



Australian Government

Department of Families, Housing,
Community Services and Indigenous Affairs

Growing up in Australia

The Longitudinal Study of Australian Children (LSAC)



LSAC TECHNICAL PAPER No 7

October 2011

Validating Income in the Longitudinal Study of Australian Children

Killian Mullan and Gerry Redmond

*Social Policy Research Centre
University of New South Wales*

Acknowledgements:

This report uses unit record data from the Longitudinal Study of Australian Children (LSAC), the Household Income and Labour Dynamics Australia survey (HILDA) and the Survey of Incomes and Housing (SIH) 2003–04. LSAC and HILDA were initiated and funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and are being undertaken in partnership respectively with the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS), and the Melbourne Institute of Social and Economic Research. SIH is used with the permission of the ABS. Comment and review from Bruce Bradbury, Ilan Katz, staff at FaHCSIA and members of the LSAC Data Expert Reference Group are gratefully acknowledged. The findings and views reported in this paper are those of the authors, who are also responsible for any errors.

This report has been completed under FaHCSIA's Social Policy Research Services Agreement (2005–2009) with the SPRC. The opinions, comments and analysis expressed in this document are those of the authors and do not necessarily represent the views of FaHCSIA or of the Minister and cannot be taken in any way as expressions of Government policy.

Gerry Redmond
Social Policy Research Centre
University of New South Wales
g.redmond@unsw.edu.au

For more information

Research Publications Unit
Research and Analysis Branch
Australian Government Department of Families, Housing, Community Services and
Indigenous Affairs
PO Box 7576
Canberra Business Centre ACT 2610
Phone: (02) 6244 5458
Fax: (02) 6133 8387
Email: publications.research@fahcsia.gov.au

Contents

Executive Summary	5
1 Introduction	8
2 How income is recorded in LSAC and other surveys	10
2.1 Income in the Longitudinal Study of Australian Children	10
2.2 Income in the Survey of Incomes and Housing Costs	14
2.3 Income in the Household Income and Labour Dynamics Australia Survey	17
2.4 Comparability of surveys: timing issues	20
3 Literature Review	22
3.1 Why is collecting information on income difficult?	22
3.2 Validating income in surveys	23
4 Analysis.....	26
4.1 Analysis plan.....	26
4.2 Comparing samples.....	26
4.3 LSAC income data: item non-response	31
4.4 Comparing averages and distributions.....	38
4.5 The distribution of men’s and women’s weekly income in LSAC, SIH and HILDA	41
4.6 Income over time: LSAC and HILDA compared	46
5 Discussion	52
6 Conclusion	56
Appendixes.....	58
References.....	66
List of shortened forms	68
Endnotes.....	68

List of Tables

Table 1:	Characteristics of families in LSAC K Cohort Wave 1, SIH 2003–04 families with children aged 3 to 4 years and aged 3 to 9 years, and HILDA Wave 4, families with children aged 4 to 5 years and aged 4 to 9 years ..	27
Table 2:	Item non-response rates for the banded income unit and the individual income questions in LSAC K Cohort Wave 1	32
Table 3:	Odds ratios for determinants of item non-response to banded income unit income question in LSAC K Cohort Wave 1	34
Table 4:	Odds ratios for determinants of item non-response to Parent 1 individual income question in LSAC K Cohort Wave 1	35
Table 5:	Odds ratios for determinants of item non-response to Parent 2 individual income question in LSAC K Cohort Wave 1	36
Table 6:	Odds ratios for determinants of item non-response to Parent 1 and/or Parent 2 individual income question in LSAC K Cohort Wave 1 (couple families only)	37
Table 7:	Multinomial logit regression coefficients relating to Parent education for a model of item non-response to individual income questions in LSAC K Cohort Wave 1 (couple parents only)	38
Table 8:	Men’s mean weekly income in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample across select socioeconomic and demographic characteristics (\$)	39
Table 9:	Women’s mean weekly income in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample across select socioeconomic and demographic characteristics (\$)	40
Table 10:	Mean weekly income in the bottom, second, third and top quartiles of the income distribution for men and women in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample (\$).....	43
Table 11:	Mean household weekly income in the bottom, second, third and top quartiles of the income distribution in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years	

sub-sample (\$).....	44
Table 12: Sequence of waves for LSAC and HILDA with the corresponding age of the LSAC study child.....	47
Table 13: Changes in ranking within the income distribution for men, women and households in LSAC and HILDA (per cent).....	50

Appendix Tables

Table A1: Comparison of characteristics of LSAC with HILDA Wave 3 and Wave 4 sub-samples (per cent, unless otherwise indicated).....	58
Table A2: Comparison of incomes of LSAC with SIH age 3 to 9 years and HILDA Wave 4 age 4 to 9 years sub-samples (\$).....	59
Table A3: Men's and women's characteristics in LSAC, SIH age 3 to 4 years sub-sample and HILDA age 4 to 5 years sub-sample (unweighted N).....	60
Table A4: Linearised standard errors associated with men's and women's mean weekly incomes in LSAC, SIH age 3 to 4 years sub-sample and HILDA age 4 to 5 years sub-sample (\$).....	61
Table A5: Changes in men's ranking in the income distribution from LSAC Waves 1 to 2 and Waves 2 to 3.....	62
Table A6: Changes in women's ranking in the income distribution from LSAC Waves 1 to 2 and Waves 2 to 3.....	62
Table A7: Changes in men's ranking in the income distribution from HILDA Waves 4 to 6 and Waves 6 to 8.....	63
Table A8: Changes in women's ranking in the income distribution from HILDA Waves 4 to 6 and Waves 6 to 8.....	63
Table A9: Changes in ranking of households in the income distribution from LSAC Waves 1 to 2 and Waves 2 to 3 (unweighted N).....	64
Table A10: Changes in ranking of households in the income distribution from HILDA Waves 4 to 6 and Waves 6 to 8 (unweighted N).....	64

