



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Watson, N;Ey, T

Title:

Comparison of HILDA Survey Estimates With the 2006–2021 Censuses

Date:

2026-03-01

Citation:

Watson, N. & Ey, T. (2026). Comparison of HILDA Survey Estimates With the 2006–2021 Censuses. *Australian Economic Review*, 59 (1), pp.75-83. <https://doi.org/10.1111/1467-8462.70028>.

Persistent Link:

<https://hdl.handle.net/11343/362460>

License:

[CC-BY](#)

ARTICLE OPEN ACCESS

Comparison of HILDA Survey Estimates With the 2006–2021 Censuses

Nicole Watson  | Taylor Ey 

Melbourne Institute of Applied Economic and Social Research, University of Melbourne, Parkville, Australia

Correspondence: Nicole Watson (n.watson@unimelb.edu.au)**Received:** 22 May 2025 | **Revised:** 18 August 2025 | **Accepted:** 20 August 2025**Keywords:** attrition | household panel | nonresponse | representativeness

ABSTRACT

Issue: The Household, Income and Labour Dynamics in Australia (HILDA) Survey has provided researchers and policymakers with invaluable insights into Australian life for more than two decades. Nevertheless, nonresponse, attrition and shifts in population coverage can affect how representative the sample is over time.

Methods: We evaluate the HILDA Survey's representativeness by comparing its estimates against those from the Census across four time periods and four age groups.

Findings: We find strong alignment across many of the variables examined. For the most part, observed differences can be explained by differences in the questions asked, recall periods or different collection methodologies. Recent immigrants are a concern as they can only join the sample by living with a HILDA household or be included via a top-up sample. Sizeable differences are identified in the proportion of immigrants aged 15–24 and 25–44 in particular, which were corrected in 2011 with the addition of a general top-up sample.

Implications: We conclude that the HILDA Survey is largely representative of the Australian population and note that the coverage of recent immigrants is in the process of being addressed. Additional immigrant-specific samples are being added in 2024 and 2025 and regular immigrant sample top-ups are needed thereafter.

1 | Introduction

Now in its 25th year, the Household, Income and Labour Dynamics in Australia (HILDA) Survey has been providing researchers and policymakers with invaluable insights into the many different opportunities, challenges and pathways of Australian life. The HILDA Survey is a nationwide household panel study that began in 2001, providing longitudinal data on many aspects of life including employment, income, wealth, relationships, family formation, health and wellbeing. There have been over 2000 journal articles published using this data, with now more than 140 published each year, along with numerous other publications.¹ The HILDA Survey is used across many disciplines, including economics, sociology, public

health, demography, psychology and environmental studies. Indeed, the former editor of the *Australian Economic Review*, Ross Williams, stated in *The Policy Providers* that the HILDA Survey was “one of the most important and influential projects in economics, and more generally in the social sciences, in Australia in the last fifty years” (Williams 2012, 144). Not only providing new data, it has also “stimulated a step change in the sophistication of empirical analysis in the social sciences in Australia” (Richardson 2013, p.216).

Given the HILDA Survey's broad applicability, it is important to systematically review its representativeness. Nonresponse and attrition can reduce the overall sample size but only introduces bias if the respondents and nonrespondents are significantly

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *The Australian Economic Review* published by John Wiley & Sons Australia, Ltd on behalf of The University of Melbourne, Melbourne Institute: Applied Economic & Social Research, Faculty of Business and Economics.

different from one another.² Previous research on nonresponse and attrition typically finds associations with demographic, economic and health variables. Non-respondents are more likely to be male, either young or elderly, single, less educated, renters, living in capital cities, and from ethnic minority groups (Cabrera-Álvarez et al. 2023; Lynn et al. 2023; Uhrig 2008; Watson and Wooden 2009). In terms of economic factors, these individuals tend to have low income (Fitzgerald et al. 1998; Hill and Willis 2001; Lynn et al. 2023; Watson 2003) and are less likely to be employed (Gray et al. 1996; Lepkowski and Couper 2002). Non-respondents are also more likely to have a serious long-term health condition (Cabrera-Álvarez et al. 2023; Watson and Wooden 2009). Similar characteristics are associated with nonresponse in cross-sectional surveys (Bethlehem et al. 2011; Groves and Couper 2012).

Moreover, over the last two decades, there have been substantial changes in the Australian population in at least four areas. First, due to lower fertility rates and increasing life expectancy, the population is ageing with the median age increasing from 35 years in June 2000 to 38 years in June 2020 (Australian Bureau of Statistics 2020b). Second, net overseas migration has outpaced natural increase during this time, with the exception of 2020–21 where international border restrictions were in place during the COVID-19 pandemic (Australia Institute of Health and Welfare 2024). Third, there has been greater participation in higher education, with a 20% increase in the last 5 years alone, and lower marriage rates, dropping 15% in the 5 years to 2019 (Australian Bureau of Statistics 2020a, 2022b). Fourth, the composition of industries that people are employed in has steadily changed over time, with growth rates slowing for Accommodation and Food Services, Construction and Retail Trade industries and an expansion of Health Care and Social Assistance (Australian Bureau of Statistics 2019). The HILDA Survey needs to keep pace with these changes so that it can continue to provide both longitudinal and cross-sectional data of high quality.

