)		(1.84)
Age from 38 to 42[a]	2.85	−12.34***	*0.00*	−5.82***
	(2.85)	(2.50)		(1.84)
Age from 43 to 47[a]	1.61	−15.38***	*0.00*	−7.05***
	(2.82)	(2.81)		(1.86)
Age from 48 to 52[a]	2.92	−14.80***	*0.00*	−6.31***
	(2.81)	(2.87)		(1.87)
Age from 53 to 57[a]	−0.19	−9.65***	*0.02*	−7.09***
	(2.83)	(2.75)		(1.90)
Age from 58 to 62[a]	−0.22	−11.91***	*0.00*	−7.49***
	(2.90)	(2.96)		(1.97)
Age from 63 to 67[a]	−1.10	−5.74*	*0.28*	−6.84***
	(3.04)	(3.09)		(2.09)
Age from 68 or over[a]	−3.70	0.72	*0.31*	−4.53**
	(3.22)	(2.91)		(2.11)
Male	−0.83	−0.53	*0.83*	−1.02
	(0.84)	(1.10)		(0.68)
Non-indigenous status	−0.35	8.53*	*0.16*	5.80*
	(3.67)	(5.17)		(2.99)
Born in Australia	−4.75***	−2.43**	*0.14*	−3.54***
	(0.92)	(1.24)		(0.75)
Poor English proficiency	3.72**	3.05**	*0.76*	3.22***
	(1.58)	(1.42)		(1.06)
Diploma/certificate[b]	−0.43	−0.41	*0.99*	−0.03
	(0.97)	(1.24)		(0.78)
Bachelor or higher[b]	−5.10***	4.59***	*0.00*	−1.34
	(1.04)	(1.38)		(0.85)
Widowed[c]	4.67*	−0.38	*0.20*	3.13
	(2.45)	(3.12)		(1.97)
Divorced[c]	3.07**	−5.30**	*0.00*	0.83
	(1.52)	(2.32)		(1.28)
Separated[c]	4.37**	−1.27	*0.10*	2.55
	(2.02)	(2.73)		(1.64)
Married[c]	1.69	1.89	*0.92*	2.55***
	(1.25)	(1.50)		(0.98)
Poor health	0.70	−2.29	*0.19*	−0.76
	(1.39)	(1.78)		(1.13)

TABLE 2 (Continued)

Variable	False negatives (1)	False positives (2)	Test for equality of coefficient in false negative and false positive equations (*p* value) (3)	Any false (4)
Mental distress	1.60	−1.69	*0.15*	0.49
	(1.39)	(1.79)		(1.13)
Disable	−0.36	−1.56	*0.43*	−0.91
	(0.90)	(1.25)		(0.75)
Inpatient treatment	−2.82**	0.97	*0.08*	−1.48
	(1.29)	(1.76)		(1.07)
Outpatient treatment	−0.44	−2.48	*0.44*	−0.93
	(1.60)	(2.09)		(1.29)
Smoker	4.53***	−4.08***	*0.00*	−0.23
	(1.16)	(1.47)		(0.93)
Part-time employed[d]	0.11	−4.04***	*0.02*	−1.82**
	(1.03)	(1.39)		(0.85)
Unemployed[d]	6.44**	−3.10	*0.02*	0.93
	(2.70)	(3.06)		(2.06)
Not in the labor force[d]	−4.79***	3.57**	*0.00*	−1.12
	(1.48)	(1.70)		(1.11)
Number of adults in household	−0.11	1.30**	*0.10*	0.49
	(0.59)	(0.62)		(0.43)
Number of children in household	1.49***	−0.74	*0.00*	0.56
	(0.48)	(0.61)		(0.38)
Household annual income (admin)	−7.66***	−5.63***	*0.38*	−6.81***
	(1.10)	(2.00)		(0.82)
Household annual income squared	0.23***	0.85***	*0.06*	0.21***
	(0.03)	(0.32)		(0.03)
Observations	6371	3548		9919
Sample mean	11.57	12.37		11.86
Test for equality of false negative and false positive equations (*p* value)			*0.00*	

Note: Results (in average marginal effects) are from a Probit regression. Coefficient estimates, standard errors and sample means are multiplied by 100 for esthetic 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 italic in Column 3. Other explanatory variables include state/territory, survey month-year dummies. Robust standard errors are in parentheses.

[a]Age from 18 to 22 years as the base group.

[b]Having year 12 or below qualification as the base group.

[c]Never married as the base group.

[d]Full-time employed as the base group.

The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

Nevertheless, to provide a more general picture of factors associated with PHI misreporting, as has been done in the previous US studies (Call et al., 2022; Lurie & Pearce, 2021; Pascale et al., 2019b), 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. Likewise, the estimate of the married variable becomes statistically significant (at 1% level), indicating that married individuals are more likely to misreport about their PHI coverage. By contrast, estimates of some variables, including those measuring whether an individual has a bachelor degree or smokes, divorced or separated marital states, 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, including those representing the migration status, poor English proficiency, and income, 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; Doiron & Kettlewell, 2020; Johar et al., 2011). 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 findings align with the broader literature on misreporting in government program receipt. This literature often suggests that characteristics predicting program uptake also predict more accurate reporting (Meyer et al., 2022).

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; Drake et al., 2022) 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.

Additionally, we find that individuals with poorer 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 difficulty 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 or above, 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 over-report their PHI coverage (Meyer et al., 2009).