List of Figures

Figure 1:	Range of possible birth dates for children in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample	21
Figure 2:	Range of dates when most interviews were carried out in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample, HILDA Wave 3, age 3 to 4 sub-sample and HILDA Wave 4, age 4 to 5 years sub-sample	21
Figure 3:	Age distribution of fathers in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample	29
Figure 4:	Age distribution of mothers in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample	30
Figure 5:	Men’s average weekly income across the income distribution.....	41
Figure 6:	Women’s average weekly income across the income distribution	42
Figure 7:	Average weekly income unit income across income percentiles.....	44
Figure 8:	Cumulative distribution of banded income unit income in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample	46
Figure 9:	Men’s mean weekly income across LSAC Waves 1 to 3 and HILDA Waves 4 to 8.....	48
Figure 10:	Women’s mean weekly income across LSAC Waves 1 to 3 and HILDA Waves 4 to 8.....	48
Figure 11:	Men’s and women’s combined mean weekly income across LSAC Waves 1 to 3 and HILDA Waves 4 to 8	49

Executive Summary

Income is one of the most important pieces of information about individuals and households available to researchers in social science; but collecting good data on income is difficult. There are many reasons for this. People may not know their income or may not be willing to divulge it. Even if they are willing, their knowledge of their income may not be complete. This is not surprising, considering the variety of sources from which individuals and households derive their income (including non-cash income), the multiple income streams which exist within households, and the variability of the periods over which income is attained. Add to this a complex and dynamic tax-benefit system, and it is possible to see why fully knowing one's personal or household income at a distinct point in time is not straightforward. Surmounting these issues is not trivial, especially in surveys where the collection of income data is not the main aim.

This report seeks to assess how the measure of income collected in the Longitudinal Study of Australian Children (LSAC) compares with measures of income from two large-scale Australian surveys designed completely, or in large part, to collect data on income. These studies are the Survey of Income and Housing Costs (SIH) and the Household, Income and Labour Dynamics in Australia (HILDA) Survey. Both surveys ask all household members (15 years and over) a detailed set of questions about their income, and both impute missing income data. Though not perfect, these surveys can be viewed as providing exemplars for income measurement in survey data in Australia. In contrast, LSAC asks fewer questions of a single respondent and makes only limited imputation of missing income data. The question for this report is whether any or all of these factors negatively affect the quality of the measure of income in LSAC.

In Wave 1, LSAC respondents are asked to provide information on their and their partner's income in dollars and to indicate the combined income of both parents in the household, from a list of 15 income bands (plus 'nil income' and 'negative income' categories). Most of this report is concerned with an analysis of non-response in the Wave 1 LSAC data and with the comparison of Wave 1 LSAC data with corresponding data in SIH and HILDA.

We find that item non-response is relatively low in LSAC for the banded income unit income question. Non-response is highest for fathers' income (which is not surprising, given that most respondents are mothers), while one in five respondents fails to provide information on the individual income of either the mother and/or the father. This means that, while item non-response is lowest for banded household income, it is highest for the measure of the combined individual income of mothers and fathers.

Families with someone who is self-employed are significantly and consistently less likely to provide information on income. Highly educated couple households are also less likely to respond to the individual income questions for either parent. Lone mothers are more likely to provide individual income data, but there is no difference between lone and partnered mothers in terms of their response for household income in bands.

The main finding of the report is positive, in that measures of income in LSAC are broadly comparable with measures derived from both SIH and HILDA. Average income for men in LSAC and SIH is quite similar, while average income for men in HILDA is about one-fifth greater than that in LSAC. Women's average incomes in all three surveys are similar, but, again, income in HILDA is slightly greater. However, average income for lone mothers in LSAC is significantly lower than for lone mothers in SIH and HILDA. The pattern for men is consistent across the income distribution; for women, however, there are significant differences in means for each quartile of incomes except the bottom. Income unit income (the combined individual income of Parents 1 and 2) in LSAC and SIH is very similar but diverges from HILDA towards the upper quartile of the income distribution. Finally, the distribution of banded income unit income is very similar across all three surveys.

The report looked also at income measured across all three waves of LSAC and compared this with corresponding waves of data in HILDA (Waves 4, 6 and 8). Here, the analysis was restricted to observations that responded in all relevant waves. We found that men's incomes in the corresponding LSAC and HILDA waves were more similar than for the overall comparison at Wave 1 only. Furthermore, they appeared to be converging towards parity by the third wave of LSAC (HILDA Wave 8). In contrast, we observed a significantly wider gap in the measure of income for women in LSAC and HILDA across all three waves than was apparent in the Wave 1 analysis. However, combined income of men and women was remarkably similar between LSAC and HILDA among respondents in couple households who provided income data in all waves. Finally, broad patterns of change in the relative rankings of men, women and households in the income distribution were similar in LSAC and HILDA. This suggests some stability over time in LSAC measures of individual income.

Some outstanding issues remain from this report. There are many instances where the wording of questions between surveys is different. Perhaps more worryingly, there are instances where the wording of questions within LSAC changes across waves. The extent to which this impacts upon comparisons across surveys or across waves within LSAC has not been considered in this report. Future research on the manner in which respondents understand the wording of the questions should be conducted. Another

limitation of this report is that it did not consider income from the infant cohort, which could perhaps be addressed in further work. The analysis of income measured across waves in this report is only a first step, and future research should build upon this, considering the importance of income and longitudinal data for social science research.