A key feature of the longitudinal design of the HILDA Survey is that the sample automatically replenishes itself in the same way as the population does. New entrants, such as babies, partners and others, join the households of sample members and the inclusion probabilities of these households are adjusted downwards for the new entrants (apart from babies and recent immigrants) to allow for the multiple pathways to observing the household (Watson 2022a). However, there is one obvious area where the sample will struggle to adequately represent the changing cross-sectional population – immigrants. Without a refreshment sample, the HILDA sample can only incorporate recent immigrants into the sample if they live with existing sample members, which is a biased sample of all immigrants. A general refreshment sample was added in 2011, resulting in a 30% increase in the overall sample size (Watson and Wooden 2013). This allowed the immigrants arriving in 2002–2011 to be represented in the sample (albeit with higher weights than average). Immigrants arriving after 2011 are presently underrepresented in the available data sets. However, this is being rectified by a new sample of immigrants being added in 2024 and 2025.³ In addition to correcting for the number of recent immigrants, a refreshment sample including immigrants will have some flow-on effects to other survey estimates.

Immigrants tend to be younger, and even when taking age into account, they also tend to have higher education levels, fewer children, less wealth and savings, be married, settle in capital cities (especially inner areas), and work fewer hours per week (Productivity Commission 2016).

Given these known issues and trends, the question then is how well does the HILDA Survey estimates align with official statistics? Comparisons of other household panel surveys to official statistics have yielded useful insights. For example, Duncan and Hill (1989) compared the 1980 (wave 13) of the Panel Study of Income Dynamics (PSID) with the Current Population Survey (CPS) and found the sample maintained its representation of the nonimmigrant population of the United States. If anything, some of the small differences noted in the initial sample (such as family income) declined over the 13 waves. Fitzgerald et al. (1998) assessed the representativeness of young adults (aged 20–38) in the 1989 PSID against the CPS and found a few percentage points of difference in education, marital status and region. More importantly, Hispanics were underrepresented in the PSID by half and mean wages and salaries were higher in the PSID than the CPS. More recently, Duffy and Sastry (2012) assessed the representativeness of children (aged 0–17) in the 2007 PSID against the American Community Survey (the replacement for the long-form in the decennial Census). This is after the PSID added their first immigrant top-up sample in 1997. They found generally good representation, apart from Asian and Hispanic children and those born to foreign-born parents. Using the UK's Understanding Society, Borkowska (2019) compared estimates from wave 8 to the 2017 Annual Population Survey and found a small under-representation of young people, those living in Greater London, and some ethnic minorities. Some hard-to-reach groups (such as, ethnic minority groups and non-UK born individuals) were more likely to drop out of the sample by wave 8, and this was not sufficiently corrected by weights. These comparisons show how the representativeness of the sample can drift over time from the contemporary population due to nonresponse, attrition and shifts in population coverage.

In this article, we compare the HILDA Survey estimates with the 5-yearly Census for 2006 to 2021. While the Census has quite a different data collection methodology to the HILDA Survey, it can provide some useful insights into how the HILDA sample may be drifting over time. We also draw on a comparison between the HILDA Survey and the General Social Survey (GSS), which are methodologically closer, to deepen our understanding of the comparison with the Census. Some of the differences highlighted through the HILDA-Census comparison may be explained by differences in the data collection methodology if the HILDA-GSS comparison of the variables in question shows similarities.

2 | Data

The methodological approaches of the HILDA Survey and the Census differ in several key aspects, but the HILDA Survey aligns more closely with the GSS methodology. A comparison of these key methodological differences can be found in Table 1. These three data sources are described in turn.

TABLE 1 | Key methodological features of the HILDA Survey and the Census.

Methodological feature	HILDA Survey	Census	General social survey
Type	Longitudinal survey	Census	Cross-sectional survey
Years compared	2006, 2011, 2014, 2016, 2021	2006, 2011, 2016, 2021	2014
Mode of data collection	Interviewer administered, at least 90% face-to-face, up to 10% telephone In 2021, 76% completed by telephone due to COVID-19 lockdowns and restrictions (24% face-to-face)	Household-completed (hardcopy or online)	Interviewer administered face-to-face
Who is interviewed	Each person aged 15+ in the household	One or more persons answer the questions for each person in the household, so have some proxy reports	Randomly selected person aged 15+ from the household
Population exclusions	People in very remote parts of Australia and those in non-private dwellings	None ^a	Very remote areas of Australia; discrete Indigenous communities; non-private dwellings
Fieldwork period	Late July to February/March	2nd Tuesday in August	March to June
Response rate	Initial wave response: 66% in 2001, 69% for the refreshment sample in 2011. Reinterview rate of previous wave respondents: 95% in 2006, 97% in 2011 and 2016, 94% in 2021	96% in 2011, 95% in 2016% and 96% in 2021 (not reported for 2006). The net undercount was 2.7% in 2006, 1.7% in 2011, 1.0% in 2016% and 0.7% in 2021	80%
Primary sources of error	Initial wave nonresponse; attrition	Nonresponse; undercount; proxy reporting error	Nonresponse

^aThe place of Usual Residence database is used, which puts people visiting other locations in Australia back into their usual Mesh-block and excludes overseas visitors. The HILDA Survey also uses the concept of usual residence (so visitors are excluded).

The HILDA Survey is a nationwide household panel study that began in 2001 (DSS & Melbourne Institute 2023; Summerfield et al. 2023). It is funded by the Australian Government Department of Social Services and is managed by the Melbourne Institute of Applied Economic and Social Research at the University of Melbourne. The fieldwork is conducted by Roy Morgan Research (from wave 9) and previously by Neilsen (waves 1 to 8). Annual interviews are conducted with all household members aged 15 years and older and are predominantly face-to-face. Most interviews (95%) are completed between August and November and include core questions on families, employment and income. The HILDA Survey sample is representative of people living in private dwellings (i.e., it excludes people living in institutions) and excludes those living in very remote areas. The household response rate in wave 1 was 66%, resulting in 7682 responding households that contained 19,914 individuals. These individuals (termed continuing sample members [CSMs]) form the basis of the sample that is followed and interviewed over time. Other people who join the household of a CSM are temporarily added to the sample. Babies of CSMs, the other parent of these babies and recent immigrants are converted to CSMs. A population-wide top-up sample was added in 2011 which importantly included, among others, immigrants arriving in Australia after 2001 (Watson and Wooden 2013). The household response rate for the initial wave

of the top-up sample was 69%, resulting in 2153 responding households and 5462 individuals being added to the sample. Reinterview rates (i.e. the percentage of respondents interviewed in one wave that are re-interviewed in the next, excluding those who have died or moved abroad) are high in both samples, rising from 87% in wave 2% to 96% or higher from wave 9 to 20 and was 94% in wave 21. For the top-up sample, the reinterview rates rise from 92% in wave 12 (2012) to 95% or higher for waves 15–19 and was 94% in waves 20 and 21 (2020 and 2021). In 2021, there were 9358 responding households containing 16,549 responding individuals (aged 15 and over).