In summary, our work documents both false negatives and false positives 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.

The finding of a systematic correlation between PHI misreporting and individual characteristics suggests that some methods which base on the assumption of no false positives to correct the misreporting when modeling PHI enrollment choices do not work well (Hausman et al., 1998; Mittag, 2019). Nevertheless, coefficient estimates of factors associating with misreporting presented here could be used to correct for binary PHI enrollment models, employing methods documented in Bollinger and David (2001) or Meyer and Mittag (2017).

Our finding that around 12% of individuals in our data misreport their PHI coverage status complicates the estimation of treatment effects. This misreporting rate is non-negligible because Kreider (2010) shows that even with health insurance misreporting rates of less than 2%, the coefficient estimate obtained from the contaminated data can be seriously biased. Moreover, the finding of the systematic association between PHI misreporting and various individual characteristics indicates that measurement error in the potentially endogenous PHI dependent variable is not classical. This finding violates the assumption of most methods correcting for misreporting and suggests that Instrumental Variable (IV) methods applied to the binary endogenous PHI variable are unlikely to give consistent treatment estimates (DiTraglia & García-Jimeno, 2019; Meyer et al., 2009; Nguimkeu et al., 2019). Indeed, Nguimkeu et al. (2019) demonstrate that IV estimates of treatment effects can be substantially biased.[19] However, it is unclear how this systematic survey error affects estimates of PHI treatment impacts. Future research should explore the magnitude and direction of such bias.[20]

4 | ADDITIONAL RESULTS

4.1 | The impact of PHI misreporting on estimates of PHI enrollment

Having explored the correlations of PHI coverage misreporting, we directly assess the effect of misreporting on estimates of PHI enrollment. Many studies have used survey data to study the determinants of PHI enrollment worldwide (Besley et al., 1999; Buchmueller et al., 2021; Frean et al., 2017; Hullegie & Klein, 2010; Nguyen & Leung, 2013). As documented above, there are numerous Australian studies employing self-reported PHI measures as a dependent variable (Bilgrami et al., 2021; Buchmueller et al., 2021; Cameron & Trivedi, 1991; Doiron et al., 2008; Ellis & Savage, 2008; Johar et al., 2011; Kettlewell et al., 2018; Palangkaraya & Yong, 2005). 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, for the first time in this literature, whether the use of administrative data provides a different understanding of the factors associating with PHI enrollment from the survey data. To do this, we concurrently run two Probit regressions of the PHI enrollment binary variable as recorded from survey or administrative data on survey covariates. As has been done previously in Australian studies (Buchmueller et al., 2013; Doiron et al., 2008; Doiron & Kettlewell, 2020; Johar et al., 2011), 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 (Doiron & Kettlewell, 2020; Johar et al., 2011). Specifically, our results indicate that individuals who are non-Aboriginal, were born in Australia, have better English proficiency, or have higher household incomes have a statistically significantly higher probability of purchasing PHI. By contrast, and in line with prior evidence (Doiron et al., 2008; Johar & Savage, 2012; Savage & Wright, 2003), 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 (Buchmueller et al., 2013; Cameron & Trivedi, 1991; Doiron et al., 2008). On one hand, individuals with poorer general 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 enrollment is noticeably similar in regressions using survey or administrative data. An exception is that the marginal effects for some age groups between 28 and 52 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 the coefficients of these age groups 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 describing gender, English proficiency, divorced marital status, disability status, not-in-labor-force status, the number of children in the household and household income 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, despite relatively high misreporting rates in survey data, using survey data would provide quite an accurate (in qualitative terms) picture of factors associated with PHI coverage. This finding aligns with the results of a previous study by Meyer et al. (2022), which investigated the misreporting of the Supplemental Nutrition Assistance Program (SNAP) in the US. Similarly, that study revealed that the qualitative conclusions drawn from highly contaminated survey data remained largely consistent with those obtained from administrative data. Moreover, and in line with that in Meyer et al. (2022), our finding is consistent with the previous finding that the characteristics that explain PHI enrollment also predict more accurate reporting of PHI status. The findings from Australian and US studies, while being conducted in distinct contexts, conform to theoretical predictions proposed by Meyer and Mittag (2017), suggesting that contaminated survey data may produce important qualitative conclusions.

However, as shown above, survey error clearly quantitatively changes what we learn about PHI enrollment determinants, especially those variables capturing age, gender, language proficiency, disability status, labor force status, the number of children and household income. Among the previously mentioned variables, two variables describing language proficiency and household income also predict reporting worse (i.e., variables with same signs in both false positives and false negatives regressions, as reported in Section 3.2). Similarly, among the variables which predict reporting more or less (i.e., those with

TABLE 3 Determinants of private health insurance coverage from survey and administrative data.