1 Introduction

LSAC is a widely used instrument for tracking children's development in Australia. Common applications of LSAC data involve analysing the relationship between parents' socioeconomic status and children's development—this is a key issue in both Australian and international research (Bradbury, 2007; Haveman and Wolfe, 1994; Katz and Redmond, 2009; Khanam et al., forthcoming; Mayer, 1997). Family income is clearly important in this stream of research, both in its own right as an indicator of a family's ability to purchase goods and services that can support a child's wellbeing and development and as a proxy for broader concepts of socioeconomic status. Indeed, it is difficult to understate the importance of income, both as an indicator of family socioeconomic status in current research on child development and as a policy lever for supporting families with children. As Duncan et al. (1998: 421) state, 'raising the incomes of poor families will enhance the abilities and attainments of their children.' However, most analyses of LSAC that use family income have tended not to consider potential shortcomings with these data in any detail. This is surprising, given that income is calculated in LSAC from just a few summary questions asked of a single household member.

This report examines how measures of income in LSAC compare with measures of income as reported in two large-scale Australian surveys: the Survey of Incomes and Housing Costs (SIH) and the Household Income and Labour Dynamics Australia (HILDA) Survey. Both these surveys attempt to measure income in as comprehensive a manner as possible. They ask a large number of detailed questions on personal and household incomes; they request that all adults (15 years and over) in the household provide information on their income; and they impute missing income data. Methodological research suggests that these steps should allow for a more accurate estimation of personal and household income.

LSAC is not primarily designed to collect income data. It therefore asks a single respondent fewer questions about income, and it carries out only limited imputation of missing income data. This report asks whether any or all of these three factors affects the quality of LSAC income measures, adjudged here as the extent to which measures of income in LSAC are comparable with those in SIH and HILDA. No survey is immune from measurement error—this is certainly true of both SIH and HILDA. However, our comparison simply acknowledges that HILDA and SIH collect more information from respondents, which may, other things being equal, lead to a better measure of income. In this sense, they provide a useful benchmark against which to evaluate income data collected in LSAC. In other words: if income data in SIH and HILDA represent 'true' pictures of the incidence and distribution of income in

Australia, how closely does income as reported in LSAC match these distributions?

Addressing this question is not straightforward. Firstly, while SIH and HILDA aim to be representative of the entire household population in Australia, LSAC aims to be representative only of two cohorts of children, the first born in 1999 and the second born in 2003. Comparison, therefore, needs to be made between LSAC and appropriate subsets of the other two datasets. Our approach to this issue is set out in detail in the report. Secondly, as this analysis makes clear, SIH uses a different income concept to that in HILDA. This raises further issues relating not only to comparability, but also to what income concept LSAC respondents have in mind when they answer questions on their own and their partner's income. We describe and discuss this issue in the report, though we cannot address it directly.

Our analysis proceeds in three stages and draws on methods proposed by Micklewright and Schnepf (2010). Firstly, we examine characteristics associated with non-response to income questions in LSAC. This allows us to assess the extent to which the lower amount of data imputation in LSAC, compared with SIH or HILDA, may impact upon the comparability of measures of income. Secondly, we compare mean incomes in LSAC with those in selected subsets of SIH and HILDA for men and women, comparing distributions of men's and women's individual incomes, and income unit incomes, in the three surveys. Finally, we briefly look at dynamics—the extent to which patterns of change in income differ in LSAC and HILDA.

The remainder of this report is organised as follows: Section 2 describes how income is recorded in LSAC, SIH and HILDA. Section 3 presents a review of the literature on income validation. Results are presented in Section 4 and discussed in Section 5. Section 6 concludes with an assessment of the strengths and weaknesses of the study and recommendations for further research.

2 How income is recorded in LSAC and other surveys

2.1 Income in the Longitudinal Study of Australian Children

The purpose of LSAC is ‘to provide the database for a comprehensive understanding of Australian children’s development in the current social, economic and cultural environment, and hence to become a major element of the evidence base for policy and practice regarding children and their families.’ (Sanson et al., 2002: v) The study is part of a growing body of large-scale, nationally representative longitudinal studies that track children’s development across a number of rich countries. Similar studies have been, or are now being, carried out in the US, New Zealand, Ireland and the UK, for example. These international studies, as well as smaller scale studies carried out previously in Australia, have provided a template for the construction of LSAC.

LSAC was launched in 2004 and is ongoing, with detailed information collected from responding families every two years. Data are being collected from two separate samples of children and their families, the first aged 3 to 17 months in 2003–04 (the B Cohort) and the second aged 4 to 5 years in 2003–04 (the K cohort). The sampling frame for the two age cohorts was taken from the Medicare Australia enrolment and activity databases held by the Health Insurance Commission (Soloff et al., 2003). We focus our attention in this analysis on the K Cohort. In general, the recruitment rate of K cohort families was moderate. Of the original sample of 10,275 children aged 4 to 5 years selected from the Medicare enrolments database, 4,983 were successfully recruited to the study, giving an overall response rate of 48 per cent. Reasons for non-response included database mismatches (3 per cent), non-contacts (14 per cent) and refusals (35 per cent) (Soloff et al., 2006). As with all longitudinal studies, attrition has reduced the number of responding families in each wave. Of 4,983 Wave 1 observations, 519 (10.4 per cent) did not participate at Wave 2, and 652 (13.1 per cent) did not participate at Wave 3. The vast majority of primary carers who completed questionnaires were mothers. Only 144 out of 4,983 primary carers at Wave 1 were fathers.

The structural features of LSAC are important for the task at hand. LSAC is not a household or family survey, but a longitudinal study of **children**. In the early years, at least, most of the information on the children is collected from parents, guardians or other responsible adults. However, since the focus of the study is not adults, priority is (quite rightly) given to reporting on the child, rather than on the adults she or he lives with.

This prioritisation of information on the child is seen in the way data on family

incomes are collected in LSAC. Information on individual and family incomes is not asked of each family member, or of each adult in the family, but of the main respondent, who is selected on the basis of being the main carer for the study child. (As noted above, in most cases, this is the mother).

At **Wave 1**, the main respondent is asked the following questions about the income of firstly herself and then her partner:

Currently, do you personally receive income from any of these sources?