The Australian Census of Population and Housing (referred to as the “Census”) is conducted by the Australian Bureau of Statistics (ABS) and occurs every 5 years on the second Tuesday in August. For the purposes of this comparison, the 2006, 2011, 2016 and 2021 Censuses are used. From 2016, online completion of the Census has been strongly encouraged, with 59% of households completing the form online in 2016 and 79% in 2021 (Harding et al. 2022). This is compared to 11% in 2006 and 34% in 2011. All other households completed the paper form. One or more individuals fill out the form on behalf of all members of the household. The response rates are very high (95% or higher) and missing observations are imputed (Australian Bureau of Statistics 2022c). The net undercount has fallen from 2.7% in

2006 to 0.7% in 2021 and the ABS have adjusted the Census numbers for this (Australian Bureau of Statistics 2022a). For the purposes of this comparison, the Place of Usual Residence database is used where residents who are visiting other locations in Australia on Census Night are put back into their usual Mesh Block (but not to a specific dwelling or family) and overseas visitors are excluded.⁴

The GSS is a repeated cross-sectional survey conducted every 4 years on a range of topics similar to the HILDA Survey. The estimates used here are drawn from the 2014 survey: even though the GSS has been conducted multiple times, only the 2014 data are available from TableBuilder (the system used to extract the estimates). The fieldwork for the GSS is conducted from March to June. The GSS sample represents people living in private dwellings, excluding those in discrete indigenous communities or very remote areas. Smaller states and lower socio-economic areas are oversampled. One person aged 15 and older was randomly selected from each household, resulting in 12,932 interviews and a response rate of 80% (Australian Bureau of Statistics 2015).

3 | Methods

To assess the representativeness of the HILDA Survey sample, a range of estimates across four cohorts are compared to Census estimates for 2006, 2011, 2016 and 2021.⁵ The four cohorts are people aged 15–24, 25–44, 45–64 and 65 and older which broadly correspond to different life stages. The estimates include:

- Demographic variables – sex, state, remoteness area, indigenous status, social marital status, year of arrival, highest level of school, highest post-school qualification.
- Employment variables – employment status, industry and occupation.

Comparisons are also made between the HILDA Survey and the GSS for 2014 to further our understanding of the results from the HILDA-Census comparisons.

All ABS estimates are extracted via TableBuilder (available at abs.gov.au), which is an online interactive tool to customise aggregate tables. For the GSS, TableBuilder generates both estimates and standard errors that account for the sample design and nonresponse. The estimates from the HILDA Survey use the relevant cross-sectional weight and the standard errors allow for the sample design, nonresponse and attrition. The percentage of missing values observed in the HILDA Survey for these variables are generally very low, less than 0.2%. There are a few exceptions with up to 2.1% of employed people missing industry and up to 1.5% of those aged 65 and older missing post-school education. Both of these are due to uncodeable responses rather than refusals. The percentage of missing values in the Census data items are often substantially higher, typically between 5% and 10% (see Table A1 in the Appendix). This is because the ABS creates records for people who did not respond to the Census and some people may return a form and not complete some questions. The ABS imputes key demographic

variables (such as age, sex, place of usual residence and registered marital status) and the remaining variables are left as ‘not stated’ (Australian Bureau of Statistics 2022c). Also, the Census uses a self-complete form where items may be missed or skipped, whereas the HILDA Survey collects data via an individual interview. The item nonresponse rates in the Census generally reduce over time as the use of the online form increases as this helps respondents correctly navigate through the form. The oldest age group has the highest rate of item nonresponse.

When comparing two estimates, x and y , the standard error of the estimate of the difference, $SE(x - y)$ is calculated as:

$$SE(x - y) = \sqrt{SE(x)^2 + SE(y)^2}$$

where x is the HILDA Survey estimate and y is from the Census, the standard error of the difference is simply the standard error of the HILDA Survey estimate.⁶ This standard error is used to calculate the 95% confidence interval of the difference between the HILDA estimate and the Census estimate.

4 | Results

The difference between the HILDA Survey and Census estimates for 2006–2021 for all variables, apart from industry and occupation, are given in Figure 1. The results for industry and occupation are given in Figure 2. A positive difference indicates that the HILDA Survey estimate is higher than the Census estimate, and a negative difference indicates it is lower. For reference, the HILDA Survey estimates are given in Table A1 in the Appendix.

For the most part, the estimates for sex, state, remoteness area and indigenous status align well with the Census across these four cohorts and four time points. The first two of these variables form part of the benchmarks used in the construction of the HILDA Survey weights, so we would expect them to match closely. The other variables will each be discussed in turn.