Variable	Survey data (1)	Administrative data (2)	Test for equality of coefficient in two equations (*p* value) (3)
Age from 23 to 27[a]	−14.95***	−6.31***	*0.00*
	(2.51)	(2.41)	
Age from 28 to 32[a]	−8.15***	0.40	*0.00*
	(2.54)	(2.44)	
Age from 33 to 37[a]	−1.94	6.33**	*0.00*
	(2.63)	(2.54)	
Age from 38 to 42[a]	−7.08***	1.59	*0.00*
	(2.63)	(2.53)	
Age from 43 to 47[a]	−4.49*	5.21**	*0.00*
	(2.63)	(2.54)	
Age from 48 to 52[a]	−4.62*	6.33**	*0.00*
	(2.66)	(2.56)	
Age from 53 to 57[a]	2.85	11.80***	*0.00*
	(2.71)	(2.61)	
Age from 58 to 62[a]	7.22***	18.37***	*0.00*
	(2.80)	(2.70)	
Age from 63 to 67[a]	12.86***	23.41***	*0.00*
	(2.99)	(2.87)	
Age from 68 or over[a]	15.77***	19.97***	*0.03*
	(3.11)	(2.93)	
Male	−2.43***	−3.86***	*0.02*
	(0.94)	(0.91)	
Non-indigenous status	13.13***	12.92***	*0.71*
	(3.47)	(3.24)	
Born in Australia	9.82***	10.03***	*0.28*
	(1.06)	(1.02)	
Poor English proficiency	−4.16**	−8.28***	*0.01*
	(1.64)	(1.60)	
Diploma/certificate[b]	2.97***	3.80***	*0.18*
	(1.06)	(1.01)	
Bachelor or higher[b]	14.11***	12.45***	*0.45*
	(1.17)	(1.13)	
Widowed[c]	−2.78	−0.15	*0.26*
	(2.79)	(2.62)	
Divorced[c]	−4.82***	−2.29	*0.05*
	(1.68)	(1.60)	
Separated[c]	−5.22**	−3.32	*0.34*
	(2.24)	(2.12)	
Married[c]	−0.70	−0.94	*0.76*
	(1.31)	(1.26)	
Poor health	−4.65***	−4.40***	*0.94*
	(1.49)	(1.42)	

TABLE 3 (Continued)

Variable	Survey data (1)	Administrative data (2)	Test for equality of coefficient in two equations (*p* value) (3)
Mental distress	−3.40**	−1.95	*0.28*
	(1.54)	(1.47)	
Disable	1.57	2.86***	*0.06*
	(1.02)	(0.98)	
Inpatient treatment	6.75***	6.26***	*0.99*
	(1.46)	(1.40)	
Outpatient treatment	−4.50***	−5.25***	*0.37*
	(1.69)	(1.61)	
Smoker	−13.36***	−11.98***	*0.62*
	(1.21)	(1.15)	
Part-time employed[d]	−2.36**	−1.18	*0.24*
	(1.14)	(1.11)	
Unemployed[d]	−7.37**	−4.17	*0.28*
	(3.04)	(2.83)	
Not in the labor force[d]	5.50***	0.61	*0.00*
	(1.61)	(1.55)	
Number of adults in household	1.69***	1.29**	*0.54*
	(0.64)	(0.61)	
Number of children in household	−2.55***	−1.55***	*0.04*
	(0.53)	(0.51)	
Household annual income (admin)	25.42***	29.75***	*0.00*
	(1.01)	(1.04)	
Household annual income squared	−0.72***	−0.85***	*0.00*
	(0.06)	(0.06)	
Observations	9919	9919	
Sample mean	61.23	64.23	
Test for equality of two equations (*p* value)			*0.00*

Note: Results (in average marginal effects) are from a Probit regression. Coefficient estimates, standard errors and sample means are multiplied by 100 for esthetic 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 italic in Column 3. Other explanatory variables include urban, state/territory, survey month-year dummies. Robust standard errors are in parentheses.

[a]Age from 18 to 22 years as the base group.

[b]Having year 12 or below qualification as the base group.

[c]Never married as the base group.

[d]Full-time employed as the base group.

The symbol *denotes significance at the 10% level, **at the 5% level, and ***at the 1% level.

opposite signs in the false positives and false negatives regressions), three of them, namely those representing education levels, marital status and smoker status, do not display noticeable differences in the coverage determinant equations. These patterns, which are not observed with other variables used in this study, are consistent with a typically common pattern found in the related literature on measurement errors in survey data. Specifically, that literature indicates that reporting worse typically leads to pronounced bias, while reporting more or less can preserve substantive results (Bound et al., 2001; Meyer et al., 2015, 2022).

The results presented here also suggest that, while the aggregate PHI coverage rate is quite accurate in the survey data (see sample mean figures reported at the bottom of Table 3), using survey data would lead to a distorted picture of which individuals are covered. Specifically, the smaller survey estimates and statistically significant differences between the survey and administrative

TABLE 4 Association between private health insurance (PHI) misreporting and responses to other PHI-related questions.

	(1) Security or protection or peace of mind	(2) Lifetime cover or avoid age surcharge	(3) Choice of doctor	(4) Allow treatment as private patient in hospital	(5) Provides benefits for ancillary services or extras	(6) Shorter wait for treatment or concerned over public hospital	(7) Always had it or parents pay it or condition of job	(8) To gain government benefits or avoid extra Medicare levy	(9) Other financial reasons	(10) Has condition that requires treatment	(11) Elderly or getting older or likely to need treatment	(12) Other reason
Panel A: Reasons for having PHI (sample of 6073 individuals with PHI reported in NHS)												
False positives	−6.44***	−8.39***	−2.08	−5.66**	−8.00***	−7.12***	−1.51	−11.62***	1.60*	−1.45	0.74	6.50***
	(2.23)	(2.54)	(2.54)	(2.59)	(2.70)	(2.62)	(2.26)	(2.75)	(0.96)	(1.60)	(1.78)	(0.74)
Sample mean	71.38	23.94	36.24	51.29	43.39	46.19	29.10	26.61	4.23	9.67	15.53	4.30