(MARK ALL THAT APPLY)

- *wages or salary*
- *profit or loss from own business or share in a partnership*
- *profit or loss from rental company*
- *dividends or interest*
- *any Government pension or allowance/Income Support*
- *Child Support or maintenance (from ex-partner)*
- *Superannuation or Annuity*
- *Workers' Compensation*
- *Other*
- *none of the above*

(if more than one source of income is marked) Which of those just mentioned is your main source of income?

(MARK ONE ONLY)

Do you or your partner currently receive any of these government benefits, allowances or other forms of assistance?

(Respondent is prompted with a list: Parenting Payment Partnered; Parenting Payment Single; Carer Allowance; Newstart Allowance; Disability Support Pension; Family Payment (Tax Benefit) 'A'; Family Payment (Tax Benefit) 'B'; Rent Assistance; Child Care Benefit; Health Care Card; Other government allowances, pensions, payments or subsidies.)

(Respondent is asked to state in turn if she receives each payment, if her partner receives it, if both of them receive it, or if neither receives it.)

(If she reports that she and her partner receive no government payments, she is asked) *Most families receive some government payments. Are you sure you do not receive anything?*

(yes/no)

Before income tax is taken out, how much do you usually receive from all sources in

total?
(*\$*)

What period does that cover?
(*week, fortnight, month, etc.*)

(These two questions are repeated with respect to the income of the respondent's partner)

Before income tax is taken out, what is your present yearly income (for you and your partner combined)?

(INCLUDE PENSIONS AND ALLOWANCES; BEFORE TAX, SUPERANNUATION OR HEALTH INSURANCE)

(Respondent is asked to place income in one of 17 bands, including negative income, nil income and 15 positive income bands up to \$2,400 per week or more).

At Wave 1, therefore, the respondent is asked about sources of income separately for herself and her partner; she is then asked about the total amount of usual income of herself, and then of her partner, before any taxes are deducted. Finally, she is asked to state into which of seventeen bands the present combined income of herself and her partner falls. These questions were designed by the survey team in close collaboration with the Australian Bureau of Statistics, drawing on experience gained in a range of other studies, including HILDA, the Western Australian Child Health Survey and the Australian Institute of Family Studies' *Living in Australia* study.

Income questions in Waves 2 and 3 follow a broadly similar sequence. However, questions on income sources are reduced to four: wages or salary; profit or loss from own unincorporated business or share in partnership; any government pension, benefit or allowance; any other regular source. At **Wave 2**, the following question is then asked:

Before income tax is taken out, how much do you usually receive (from this source/these sources) in total?

(*\$*) (If respondent is unable to answer, interviewer is asked to prompt for their best estimate)

The respondent is then asked substantially the same set of questions about receipt of government payments and allowances as at Wave 1, except that she is not asked if she is sure if she reports receiving no such payments. Finally, she is asked about her main source of income: wages or salary; profit or loss from own unincorporated business; profit or loss from rental company; dividends or interest; any government pension or allowance; child support or maintenance; superannuation or annuity; workers'

compensation; other. She is then asked the same set of questions with respect to her partner's income.

The final questions on income at Wave 2 are also different to those at Wave 1:

The next question is about the income of members of your household aged 15 years or over, excluding yourself <and your partner>. Before income tax is taken out, how much income in total do these people usually receive from all sources?

(\$)

What period does that cover?

(week, fortnight, month, etc.)

In other words, the respondent is no longer asked to place her and her partner's combined income into one of 17 bands as in Wave 1, but instead is asked more precise information about the income of other household members. This change in the way data on incomes are collected came about as a result of new voices being added to the study in the design of Wave 2 and subsequent waves, which brought into contention the income content for these waves and introduced variation into the collection of income data. The bracketed combined family income variable was dropped in favour of what was seen to be increasing precision. As becomes apparent in later Sections, this change in method has some implications for LSAC in the context of this analysis.

Sequencing and wording of questions at **Wave 3** generally follows that used at Wave 2. There are, however, two important differences. The main question on the amount of income of the respondent with respect to herself and then her partner is worded as follows:

Before income tax, salary sacrifice or anything else is taken out, how much do you usually receive from all sources in total?

(\$)

In addition, the respondent is asked the following question with respect to herself and then her partner:

*Before income tax, salary sacrifice or anything else is taken out, how much do you usually receive from **wages and salary** (in ALL jobs) in total?*

(\$)

Finally, the question on the incomes of other household members is also worded

slightly differently:

The next question is about the income of members of your household aged 15 years or over, excluding yourself and Parent 2. Before income tax, salary sacrifice or anything else is taken out, how much income in total do these people usually receive from all sources?

(\$; if respondent is unable to answer, interviewer is asked to prompt for their best estimate.)

There are, therefore, subtle differences in the ways that questions on income are asked in all LSAC waves. It is difficult to estimate the effects of these small changes on how respondents report incomes in each wave; for example, whether asking respondents about their incomes before salary sacrifice (Wave 3) would prompt a different response in comparison with a question where no mention is made of salary sacrifice (Waves 1 and 2). For the most part, we do not address these detailed issues in this analysis. Rather, we focus on responses to income questions at Wave 1 (where the overall LSAC sample should be most representative of the relevant cohort of children in the Australian population). Where we examine trends in income, we assume that data as collected are comparable across the three waves.

A limited amount of post-collection imputation is carried out on income data as reported by respondents in LSAC. Most of the imputation concerns outliers—very large amounts or negative amounts. All reported negative amounts are recoded to minus 99. In effect, therefore, no information on negative incomes is available to LSAC survey users. Where respondents report only government payments as their income source, but also report zero or negative income or income above \$750 per week, their income total is set to missing. Very large amounts of profit or loss are also set to missing where no salary is reported. In a number of cases, where individual incomes are reported for both parents (or for just the sole parent as appropriate), but information on family income is not given, then family income is imputed from individual incomes. No income elements or income taxes are imputed.