4.1 | Social Marital Status

The HILDA Survey has higher estimates of people in de facto relationships in the youngest two cohorts and of married people in the oldest cohort compared to the Census. The likely reason for these differences is way the relationship questions are structured. Both ask for the individual’s registered marital status, but the way relationships in the household are identified is different. The HILDA Survey asks those not married whether they are living with someone in a relationship. The Census asks the household to nominate Person 1 (giving some instructions on how to do this) and then asks about the relationships of other members of the household to Person 1 and their partner (who should be listed as Person 2). This will mean that partners who are away on Census night will be missed along with some of those living in complex households or where Person 1 and 2 were not selected correctly. The relationship questions asked in ABS surveys where an interviewer is involved, such as the GSS,

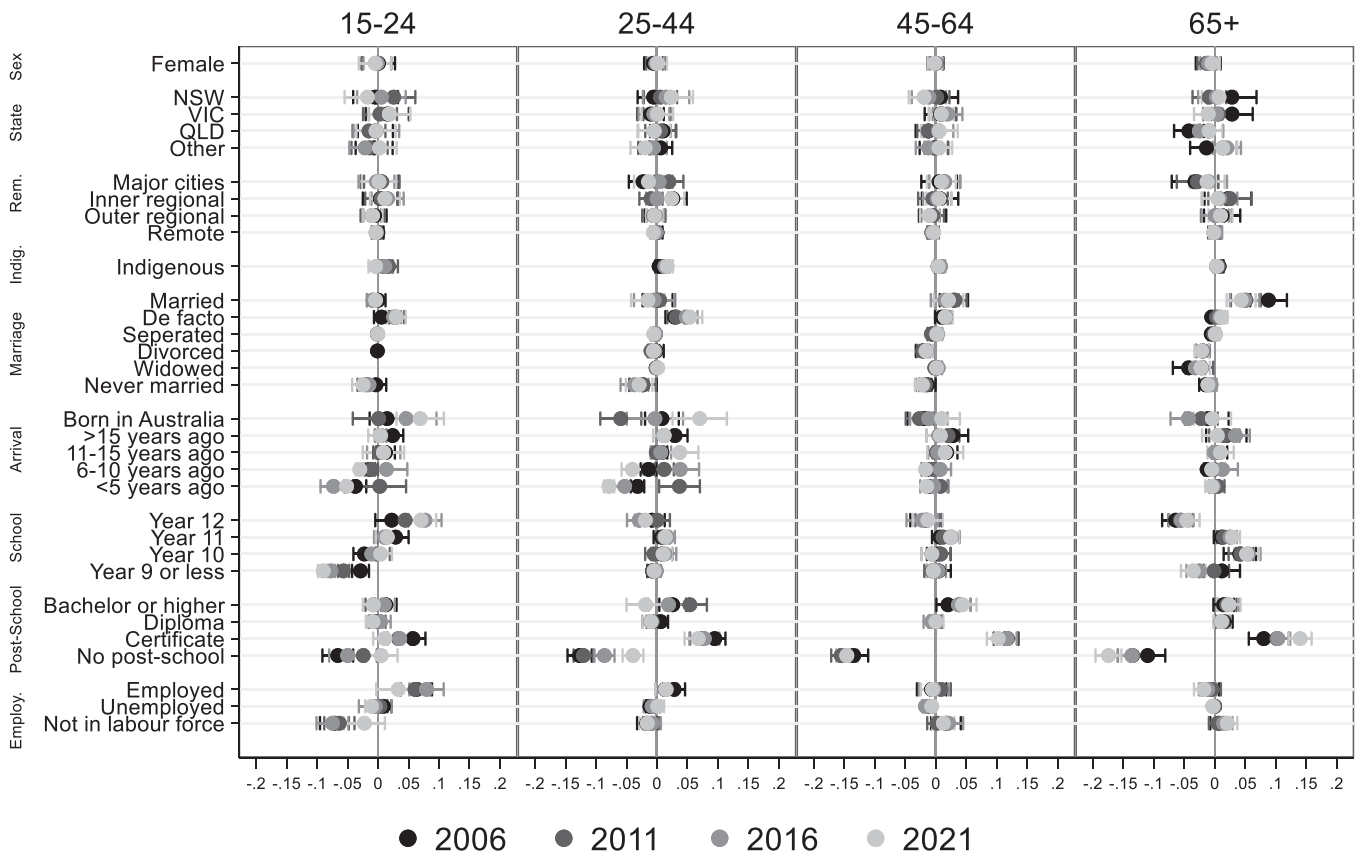


FIGURE 1 | Difference between HILDA Survey demographic estimates and 2006–2021 Census, by age group.