	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 private health insurance 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
Panel B: Reasons for not having PHI (sample of 3339 individuals without PHI reported in NHS)											
False negatives	6.50**	1.21***	0.97	−4.62*	−0.12	−1.16	−0.11	−2.06	−1.01	1.59	0.15
	(2.80)	(0.43)	(2.36)	(2.70)	(1.95)	(1.24)	(1.69)	(1.49)	(1.28)	(1.58)	(1.47)
Sample mean	57.74	0.95	20.96	29.44	12.97	4.58	10.36	7.01	4.88	8.83	7.85

	Type of cover		Type of membership				Length of coverage			
	Hospital cover only	Both hospital and ancillary cover	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
Panel C: Characteristics of PHI policy reported in NHS (sample of 6073 individuals with PHI reported in NHS)										
False positives	6.01***	−6.01***	2.04	1.63	0.04	−2.41	1.35**	3.06***	0.65	−6.84***
	(1.34)	(1.34)	(1.98)	(1.69)	(0.76)	(1.73)	(0.65)	(0.80)	(1.43)	(1.54)
Sample mean	10.29	89.71	46.14	20.07	2.50	31.27	2.57	4.01	9.22	84.55

Note: "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. Results (in average marginal effects) are from a Probit regression. Marginal impact estimates, standard errors and sample means are multiplied by 100 for esthetic 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.

estimates of some variables indicate that the coverage is under-represented in survey data for individuals who are older, divorced, disabled, or have fewer children or higher household income. By contrast, and for analogous reasons, using survey data would over-report PHI coverage for individuals who are female, not in the labor force, or have poorer English proficiency.

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.[21] 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 their 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 (ABS, 2017a; Buchmueller et al., 2013; Viney et al., 2006; Zhang & Prakash, 2021).[22] Bollinger and Tasseva (2023) point out that misreporting participation results in sample selection bias for follow-up questions. However, there is no evidence on how PHI misreporting affects the responses to these follow-up survey 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, in addition to a comprehensive set of individual and household level explanatory variables as described in the above sections, we introduce a variable describing PHI misreporting status as an independent variable. To accommodate the fact that other PHI-related questions are asked conditionally on responses to the question on PHI coverage, we necessarily split the sample based on survey PHI coverage status. Specifically, we first focus on a sample of individuals who self-identify as being insured and include a variable describing whether the individual provides a false positive answer in the regressions of reasons for having PHI, or characteristics of PHI policies. Furthermore, we consider a sample of individuals who self-report as being uninsured and include a variable representing a false negative reporting status in the regressions of reasons for not having PHI.

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., false positive 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 cover” (vs. “Hospital cover only”) or being covered by PHI for 5 years or more (see Panel C).

However, we find no statistically significant association between the false negative reporting status and reasons for not purchasing PHI (see Panel B). Two exceptions are that, relative to the truly uninsured persons, individuals who misreport as being uninsured are statistically significantly (at 5% level or higher) more likely to select “Cannot afford it/too expensive” or “high risk category” as one of main reasons for not purchasing PHI. Our finding of a statistically significant association between misreporting of PHI enrollment status and responses to follow-up PHI related questions suggests complicated biases, including non-sensical responses, in other studies that use such responses.

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 enrollment using survey data. Our results indicate the signs of most determinants of PHI enrollment 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 enrollment, especially those capturing age, gender, language proficiency, labor force status, disability status, the number of children or household income. Finally, we show that misreporting of PHI enrollment 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 enrollment in survey data are non-classical. These non-classical errors suggest complicated biases in other studies that use self-reported PHI enrollment as an independent variable in regressions, including those evaluating effects of PHI enrollment 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 enrollment like ours, is worthwhile. This would provide a more robust evidence base for health-related policies.

ACKNOWLEDGMENTS

We would like to thank Associate Editor David Johnston, anonymous reviewers of this journal and Nathan Kettlewell for useful comments and suggestions. This research was partly funded by the Australian Research Council Centre of Excellence for Children and Families over the Life Course (#CE200100025).

Open access publishing facilitated by The University of Western Australia, as part of the Wiley - The University of Western Australia agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

This paper uses unit record data from linked Australian National Health Survey and administrative Personal Income Tax, provided from the Australian Bureau of Statistics' Multi-Agency Data Integration Project (MADIP). These data are proprietary and researchers wishing to use them must seek approval from the relevant institutions. The authors are willing to offer guidance about the process of seeking approval.

ETHICS STATEMENT

The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. This research is partly funded by the Australian Research Council Centre of Excellence for Children and Families over the Life Course (CE200100025). However, the funder does not involve in study design; in the collection, analysis and interpretation of data; in the writing of the articles; and in the decision to submit it for publication.

DISCLAIMER

The results of these studies are based, in part, on tax data supplied by the ATO to the ABS under the *Taxation Administration Act 1953*, which requires that such data is only used for the purpose of administering the *Census and Statistics Act 1905*. Any discussion of data limitations or weaknesses is in the context of using the data for statistical purposes, and is not related to the ability of the data to support the ATO's core operational requirements. Legislative requirements to ensure privacy and secrecy of these data have been followed. For access to MADIP data under Section 16A of the *ABS Act 1975* or enabled by Section 15 of the *Census and Statistics (Information Release and Access) Determination 2018*, source data are de-identified and so data about specific individuals has not been viewed in conducting this analysis. In accordance with *the Census and Statistics Act 1905*, results have been treated where necessary to ensure that they are not likely to enable identification of a particular person or organization.