2.2 Income in the Survey of Incomes and Housing Costs

SIH has been carried out regularly by the Australian Bureau of Statistics (under a few different names) since 1982. Its purpose is to track trends in personal, family and household cash incomes and the distribution of incomes in Australia. It is conducted under the Census and Statistics Act (1905), which means that sampled households are legally obliged to participate in the survey. The survey was carried out every four or five years between 1982 and the mid 1990s but has been carried out roughly every two years since then, most recently in 2007–08. This is a household survey, where all

individuals aged over 15 in sampled households are interviewed. The total number of households participating in the surveys has varied considerably but has mostly been around 14,000 in the earlier years and 8,000 in the more recent years.

Although SIH data have been regularly used to examine poverty, inequality and changes in incomes in Australia (Austen and Redmond, 2008; Harding, 1997; Harding and Szukalska, 1999; Redmond, 1999) and incomes in Australia in international comparison (Smeeding, 2002; UNICEF, 2005), research has also cautioned about the comparability of this data series over time, because of changes in sampling methods and changes in questions on incomes (Saunders and Bradbury, 2006). Nonetheless, because the series stretches from 1982, this dataset remains perhaps the most commonly used in Australia for examining trends in income distribution. Therefore, it is seen as the ‘gold standard’ (albeit a flawed one) against which other income surveys should be compared.

While the broad structure of the Section on incomes in SIH has remained reasonably constant through the years, questions on individual income items have changed. Here we describe questions asked of respondents to the 2003–04 SIH, the survey year that we examine most extensively in this analysis. This survey was run concurrently with, and on the same sample as, the Household Expenditure Survey. Respondents are asked about their pay to the nearest dollar from their main job, in the following sequence:

What was the total amount of your most recent pay before tax or anything else was taken out?

(\$)

Is that your usual pay?

(yes/no)

(if ‘no’) How much do you usually receive each pay?

(\$)

What period does that cover?

(week, fortnight, month, etc.)

Including leave loading, do you receive any regular bonuses from your employer which are not included in every pay?

(yes/no)

(if ‘yes’) What was the total amount of these bonuses you received in the last 12 months before tax was taken out?

(\$)

These questions are followed by a series of questions about benefits that respondents might receive from their employer. Respondents are specifically asked about housing, telephone calls and motor vehicles and whether benefits are provided through salary sacrifice schemes. Respondents are then asked about other salary sacrifice arrangements with their employers, including superannuation, computers and child care, and are also asked:

*Did you include the amount of income you salary sacrifice as part of the amount you reported earlier for wage and salary income?
(yes/no)*

Respondents are also asked about non-cash benefits received from employers, including superannuation, computers, child care, car parking, shares and low interest loans, and are asked to estimate the cash value of these benefits. In the definition of income used in this report, these non-income elements are all added to the total. Questions on most recent and usual pay (but not questions of salary sacrificing or in kind benefits) are then repeated for respondents who report having a second job as an employee. These are followed by questions on wages or salary in the last financial year and detailed questions on income from business or self-employment and the value of the business. Respondents are asked to consult tax assessments and, where necessary, receive subsequent phone calls about their tax assessments.

Questions on self employment are followed by questions on Family Tax Benefits (including whether the most recent payment included a lump sum advance or a reduction because of an earlier lump sum advance), Age Pensions and other pensions and allowances; most recently, and then in the previous financial year. Finally, respondents are asked about income from investments, rental properties, workers' compensation and other sources in the last financial year.

To summarise, the main differences in approach to the collection of income data between LSAC and SIH are as follows.

1. In LSAC, all income information for the family is collected from the person with the prime responsibility for caring for the child, while in SIH, information is collected separately from each household member aged 15 years and over.
2. While fairly detailed information on income sources is collected in LSAC, only one question is asked of the respondent about the amount of her current income and that of her partner (if she has one). In SIH, on the other hand, information is sought from each respondent in the household on his or her receipt of different

income elements and the amount received.

3. The respondent in LSAC is asked to report the present income unit income in one of fifteen bands (plus 'nil income' and 'negative income' categories). Respondents in SIH are not asked about income unit income at all. Rather, information elicited from each individual respondent in the household is used by ABS to calculate measures of income unit, family and household income.
4. While respondents to LSAC are asked about their 'current' or 'present' income, respondents to SIH are asked, for the most part, about their most recent income amount. If their most recent amount is not their usual amount, they are also asked about their 'usual' income now, as well as income in the most recent financial year.
5. Imputation strategies in the two surveys are different. Notably, income taxes are imputed in SIH, allowing for the estimation of net individual, family and household income. Where necessary, government allowances and pensions are also imputed where respondents do not know the amount they receive or give patently false amounts. In LSAC, by contrast, no taxes are imputed, and it is only possible to estimate individual and income unit gross income at Wave 1.¹
6. As data are collected from all household members in SIH, it is possible to estimate (gross or net) income at the level of: the individual; the income unit (comprising only parents and their dependent children); and the level of the household, which may include both extended family (parents, grandparents, non-dependent children, etc.) and non-family members residing in the same household. In LSAC, on the other hand, it is only possible in Wave 1 to estimate income at the level of the individual parent or at the level of the income unit in which the child lives.

2.3 Income in the Household Income and Labour Dynamics Australia Survey

The HILDA Survey is a household-based panel study that follows almost 20,000 individuals, spread across around 7,600 households, who have been interviewed each year since 2001. In each of the eight waves of interviews carried out to date, information is collected on respondents' demographic characteristics, their family arrangements, education, employment, income and assets, and subjective wellbeing. The fact that this information is repeatedly collected for the same individuals means that it is possible to track important changes in their lives, including transitions in childhood and from education to the labour market, movement between jobs and labour force status, changes in income, family formation and dissolution, and the birth

and coming-of-age of children.