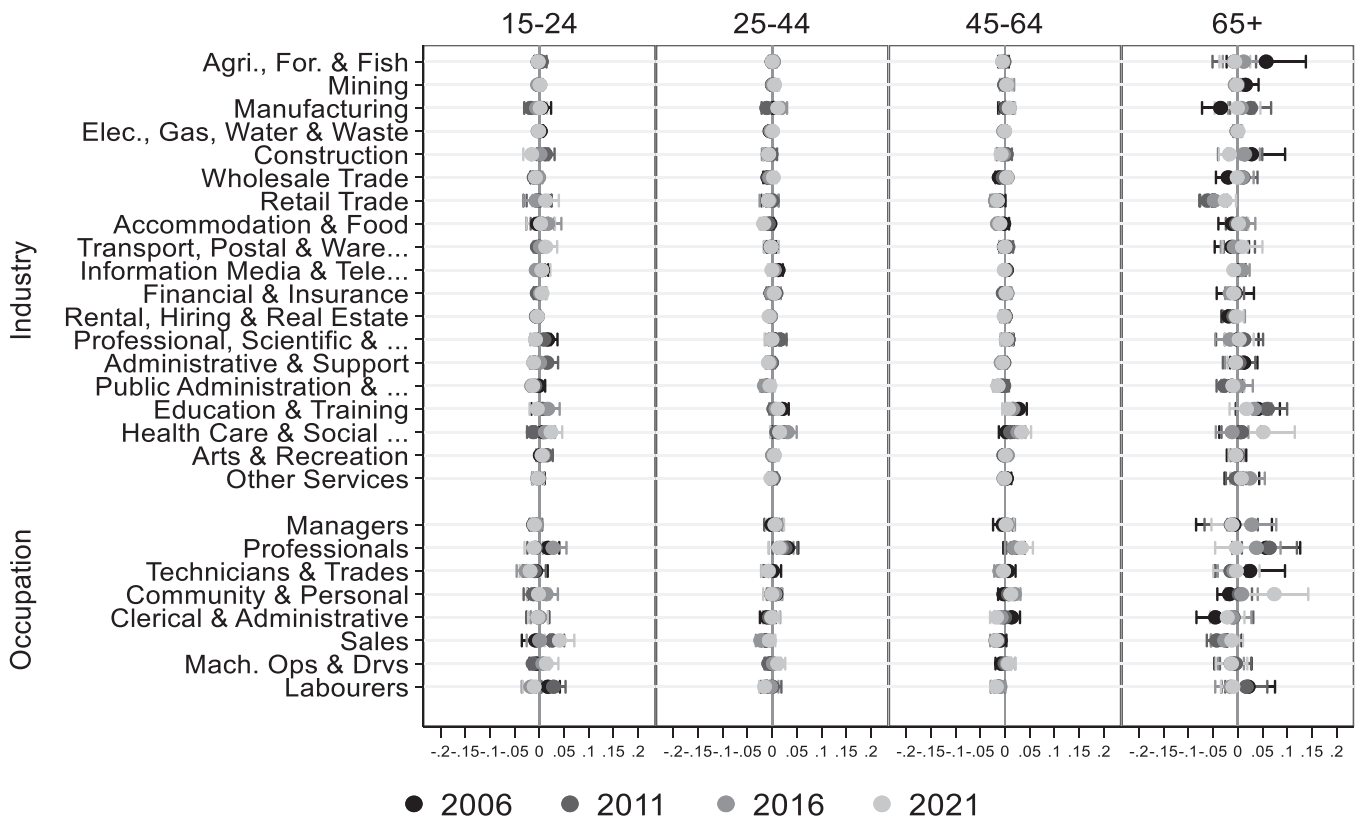


FIGURE 2 | Difference between HILDA Survey estimates and 2006–2021 Census for industry and occupation of employed persons, by age group.

have the interviewer help guide the selection of Person 1 and 2 in the household, so will provide a more accurate picture of the relationships in the household. A comparison between the HILDA Survey estimates and the GSS is provided in Figure 3 and shows very good alignment for the social marriage and registered marriage variables.⁷

4.2 | Year of Arrival

The HILDA Survey does not have a natural mechanism to add recent immigrants into the sample (aside from immigrants joining HILDA households or the top-up sample that was added in 2011). As a result, we expect to see some drift in the proportion of recent arrivals. In 2006, the HILDA Survey estimate of the proportion of people arriving in the last 5 years is lower than reported in the Census. This is concentrated in the two youngest cohorts as immigrants are typically younger than the general population.

In 2011, the HILDA Survey estimates for the year of arrival realign with the Census estimates due to the recent immigrants recruited as part of the top-up sample. Note, however, that these immigrants are given higher weights than other sample members as they only have one chance of selection (in 2011) compared to other people who have had two chances of selection (in 2001 and 2011). This results in wider confidence intervals for the estimates of recent immigrants.

In 2016 and 2021, the gap between the HILDA Survey and Census estimates for those arriving after 2011 widens. Future immigrant top-up samples will address this gap.

4.3 | Highest Year of School Completed

For the most part, there is strong alignment between the HILDA Survey and Census for the highest year of school completed. We note two exceptions to this. First, the proportion of people reporting they have completed Year 12 in the 15–24 age cohort is higher in the HILDA Survey than in the Census. However, when restricting the analysis to those aged 19–24 these differences are largely eliminated which suggests the problem is with the 15–18 year olds still at school. Second, there is a lower proportion of people in the oldest cohort reporting they have completed Year 12 in the HILDA Survey compared to the Census. It is not clear why this is, but may be due to the higher rate of item nonresponse in the Census (see Table A1 in the Appendix), where those with lower levels of education being perhaps more likely to miss answering this question.

4.4 | Highest Post-School Qualification

Comparison of the estimates for the highest post-school qualification indicates that the HILDA Survey has a higher proportion of people with some form of post-school qualification than the Census. This is primarily due to differences in the proportion of people with certificates (for the three oldest cohorts) but there is also evidence of a slightly higher proportion of people aged 45–64 with a bachelor (or higher) degree. The difference in certificates may stem from how the two sources ask about educational qualifications. In the HILDA Survey, the respondent is asked to list all of their qualifications in their first interview, and this is added to year-on-year as they study and obtain new qualifications. In the Census, the respondent is