ORCID

Ha Trong Nguyen https://orcid.org/0000-0002-2240-8942
Luke Connelly https://orcid.org/0000-0002-1734-4809

ENDNOTES

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

[2] Particularly, US studies use aggregate data (Lurie & Pearce, 2021) or smaller and less comprehensive datasets (Call et al., 2022; Nelson et al., 2000; Pascale et al., 2019b) than ours.

[3] The US has a program also called Medicare which primarily provides health insurance for individuals aged 65 years and older, as well as certain younger individuals with disability (Call et al., 2022).

[4] There is a small number (about 30) of observations surveyed in July 2015. For them, we consider as if they were surveyed in June 2015 - the last month of the 2014–15 financial year - to have more reliable estimates for month-year dummy variables in all regressions. Dropping these observations makes no substantive difference to our findings. 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 (Buchmueller et al., 2013; Cameron et al., 1988; Cameron & Trivedi, 1991; Doiron et al., 2008; Johar et al., 2011; Kettlewell, 2019b; Palangkaraya & Yong, 2005; Savage & Wright, 2003). Other survey data sources include Household, Income and Labor Dynamics in Australia (Bilgrami et al., 2021; Buchmueller et al., 2021; Cheng, 2014; Kettlewell, 2019a) and the 45-and-Up Study (Doiron et al., 2014; Johar & Savage, 2012; Kettlewell et al., 2018). A few Australian studies have used data from PIT, the same data source as one of our data sources, to document PHI enrollment (Kettlewell & Zhang, 2021; Liu & Zhang, 2022; Stavrunova & Yerokhin, 2014).

[5] An appropriate level of cover must have an excess of $750 or less for singles and $1500 or less for couples or families (ATO, 2022). "Ancillary" cover, also known as "extra", which covers items such as optical, dental and physiotherapy or chiropractic treatment, is not private patient hospital cover.

[6] Type of cover is derived from responses to a question asking "Which best describes what [your/his/her] private health insurance covers?". This question is only asked for respondents answering "Yes" to the PHI coverage question, as documented above. Among a sample of all insured individuals in 2014–15 NHS data, around 11%, 8%, 80%, and 1% of them reported as having a PHI policy under "hospital cover only", "ancillary cover only", "both hospital and ancillary cover", and "insured but type of cover not known", respectively.

[7] It remains unclear how this potential difference in concepts may affect the results. Assigning individuals with an "ancillary cover only" as being insured in the survey data, as Nguyen et al. (2022) did, would underreport the false negative rates and overreport the false positive rates. However, doing so does not change other regression results presented in this current paper in any significant way.

[8] Briefly, Medicare Levy Surcharge (MLS) is a means-tested insurance mandate where individuals who do not purchase private insurance covering hospital care are subject to a tax surcharge on their total income. Lifetime Health Cover requires that individuals who do not have PHI hospital cover have to pay a 2% loading on top of the hospital premium for every year individuals are aged over 30. Finally, premium subsidy policies provide rebates for private hospital cover (Duckett & Nemet, 2019).

[9] For instance, PIT data have no information about the coverage duration during this financial year. Moreover, survey information on coverage duration is not detailed or accurate enough (see Section 4.2 for details).

[10] Almost all individuals in our sample (over 99%) were born in Australia or arrived in Australia before 2013. This means that they are expected to have lived in Australia for the entire study period, which reduces the likelihood that any of them were covered by an overseas provider.

[11] See Appendix Table A1 for variable description and summary statistics of main variables.

[12] We are interested in the raw reporting numbers mainly because, as described above, our focused sample is not representative of the whole population. Moreover, 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.

[13] We use household income which is the means-test base for coupled individuals according to most Australian PHI policies (Duckett & Nemet, 2019). For single individuals, household income refers to their own income. Particularly, household income is calculated from the respondent's and their spouse's taxable incomes which are obtained from PIT data. For tax assessment purposes, as part of the tax lodgement process, all tax filers who have a spouse during the financial year are legally required to complete a spouse details section which asks about the spouse's taxable income, among other details (ATO, 2015). The spouse's taxable income is set to zero if the spouse does not have any taxable income, for example, because the spouse does not work, or the tax filler does not have a spouse during the financial year.

[14] 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 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. In practice, probably due to a lack of a proper identification strategy or suitable data, most studies, including this current study, have to speculate about the causes of misreporting (Celhay et al., 2022).

[15] The 2014–15 NHS shows that about two thirds of insured individuals in our sample are covered under a household-based (i.e., family or couple) membership. As expected, the proportion of insured individuals with household-based coverage is higher among married individuals (92%) than non-married individuals (31%).

[16] 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.

[17] Remaining results, reported in Appendix Table A4, show noticeable geographical differences in both types of misreporting of PHI coverage. Moreover, the negative and statistically significant estimate of the linkage quality variable in the false positive equation suggests that a higher matching quality between NHS and MADIP asset is associated with a lower probability of observing a false positive report. However, the estimate of the linkage quality variable is statistically insignificant in the false negative equation. Similarly, there appear no temporal differences in misreporting as all estimates of month-year dummy variables are statistically insignificant. This alleviates the concern that differences in survey timing may drive our results.