Questions on income in HILDA roughly follow the main sequence of questions asked in SIH. However, rather less detail is sought from respondents.

Do you currently receive income from wages or salary?

(yes/no)

(if 'yes') What was the total gross amount of your most recent pay before tax or anything else was taken out?

(\$)

Do you know what your income from wages and salaries in this job is after tax and other deductions are taken out?

(yes/no)

(if 'yes') What was the total amount of your most recent pay after these deductions?

(\$)

And what were these deductions?

- *Taxation*
- *Superannuation contributions (employee)*
- *Union dues*
- *Health fund contributions*
- *Insurance Premium*
- *Other (please specify)*

Respondents are asked about the period covered by the last pay, and then:

And is that your usual pay?

(yes/no)

(if 'no') Looking over the last month, on average how much would you have received each week or fortnight?

(\$)

And is that before tax or after tax is taken out?

(before/after)

The entire sequence of questions is repeated for respondents who report having more than one job. Respondents are then asked about government pensions and allowances:

which ones they receive and how much they receive each fortnight. This is followed by questions (in the same format as outlined above) about respondents' wage or salary in the last financial year. Respondents are asked about income from business or self employment, from investments and royalties, and from rental properties in the last financial year. They are also asked about income from a wide range of government payments in the last financial year or about fortnightly amounts received and number of weeks for which they were received. Finally, respondents are also asked about income from workers' compensation, child support and other sources in the last financial year. Therefore, the most detailed information on respondent incomes in HILDA is collected for the last financial year. Only information on wages/salaries and government payments is collected on both a current and an annual basis. No information is recorded on Family Tax Benefit payments; these are imputed.

The HILDA questionnaire, therefore, implicitly assumes a somewhat different definition of income to that in SIH. In particular:

1. While it is possible to calculate a fairly comprehensive total of personal income on a current or past financial year basis in SIH, comprehensive income information is only available on a past-financial-year basis in HILDA.
2. HILDA ignores salary sacrificing. It is not clear from the HILDA questions whether it is expected that respondents include or exclude salary sacrifice amounts from gross earnings 'before tax or anything else is taken out', especially since the deductions that respondents are asked about do not explicitly include some items (such as motor vehicles) that could be salary sacrificed.
3. HILDA ignores income in kind. Therefore, this is not added to total income in HILDA.

The difference between SIH and HILDA in these three issues points to a dilemma that the LSAC user faces. Since only summary information is collected on income in LSAC, it is not clear whether the income that the LSAC respondent reports at Waves 1 and 2 might bear more resemblance to income as defined in SIH or income as defined in HILDA. In theory, the income reported at LSAC Wave 3, where respondents are explicitly asked to report income before salary sacrifice, should be closer in concept to that in SIH. However, it is still unclear in Wave 3 how respondents actually interpret the questions they are asked about their incomes. We discuss research literature on how survey respondents understand income questions in Section 3.

2.4 Comparability of surveys: timing issues

To conduct this study, it is necessary to ensure, as far as possible, that samples are comparable across data sources. Particular features of LSAC data introduce complexity to the task of matching samples, exacerbated by limited (but not identical) information in some respects in both HILDA and SIH. In this Section, we outline our approach to matching the samples and provide some background information on sample design pertinent to our statistical analyses.

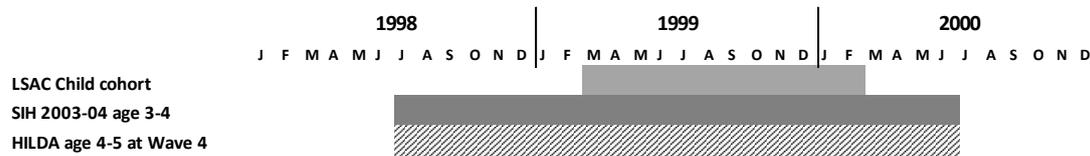
Dates of birth

The child cohort of LSAC includes children born between March 1999 and February 2000. Interviews at Wave 1 were carried out between March 2004 and March 2005, but around 85 per cent of interviews were conducted prior to September 2004. Interviews at Waves 2 and 3 were carried out in the years starting March 2006 and 2008, respectively, again with the vast majority being completed by September of those years.

Interviews in SIH were evenly carried out over a twelve-month period between July 2003 and June 2004, to ensure that data were representative of income across the year. Interviews for each wave of HILDA commence in August of each year, and over 95 per cent of interviews are completed by December of the same year. Therefore, at Wave 3, HILDA interviews commenced in August 2003 and had largely been completed by December 2003; at Wave 4, most interviews took place between August and December 2004.

While the population for the LSAC child cohort sample comprises children born between particular dates, the populations for SIH and HILDA samples comprise people of all ages living in private households in Australia at particular dates. In order to compare LSAC with SIH and HILDA, therefore, it is necessary to select observations from these latter two datasets that are most representative of the population from which the LSAC sample is drawn. Ideally, this would mean identifying children in these two samples who were born between March 1999 and February 2000.² Data limitations mean that it is not possible to find such children. In the public access SIH, it is only possible to identify children who were aged 3 to 4 years (age last birthday) at the time of interview. In HILDA, it is possible to identify children by their age on 30 June preceding interview. Figure 1 shows graphically how the ranges of dates of birth of children in the three surveys compare. In effect, births of children aged 3 to 4 in SIH and aged 4 to 5 at HILDA Wave 4 take place over a longer time interval than those in LSAC.

Figure 1: Range of possible birth dates for children in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample



Source: Watson (2010), Australian Institute of Family Studies (2010), and Australian Bureau of Statistics (2005).