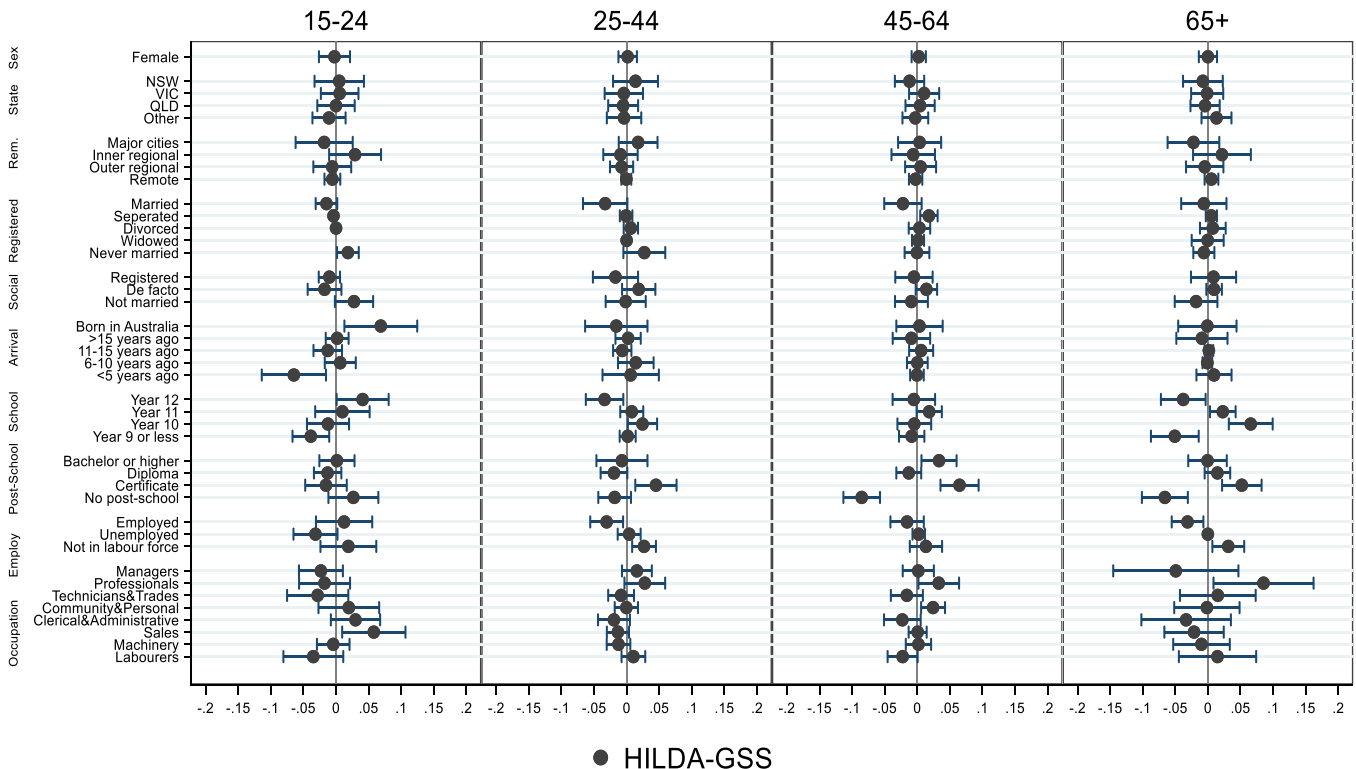


FIGURE 3 | Difference between HILDA Survey demographic estimates and General Social Survey (GSS), 2014, by age group.

asked to state their highest qualification and that of the other members of their household. It may be that some qualifications (the ones that take a shorter amount of time to obtain or perhaps are less relevant to them now) are forgotten in this process. There are four pieces of evidence that support this claim. First, the proportion of certificates and 'no-post school qualification' best align in the 15–24 age cohort, thereafter slowly and increasingly drifting apart as the cohort gets older and are more likely to forget their certificates. Second, there is a high rate of nonresponse to this question in the Census of 6%–13% for the youngest three cohorts and 14%–20% for the oldest cohort (see Table A1 in the Appendix), thereby indicating people have difficulty answering this question. Third, the comparison between the HILDA Survey estimates with those from the GSS provided in Figure 3 also shows differences in the proportion of people with certificates, but the differences are not as large as those found with the Census, suggesting that the presence of an interviewer to help probe for correct response reduces these errors. Fourth, the comparison of HILDA Survey and Census estimates from 2001 (shown in Figure A2 in the Appendix) also shows the HILDA Survey estimates for people with certificates are higher than the Census, suggesting that asking about all qualifications rather than just the highest may elicit a different result for the highest qualification.

Regarding the higher estimates of people aged 45–64 with a bachelor degree, this may be due to selective nonresponse in the HILDA Survey or selective reporting in the Census. In the comparison of the HILDA Survey to the GSS, we also see a higher estimate in the HILDA sample for this age group (HILDA 2014: 0.272; GSS 2014: 0.238; HILDA 2016: 0.279; Census 2016: 0.242). Both the Census and the GSS ask the respondent to report their highest qualification, and some people may report the most relevant qualification (to their occupation) rather than the highest one, whereas in the HILDA Survey respondents are asked to report all of their qualifications, which should help eliminate this problem. If the cause is selective nonresponse, we would expect to see this same issue in at least some of the other cohorts. It could actually be a combination of both issues.

4.5 | Employment

The HILDA Survey and Census estimates for employment align quite well with the exception of the youngest cohort (aged 15–24). The GSS estimates for this youngest cohort; however, are close to the HILDA Survey estimates (shown in Figure 3). This indicates that the likely reason for the differences in estimates identified here is due to differences in methodology. The Census questions are typically answered by one person on behalf of others in the household, thus the responses are typically less reliable than asking each person individually (as is done in the HILDA Survey and the GSS). Moreover, the questions asked in the HILDA Survey and the GSS to determine employment status are much more comprehensive than those in the Census as they seek to tease out various nuances of being employed, unemployed or not in the labour force. Item non-response in the Census of around 5% compared to 0.2% in the HILDA Survey which may also have an impact.

4.6 | Industry and Occupation

The industry and occupation estimates from the HILDA Survey are similar to those from the Census. Note that the greater variability in the difference between the two sources for the youngest and oldest cohorts is due to the smaller number of people employed in these age groups.