[18] This age profile of misreporting is consistent with that in a modified empirical model in which we introduce age in 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 age up to the age of 47 years, before increasing. Likewise, and in line with the baseline results, the probability of providing a false report decreases with age up to 57 years of age, 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.

[19] See Nguimkeu et al. (2019) for a formal proof and an empirical example. While our estimates can be used as inputs for the theoretical framework proposed by Nguimkeu et al. (2019) to correct for the bias in estimated PHI treatment effects, this framework is not readily applied to our case for two main reasons. First, the theoretical framework proposed by Nguimkeu et al. (2019) only focuses on the case of one-side endogenous misreporting (e.g., the case of false negatives which is the predominant case of misreporting of public program receipt as documented in Meyer et al. (2009)). By contrast, our results show that misreporting of PHI coverage is bidirectional (i.e., false negatives and false positives). Second, employing their model requires one to find two valid instruments: One for the endogenous PHI enrollment and the other to address PHI misreporting. Unfortunately, it is difficult to find two plausible instruments in our case.

[20] This research question is particularly important given that, in the absence of randomized controlled trials, such as the RAND experiment in the US (Manning et al., 1987), observational studies mainly rely on an IV method to address the potential endogeneity of self-reported PHI enrollment when examining it causal impacts (Hullegie & Klein, 2010; Nguyen & Connelly, 2017). Australian studies are not an exception because all existing IV studies employ self-reported PHI coverage measures (Cheng, 2014; Doiron & Kettlewell, 2018; Eldridge et al., 2017; Hopkins et al., 2013; Kettlewell, 2019b; Srivastava et al., 2017).

[21] 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?". As documented previously, 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?".

[22] Specifically, these studies employ responses to these follow-up questions to document reasons for having PHI or not (ABS, 2017a; Zhang & Prakash, 2021), investigate factors associating with stated reasons for having PHI (Buchmueller et al., 2013), or explore whether reasons for purchasing PHI influence behaviors (Viney et al., 2006).

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*(5), 1451–1467. https://doi.org/10.1162/rest_a_00352

ABS. (2017a). *Health service usage and health related actions, Australia, 2014-15*. Australian Bureau of Statistics.

ABS. (2017b). *National health survey 2014-15 questionnaire*. Australian Bureau of Statistics (ABS).

ABS. (2017c). *National health survey: Users' guide, 2014-15*. Australian Bureau of Statistics (ABS).

ABS. (2020a). *Integration of the 2017-18 national health survey and the personal linkage spine*. Australian Bureau of Statistics (ABS).

ABS. (2020b). *Integration of the national health survey with the multi-agency data integration project*. Australian Bureau of Statistics (ABS).

ABS. (2020c). *National health survey: Persons accessing pharmaceutical benefits scheme subsidised prescriptions, 2014-15*. Australian Bureau of Statistics (ABS).

ABS. (2023). *Multi-agency data integration Project (MADIP)*. Australian Bureau of Statistics (ABS). Retrieved from: https://www.abs.gov.au/about/data-services/data-integration/integrated-data/multi-agency-data-integration-project-madip

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

ATO. (2015). *Tax return for individuals 2014-15*. Australian Taxation Office (ATO). Retrieved from: https://www.ato.gov.au/Forms/Tax-return-for-individuals-2014-15/

ATO. (2022). *Taxation statistics*. Australian Taxation Office (ATO). Retrieved from: https://www.ato.gov.au/About-ATO/Research-and-statistics/In-detail/Taxation-statistics/

Baker, M., Stabile, M., & Deri, C. (2004). What do self-reported, objective, measures of health measure? *Journal of Human Resources*, *39*(4), 1067–1093. https://doi.org/10.2307/3559039

Battistin, E., De Nadai, M., & Sianesi, B. (2014). Misreported schooling, multiple measures and returns to educational qualifications. *Journal of Econometrics*, *181*(2), 136–150. https://doi.org/10.1016/j.jeconom.2014.03.002

Besley, T., Hall, J., & Preston, I. (1999). The demand for private health insurance: Do waiting lists matter? *Journal of Public Economics*, *72*(2), 155–181. https://doi.org/10.1016/s0047-2727(98)00108-x

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. *The Economic Record*, *97*(317), 170–211. https://doi.org/10.1111/1475-4932.12603

Bingley, P., & Martinello, A. (2017). Measurement error in income and schooling and the bias of linear estimators. *Journal of Labor Economics*, *35*(4), 1117–1148. https://doi.org/10.1086/692539

Bollinger, C. R., & David, M. H. (2001). Estimation with response error and nonresponse. *Journal of Business and Economic Statistics*, *19*(2), 129–141. https://doi.org/10.1198/073500101316970368

Bollinger, C. R., & Tasseva, I. V. (2023). Income source confusion using the SILC. *Public Opinion Quarterly forthcoming*.