Timing of interviews

Since respondents in LSAC are asked about their current or present incomes, timing of interviews is also an issue in the comparison of the surveys. In all three surveys, information on current income (that is, around the time of interview) is compared. Figure 2 shows that, in SIH, the time period within which most interviews took place is generally earlier than the LSAC interview period. For HILDA, the time period for Wave 3 interviews is earlier, while the time period for Wave 4 interviews is later than that for LSAC. All other things being equal, these differences would suggest lower incomes in HILDA Wave 3 data (covering the financial year 2002–03) than in SIH or LSAC, but similar timing for previous financial year income in HILDA Wave 4 data and current income in SIH (covering the financial year 2003–04), with data on current incomes in LSAC covering a slightly later period. We compare incomes in these three surveys in Section 4.

Figure 2: Range of dates when most interviews were carried out in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample, HILDA Wave 3, age 3 to 4 sub-sample and HILDA Wave 4, age 4 to 5 years sub-sample



Note: In LSAC and HILDA, the shaded areas cover the time periods when the majority of interviews (80–95 per cent) were carried out. In SIH, the shaded area covers the period when all interviews were carried out.

Source: Watson (2010), Australian Institute of Family Studies (2010), and Australian Bureau of Statistics (2005).

3 Literature Review

What should we expect from the different methods of data collection in the three surveys? Firstly, we explore the literature that reports on how respondents interpret survey questions about income. Then we investigate literature that statistically examines the validity of income data in a range of surveys.

3.1 Why is collecting information on income difficult?

As Moore et al. (2000) show in their analysis, the task of eliciting income information from survey respondents is complex:

Consider the magnitude of the survey designer's task in designing a survey to collect income data. The most difficult step of this process may, in fact, be the first: defining the construct for survey respondents in clear, simple, and easily understood language. The fact that income has so many varied components partially explains the complexity of the task. Not only are there many different forms of remuneration (e.g., pay-in-kind, non-wage cash payments, cash and noncash program benefits, take-home pay and gross income, fringe benefits, assets, etc.) that may or may not be included within the definition, but there are also varying recipients (e.g., households, families, couples, individuals) and receipt periods (e.g., weekly, bimonthly, monthly, quarterly, annual) that must be defined and conveyed. (Moore et al., 2000, p.349).

Collins and White (1996) put the problem another way—more from the respondent's perspective:

It is important to understand the way in which answers to questions on income are produced by respondents as this helps us identify problems with question wording, presentation, and acceptability. Thus...we can see that a number of cognitive processes are in evidence. Firstly, the respondent has to interpret the question, specifically, what is meant by gross income. Secondly, he or she must retrieve the information from memory, thirdly, make a judgement about the information, and finally, find the appropriate answer category to tick. Within this process of answer generation there are other processes. If respondents are paid at different intervals, such as monthly, to the intervals presented in the questions (weekly and annual amounts) they will have to convert their

answers to the appropriate interval to enable them to endorse the correct category. Furthermore, for those who receive more than one source of income, the calculation of the amount becomes even more complex, as for each source, the amount has to be calculated and converted into the appropriate time interval before a total figure can be produced. (Collins and White, 1996, p.3)

Davern et al. (2005, p.1535) raise a further issue: ‘that people do not like to divulge how much money they earn.’ The purpose of these perspectives in highlighting the difficulties faced by survey designers and respondents in devising and interpreting questions about income is to alert data users to flow-on problems in interpreting the answers that they give. Moore et al. (2000) point out that technical terms such as ‘non-wage income’ are not well understood by respondents. Collins and White (1996) argue that ‘gross income’ seems to be more widely understood than ‘net income’. Both papers suggest that people tend to exclude certain items from their calculation of total personal or family income; for example, small amounts of earnings (even if regular) or income in kind. It seems that people rarely consult payslips or bank statements in reporting their incomes at interview, relying rather on memory and mental calculations. Therefore, the rationale behind detailed questioning on income, as occurs with SIH and HILDA respondents, appears to be that people are less likely to forget individual income sources and more likely to report amounts that approach their total ‘real’ income. This, presumably, is also the rationale behind asking LSAC respondents, particularly at Wave 1, about their income sources before asking them about their total income in dollars from these sources (as discussed in Section 2). If this is the case, it raises the question of whether reducing the number of categories in the ‘income sources’ question at Waves 2 and 3 of LSAC may have affected respondents’ estimation of their total income in dollar terms. Certainly, the literature suggests that prompting does make a difference (Davern et al., 2005; Moore et al., 2000).

3.2 Validating income in surveys

Moore et al. (2000) state that surveys where detailed questions are asked can achieve better results, in comparison with aggregates from other sources, than surveys where fewer questions are asked. The former also have what Davern et al. (2005, p.1537) call ‘more face validity’; that is, they look like they are measuring what they are supposed to be measuring. Moore et al. (2000) note, however, that income appears to be under reported in *all* surveys. They put this down to a number of factors, including individual item non-response and underestimation of income from some sources. This is indeed what Siminski et al. (2003) find in their analysis of the consistency of household income data in surveys carried out by the ABS. In all Household

Expenditure Surveys and Surveys of Incomes and Housing Costs carried out in 1981–82 and 1997–98, they find that aggregate income never totals more than four-fifths of that estimated in the Australian System of National Accounts, and sometimes a good deal less. While they put this down partly to differences in income concepts between the two sources, they also suggest some other differences. Specifically, they argue that estimates of current weekly income fall particularly short of national accounts aggregates because respondents are asked about their ‘usual’ weekly income, which in principle excludes non-regular payments. These payments might, however, be reported in annual income. An important conclusion to be drawn from these studies is that there is no ‘gold standard’ in income measurement—all measures are, to some extent, problematic. However, the more questions that are asked, the better the information appears to be.