Some small differences may occur due to differences in how the industry and occupation questions are asked and how the responses are collected and then coded. In the HILDA Survey, the interviewer probes the respondent to ensure sufficient details are provided. For the industry question, the Census asks for the employer's business name, a description of the industry and the main goods produced or services provided, whereas the HILDA Survey asks just for the description of the industry (with relevant probing). The HILDA Survey and Census questions for occupation ask for the occupation title along with tasks and duties (with relevant probing for the HILDA response). These text responses are then coded to the standardised code frames. A comparison between the occupation estimates from the HILDA Survey and the GSS (where the collection methodology is closer) is provided in Figure 3, and it shows strong alignment between the two surveys.

5 | Conclusions

In this comparison of HILDA Survey estimates with Census estimates across four time points and four age groups, we find strong alignment across many of the variables examined. Some differences likely occur because of the different questions or recall periods used (e.g., for certificate qualifications) or differences in the collection methodology (e.g., employment of 15–24 year olds). Further evidence of the representativeness of the HILDA Survey was confirmed through comparisons to the GSS, whose methodology is closer to the HILDA Survey. As a result of these comparisons with the Census (and the GSS), we have demonstrated that the HILDA Survey is largely representative of the Australian population.

There is, however, one area where the differences between the two sources are real and of concern. This is the proportion of immigrants aged 44 and under. This is because the HILDA Survey does not have a natural way to include recent immigrants into the sample unless they live in a HILDA household or they are incorporated as part of a top-up sample. A difference is revealed for these younger age groups as immigrants tend to be young (Productivity Commission 2016). A correction was made for this in 2011 in the HILDA Survey with the addition of the general top-up sample, but the issue reappears in 2016 and 2021. To address this issue, a large top-up sample of recent immigrants is being added to the HILDA Survey sample in 2024 and 2025. Regular immigrant sample top-ups are needed thereafter to help ensure the HILDA Survey sample retains its contemporary representativeness.

A limitation of this article is that the comparisons are purely cross-sectional. While we would ideally like to compare the HILDA Survey estimates to longitudinal data sources, there is a

paucity of data in this area. Nevertheless, this is changing with linked administrative data such as ATO ALife and the Person Level Integrated Data Asset becoming available for research purposes (Polidano et al. 2020; Productivity Commission 2024). Fischer et al. (2025) recently compared the ATO ALife Family File with the HILDA Survey and, although the comparisons were still only cross-sectional, they found the comparison favourable.

Finally, it is worth noting that this article is part of a broader research agenda comparing the HILDA Survey results to official statistics. A similar analysis was undertaken for the HILDA Survey's child sample incorporating those who were too young to be interviewed in wave 1 and those subsequently born into the sample (Watson 2022b). Another analysis compared the rates of moving house in the HILDA Survey to those from official data sources (Watson 2020). Analyses of the representativeness of the working age sample members and older age sample members are underway.

Acknowledgements

This article uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Survey is conducted by the Australian Government Department of Social Services (DSS). The findings and views reported in this article, however, are those of the authors and should not be attributed to the Australian Government, DSS, or any of DSS' contractors or partners. Open access publishing facilitated by The University of Melbourne, as part of the Wiley - The University of Melbourne agreement via the Council of Australian University Librarians.

Data Availability Statement

The HILDA data are available from the Australia Data Archive (ada.edu.au).

Endnotes

¹For a list of publications that the HILDA team at the Melbourne Institute are aware of, see <https://melbourneinstitute.unimelb.edu.au/hilda/publications>.

²Fortunately, this attrition has not translated to a sizeable drop in the overall sample size as it has been offset by the inclusion of new entrants, wave 1 children becoming an age to be interviewed, and (eventually) new births.

³A small number of immigrant households in New South Wales and Victoria were added in 2023 via a trial of the method adopted in 2024–2025.

⁴Mesh Blocks are the smallest geographic area defined by the ABS with most containing 30 to 60 dwellings.

⁵A similar comparison to the 2001 Census is provided at the end of the Appendix for completeness but could not be included with the results for subsequent Censuses due to differences in the variables, age groups and tables available. The findings are very similar.

⁶The HILDA Survey standard errors are calculated via the Taylor Series approximation method using the weight, area identifier and stratification.

⁷Note the 2021 Census form asked additional questions about individuals away on Census night including their relationship to Peron 1 and 2. This may help align the Census and GSS estimates for social marital status.