Bonsang, E., & Costa-Font, J. (2022). Buying control? 'Locus of control' and the uptake of supplementary health insurance. *Journal of Economic Behavior and Organization*, *204*, 476–489. https://doi.org/10.1016/j.jebo.2022.10.035

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*(6), 1973–1995. https://doi.org/10.1111/1475-6773.12308

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

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. https://doi.org/10.1016/j.jhealeco.2020.102403

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*(5), 757–767. https://doi.org/10.1016/j.jhealeco.2013.05.001

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*(2), 519–529. https://doi.org/10.1016/j.jhealeco.2007.05.005

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*(2pt1), 652–664. https://doi.org/10.1111/j.1475-6773.2012.01446.x

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*(4), 438–456. https://doi.org/10.5034/inquiryjrnl_45.04.438

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*(4), 930–943. https://doi.org/10.1111/1475-6773.13874

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. https://doi.org/10.1016/0047-2727(91)90045-4

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*(1), 85–106. https://doi.org/10.2307/2297531

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*(4), 1739–1757. https://doi.org/10.1111/j.1475-6773.2006.00662.x

Cawley, J., Maclean, J. C., Hammer, M., & Wintfeld, N. (2015). Reporting error in weight and its implications for bias in economic models. *Economics and Human Biology*, *19*, 27–44. https://doi.org/10.1016/j.ehb.2015.07.001

Celhay, P. A., Meyer, B. D., & Mittag, N. (2022). What leads to measurement errors? Evidence from reports of program participation in three surveys. *NBER Working Paper No 29652*. https://doi.org/10.2139/ssrn.4010501

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. https://doi.org/10.1016/j.jhealeco.2013.11.007

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*.

Connelly, L. B., Paolucci, F., Butler, J. R. G., & Collins, P. (2010). Risk equalisation and voluntary health insurance markets: The case of Australia. *Health Policy*, *98*(1), 3–14. https://doi.org/10.1016/j.healthpol.2010.06.002

DiTraglia, F. J., & García-Jimeno, C. (2019). Identifying the effect of a mis-classified, binary, endogenous regressor. *Journal of Econometrics*, *209*(2), 376–390. https://doi.org/10.1016/j.jeconom.2019.01.007

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. https://doi.org/10.1016/j.jhealeco.2014.06.006

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*(3), 317–334. https://doi.org/10.1002/hec.1267

Doiron, D., & Kettlewell, N. (2018). The effect of health insurance on the substitution between public and private hospital care. *The Economic Record*, *94*(305), 135–154. https://doi.org/10.1111/1475-4932.12394

Doiron, D., & Kettlewell, N. (2020). Family formation and the demand for health insurance. *Health Economics*, *29*(4), 523–533. https://doi.org/10.1002/hec.4000

Drake, C., Ryan, C., & Dowd, B. (2022). Sources of inertia in the individual health insurance market. *Journal of Public Economics*, *208*, 104622. https://doi.org/10.1016/j.jpubeco.2022.104622

Duckett, S., & Nemet, K. (2019). *The history and purposes of private health insurance*. Grattan Institute.

Eldridge, D. S., Onur, I., & Velamuri, M. (2017). The impact of private hospital insurance on the utilization of hospital care in Australia. *Applied Economics*, *49*(1), 78–95. https://doi.org/10.1080/00036846.2016.1192273

Ellis, R., & Savage, E. (2008). Run for cover now or later? The impact of premiums, threats and deadlines on private health insurance in Australia. *International Journal of Health Care Finance and Economics*, *8*(4), 257–277. https://doi.org/10.1007/s10754-008-9040-4

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*(8), e0219731. https://doi.org/10.1371/journal.pone.0219731

Feng, S., & Hu, Y. (2013). Misclassification errors and the underestimation of the US unemployment rate. *The American Economic Review*, *103*(2), 1054–1070. https://doi.org/10.1257/aer.103.2.1054

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. https://doi.org/10.1016/j.jhealeco.2017.02.004

Hausman, J. A., Abrevaya, J., & Scott-Morton, F. M. (1998). Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics*, *87*(2), 239–269. https://doi.org/10.1016/s0304-4076(98)00015-3

Hopkins, S., Kidd, M. P., & Ulker, A. (2013). Private health insurance status and utilisation of dental services in Australia. *The Economic Record*, *89*(285), 194–206. https://doi.org/10.1111/1475-4932.12040

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*(9), 1048–1062. https://doi.org/10.1002/hec.1642

Hurst, E., Li, G., & Pugsley, B. (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *The Review of Economics and Statistics*, *96*(1), 19–33. https://doi.org/10.1162/rest_a_00363

Jenkins, S. P., & Rios-Avila, F. (2023). Reconciling reports: Modelling employment earnings and measurement errors using linked survey and administrative data. *Journal of the Royal Statistical Society - Series A: Statistics in Society*, *186*(1), 110–136. https://doi.org/10.1093/jrsssa/qnac003

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*(S1), 68–86. https://doi.org/10.1002/hec.1707

Johar, M., & Savage, E. (2012). Sources of advantageous selection: Evidence using actual health expenditure risk. *Economics Letters*, *116*(3), 579–582. https://doi.org/10.1016/j.econlet.2012.06.002

Johnston, D. W., Propper, C., & Shields, M. A. (2009). Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics*, *28*(3), 540–552. https://doi.org/10.1016/j.jhealeco.2009.02.010

Kettlewell, N. (2019a). Risk preference dynamics around life events. *Journal of Economic Behavior and Organization*, *162*, 66–84. https://doi.org/10.1016/j.jebo.2019.04.018

Kettlewell, N. (2019b). Utilization and selection in an ancillaries health insurance market. *Journal of Risk & Insurance*, *86*(4), 989–1017. https://doi.org/10.1111/jori.12250

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*(2), 96–101. https://doi.org/10.1080/13504851.2017.1299094

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*.