References

- Australian Institute of Health and Welfare. 2024. *Profile of Australia's Population*. AIHW. www.aihw.gov.au/reports/australias-health/profile-of-australias-population.
- Australian Bureau of Statistics. 2015. *General Social Survey: Summary Results, Australia, 2014* (Cat. No. 4159.0). <https://www.abs.gov.au/statistics/people/people-and-communities/general-social-survey-summary-results-australia/2014>.
- Australian Bureau of Statistics. 2019. *Jobs Growth Over the Past 25 Years*. Australian Bureau of Statistics. www.abs.gov.au/articles/jobs-growth-over-past-25-years.
- Australian Bureau of Statistics. 2020a. Marriages and Divorces, Australia 2019. <https://www.abs.gov.au/statistics/people/people-and-communities/marriages-and-divorces-australia/latest-release>.
- Australian Bureau of Statistics. 2020b. *Twenty Years of Population Change*. Australian Bureau of Statistics. www.abs.gov.au/articles/twenty-years-population-change.
- Australian Bureau of Statistics. 2022a. *2021 Census Overcount and Undercount*. ABS Website. <https://www.abs.gov.au/statistics/people/population/2021-census-overcount-and-undercount/latest-release>.
- Australian Bureau of Statistics. 2022b. Education in Australia – From abc to A's, B's and C's. Retrieved 27 January 2024. <https://www.abs.gov.au/articles/education-australia-abc-bs-and-cs#primary-and-secondary-school>.
- Australian Bureau of Statistics. 2022c. *Understanding Data Quality*. ABS. <https://www.abs.gov.au/census/guide-census-data/census-methodology/2021/understanding-data-quality#dwelling-response-rates>.
- Bethlehem, J., F. Cobben, and B. Schouten. 2011. *Handbook of Non-response in Household Surveys*. John Wiley & Sons.
- Cabrera-Álvarez, P., N. James, and P. Lynn. 2023. *Panel Attrition in the General Population Sample and the Immigrant and Ethnic Minority Boost of Understanding Society* (Understanding Society Working Paper Series), Issue. <https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2023-03.pdf>.
- DSS & Melbourne Institute. 2023. The Household, Income and Labour Dynamics in Australia (HILDA) Survey, General Release 22 (Waves 1–22), ADA Dataverse, V7. <https://doi.org/10.26193/R4IN30>.
- Duffy, D., and N. Sastry. 2012. *An Assessment of the National Representativeness of Children in the 2007 Panel Study of Income Dynamics* (12-01). <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.5684&rep=rep1&type=pdf>.
- Duncan, G. J., and D. H. Hill. 1989. "Assessing the Quality of Household Panel Data: The Case of the Panel Study of Income Dynamics." *Journal of Business & Economic Statistics* 7, no. 4: 441–452.
- Fischer, D., J. Holland, M. Chan, C. Polidano, H. Vu, and R. Wilkins. 2025. "The ATO ALife Family File." *Australian Economic Review, Early View* 58, no. 2: 163–172.
- Fitzgerald, J., P. Gottschalk, and R. Moffitt. 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *Journal of Human Resources* 33: 251–299.
- Gray, R., P. Campanelli, K. Deepchand, and P. Prescott-Clarke. 1996. "Exploring Survey Non-Response: The Effect of Attrition on a Follow-Up of the 1984-85 Health and Life Style Survey." *Statistician* 45, no. 2: 163–183.
- Groves, R. M., and M. P. Couper. 2012. *Nonresponse in Household Interview Surveys*. John Wiley & Sons.
- Harding, S., L. Liddle, P. McDonald, P. Morrison, D. Trewin, and S. Walters. 2022. *Report on the Quality of 2021 Census Data*. <https://www.abs.gov.au/census/about-census/census-statistical-independent-assurance-panel-report>.
- Hill, D. H., and R. J. Willis. 2001. "Reducing Panel Attrition: A Search for Effective Policy Instruments." *Journal of Human Resources* 36: 416–438.

Lepkowski, J. M., and M. P. Couper. 2002. "Nonresponse in the Second Wave of Longitudinal Household Surveys." In *Survey Nonresponse*, edited by R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little. Wiley.

Lynn, P., P. Cabrera-Álvarez, and P. Clarke. 2023. "Sample Composition and Representativeness on Understanding Society." *Fiscal Studies* 44, no. 4: 341–359. <https://doi.org/10.1111/1475-5890.12357>.

Polidano, C., A. Carter, M. Chan, et al. 2020. "The ATO Longitudinal Information Files (ALife): A New Resource for Retirement Policy Research." *Australian Economic Review* 53, no. 3: 429–449.

Productivity Commission. 2016. *Migrant Intake Into Australia*. From Productivity Commission Website: <https://www.pc.gov.au/inquiries/completed/labour-mobility/report>.

Productivity Commission. 2024. *A-PLIDA-nalysis: Using PLIDA for Public Policy Research and Reporting*. <https://www.pc.gov.au/research/completed/plida/plida.pdf>.

Richardson, S. 2013. "A Reflection on the Household, Income and Labour Dynamics in Australia Survey." *Australian Economic Review* 46, no. 2: 216–222.

Summerfield, M., B. Garrard, R. Kamath, et al. 2023. *HILDA User Manual – Release 22*. Melbourne Institute of Applied Economic and Social Research. https://www.melbourneinstitute.unimelb.edu.au/_data/assets/pdf_file/0020/4815110/HILDA-User-Manual-Release-22.0.pdf.

Uhrig, S. C. N. 2008. The Nature and Causes of Attrition in the British Household Panel Study (*ISER Working Paper Series, Issue*. U.o.E. Institute for Social and Economic Research. <https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2008-05.pdf>.

Watson, D. 2003. "Sample Attrition Between Waves 1 and 5 in the European Community Household Panel." *European Sociological Review* 19, no. 4: 361–378.

Watson, N. 2020. "Measuring Geographic Mobility: Comparison of Estimates From Longitudinal and Cross-Sectional Data." *Survey Research Methods* 14, no. 1: 1–18.

Watson, N. 2022a. "Choosing Who to Follow: The Long-Run Impact of Following Rules on the Sample Size and Composition of Household Panel Surveys." *Survey Research Methods* 16, no. 3: 301–328.

Watson, N. 2022b. "New Generations of Respondents: Assessing the Representativity of the HILDA Survey's Child Sample." *Longitudinal and Life Course Studies* 13, no. 3: 465–489. <https://doi.org/10.1332/175795921X16349086588358>.

Watson, N., and M. Wooden. 2013. "Adding a Top-Up Sample to the Household, Income and Labour Dynamics in Australia Survey." *Australian Economic Review* 46, no. 4: 489–498.

Watson, N., and M. Wooden. 2009. "Identifying Factors Affecting Longitudinal Survey Response." In *Methodology of Longitudinal Surveys*, edited by P. Lynn, 157–181. Wiley.

Williams, R. 2012. *The Policy Providers: A History of the Melbourne Institute of Applied Economic and Social Research, 1962–2012*. Melbourne University Publishing.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting Figure A1: Difference between HILDA Survey estimates and 2001 Census, by age group. **Supporting Table A1:** Percentage of item non-response in the Census variables. **Supporting Table A2:** HILDA Survey estimates for 2006–2021, by age group. **Supporting Table A3:** HILDA Survey estimates for 2001, by age group.