Kreider, B. (2010). Regression coefficient identification decay in the presence of infrequent classification errors. *The Review of Economics and Statistics*, *92*(4), 1017–1023. https://doi.org/10.1162/rest_a_00044

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*(2), 164–184. https://doi.org/10.1086/712213

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., Leibowitz, A., & Marquis, M. S. (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *The American Economic Review*, *77*, 251–277.

McGuire, T. G. (2011). Chapter five - demand for health insurance. In V. Mark, T. G. M. Pauly, & P. B. Pedro (Eds.), *Handbook of health economics* (pp. 317–396). Elsevier.

Meyer, B. D., & Mittag, N. (2017). Misclassification in binary choice models. *Journal of Econometrics*, *200*(2), 295–311. https://doi.org/10.1016/j.jeconom.2017.06.012

Meyer, B. D., & Mittag, N. (2019a). Misreporting of government transfers: How important are survey design and geography? *Southern Economic Journal*, *86*(1), 230–253. https://doi.org/10.1002/soej.12366

Meyer, B. D., & Mittag, N. (2019b). Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness and holes in the safety net. *American Economic Journal: Applied Economics*, *11*(2), 176–204. https://doi.org/10.1257/app.20170478

Meyer, B. D., & Mittag, N. (2021). Combining administrative and survey data to improve income measurement. In A. Y. Chun, M. D. Larsen, G. Durrant, & J. P. Reiter (Eds.), *Administrative records for survey methodology* (pp. 297–322).

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*(5), 1605–1644. https://doi.org/10.3368/jhr.58.1.0818-9704r2

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. *The Journal of Economic Perspectives*, *29*(4), 199–226. https://doi.org/10.1257/jep.29.4.199

Mittag, N. (2019). Correcting for misreporting of government benefits. *American Economic Journal: Economic Policy*, *12*(2), 142–164. https://doi.org/10.1257/pol.20160618

Nelson, D. E., Thompson, B. L., Davenport, N. J., & Penaloza, L. J. (2000). What people really know about their health insurance: A comparison of information obtained from individuals and their insurers. *American Journal of Public Health*, *90*, 924–928.

Nguimkeu, P., Denteh, A., & Tchernis, R. (2019). On the estimation of treatment effects with endogenous misreporting. *Journal of Econometrics*, *208*(2), 487–506. https://doi.org/10.1016/j.jeconom.2018.10.005

Nguyen, H. T., & Connelly, L. B. (2017). Cost-sharing in health insurance and its impact in a developing country–Evidence from a quasi-natural experiment. *Bankwest Curtin Economics Centre (BCEC) working paper number 17-02*.

Nguyen, H. T., Le, H. T., Connelly, L., & Mitrou, F. (2022). *Accuracy of self-reported private health insurance coverage*. GLO Discussion Paper No 1215, Global Labor Organization (GLO), Essen.

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*(2), 270–288. https://doi.org/10.1002/hec.4200

Nguyen, T.-H., & Leung, S. (2013). Dynamics of health insurance enrollment in Vietnam, 2004–2006. *Journal of the Asia Pacific Economy*, *18*(4), 594–614. https://doi.org/10.1080/13547860.2013.803842

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*(1), 34–43. https://doi.org/10.1111/1475-6773.13058

Palangkaraya, A., & Yong, J. (2005). Effects of recent carrot-and-stick policy initiatives on private health insurance coverage in Australia. *The Economic Record*, *81*(254), 262–272. https://doi.org/10.1111/j.1475-4932.2005.00260.x

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*(2), 409–460. https://doi.org/10.2478/jos-2019-0019

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*(5), 1099–1109. https://doi.org/10.1111/1475-6773.13191

Pascale, J., Roemer, M. I., & Resnick, D. M. (2009). Medicaid underreporting in the CPS: Results from a record check study. *Public Opinion Quarterly*, *73*(3), 497–520. https://doi.org/10.1093/poq/nfp028

Propper, C., Rees, H., & Green, K. (2001). The demand for private medical insurance in the UK: A cohort analysis. *The Economic Journal*, *111*(471), C180–C200. https://doi.org/10.1111/1468-0297.00627

Savage, E., & Wright, D. J. (2003). Moral hazard and adverse selection in Australian private hospitals: 1989-1990. *Journal of Health Economics*, *22*(3), 331–359. https://doi.org/10.1016/s0167-6296(02)00104-2

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources*, *50*(2), 301–316. https://doi.org/10.3368/jhr.50.2.301

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. *Bulletin of the World Health Organization*, *90*(9), 685–692. https://doi.org/10.2471/blt.12.102301

Srivastava, P., Chen, G., & Harris, A. (2017). Oral health, dental insurance and dental service use in Australia. *Health Economics*, *26*(1), 35–53. https://doi.org/10.1002/hec.3272

Stavrunova, O., & Yerokhin, O. (2014). Tax incentives and the demand for private health insurance. *Journal of Health Economics*, *34*, 121–130. https://doi.org/10.1016/j.jhealeco.2014.01.001

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. Jossey-Bass.

Viney, R., Savage, E., & Fiebig, D. (2006). *Does the reason for buying health insurance influence behaviour?* (Working paper 2006/1). Centre for Health Economics Research and Evaluation.

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

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nguyen, H. T., Le, H. T., Connelly, L., & Mitrou, F. (2023). Accuracy of self-reported private health insurance coverage. *Health Economics*, *32*(12), 2709–2729. https://doi.org/10.1002/hec.4748