Micklewright and Schnepf (2010) address this assumption directly in their analysis of the reliability of income data in the UK, comparing data collected with a single question with income measured using a large number of questions. They look at a single-household income question in the British Social Attitudes Survey (BSA) and a single individual income question in the Office for National Statistics Omnibus Survey. They appear to take the view that large government surveys such as SIH in Australia or the Family Resources Survey in the UK, while imperfect in a number of respects (including in terms of how they measure income), are nonetheless as close as researchers are currently likely to get to a ‘gold standard’ (or they have ‘face validity’, as Davern et al., 2005 p.1537 put it). They therefore compare income data from these surveys with income data in other surveys, where collecting data on incomes is not the main purpose, multiple questions about people’s incomes are not asked, information is only collected from a single respondent, and missing income data is not imputed. This is directly relevant to the purpose of the present analysis.

Micklewright and Schnepf’s analysis considers item non-response and compares the distributions of both household and individual income from single questions (corresponding to the BSA and Omnibus surveys respectively). They describe non-response on income as ‘low’—14 per cent for the single household income question—and report a figure of 9 per cent for the single individual income question. Using logistic regression, they show that item non-response to income questions in both surveys is not random. Women are found to be less likely than men to respond to questions on household income. Non-response to household income is also more likely in larger households. Younger respondents (under 30) are more likely to respond to questions on individual income but less likely to respond to questions on household income. Older respondents, on the other hand, show a different pattern: they are less likely to respond to any questions on income. Nonetheless, the authors

caution that their results should not be over-interpreted, since their logistic regression models only explain a small number of total response outcomes in the two surveys.

Micklewright and Schnepf then go on to compare valid responses to the single household and individual income questions with measures of income from two surveys which ask multiple questions on income: the Family Resources Survey (FRS) and the Expenditure and Food Survey (EFS). We shall concentrate on their comparisons with the FRS data. They find that men's individual income measured using a single question (Omnibus survey) is very similar to estimates from surveys containing a large number of income questions. Similarities between the surveys are strongest among active, working age men, suggesting that single questions perform better when income is derived from employment. The measure of women's individual income, on the other hand, is significantly lower for the single question than for multiple questions. The difference between surveys tends, however, to be smaller for women without children.

In Micklewright and Schnepf's comparison with the FRS, household income reported in a single question (BSA) tends to be lower, especially where the woman is a respondent. Even when the BSA data are reweighted so that the proportion in employment is similar to that in the FRS, large gaps remain in the measure of household income from a single question compared with multiple questions. While reports of household income from a single question in lone person households are found to match those in the FRS quite well, reports with respect to households with multiple adults are found to be more problematic. Micklewright and Schnepf explain this in terms of the difficulty of one person's knowing the income of all adults in a multi-adult household. Therefore, differences in the measure of income using a single question versus multiple questions are largest for respondents, and especially women respondents, in multi-adult households.

Broadly, Micklewright and Schnepf's findings with respect to women, and mothers in particular, suggest that there may be a problem with income in LSAC, considering that it targets a population of children and that the primary respondent, in the vast majority of cases, is the mother. However, as pointed out in Section 2, LSAC does not ask a single income question but asks a number of questions, and this may help to improve the measure of income. Furthermore, the majority (if not all) of LSAC parents are of working age, so problems with older men's income highlighted by Micklewright and Schnepf are not likely to be an issue for LSAC income.

4 Analysis

4.1 Analysis plan

In our analysis, as stated earlier, we closely follow Micklewright and Schnepf (2010). We begin by describing the characteristics of LSAC, SIH and HILDA samples, in Subsection 4.2. We then conduct a descriptive and multivariate analysis of item non-response to the individual and household income questions in LSAC, in Subsection 4.3. Following this, in Subsection 4.4, we compare measures of individual income in LSAC with measures in SIH and HILDA. Specifically, we compare men’s and women’s average individual income in all three surveys in total and across a range of socioeconomic and demographic characteristics. In Subsection 4.5, we examine differences across the entire distribution of men’s and women’s income and differences in means for income quartiles. In this Subsection, we also compare the distribution of combined income of men and women in households and compare the banded household income question in LSAC with corresponding measures in SIH and HILDA. Finally, in Subsection 4.6, we analyse dynamics of income in LSAC and HILDA.

Aside from the analysis of non-response in LSAC, which is carried out on unweighted data, we perform our comparisons on weighted data—that is, stratum weights (to match the sampling frame) in LSAC, and replicate weights in SIH and HILDA. Since our comparisons are mostly cross-sectional, we mostly use cross-sectional weights in both LSAC and HILDA. We conduct t-tests to assess whether differences between two estimates (x and y) from two independent samples are statistically significant. The formula is given here:

$$T = \frac{x - y}{se(x - y)}$$

Where:
$$se(x - y) = \sqrt{[se(x)]^2 + [se(y)]^2}$$

We use linearised standard errors (se) to calculate this test statistic. If the test statistic is greater than or equal to 1.96 (or $P < 0.05$), we conclude that the difference between estimates is statistically significant.

4.2 Comparing samples

In this Section, we describe basic demographic and socioeconomic characteristics of the LSAC, SIH and HILDA samples. Table 1 presents descriptive statistics on the composition of the samples with respect to key demographic and socioeconomic characteristics especially relevant to income. We start by comparing the LSAC

sample with two alternative sub-samples in SIH and two in HILDA. In SIH, the two sub-samples comprise families with a child aged 3 to 4 years at the time of interview and families with a child aged 3 to 9 years. In the HILDA Wave 4 sample, the two sub-samples are families with a child aged 4 to 5 years on 30 June before the interview, and families with a child aged 4 to 9 years. We chose to examine initially the characteristics of these two sub-samples in SIH and HILDA for sample size reasons. As the bottom rows on Table 1 show, while the sample size in LSAC is large (4,946 women and 4,320 men) the age 3 to 4 and age 4 to 5 sub-samples in SIH and HILDA are much smaller, with 661 men and 789 women in the former and 376 men and 436 women in the latter.

Table 1: Characteristics of families in LSAC K Cohort Wave 1, SIH 2003–04 families with children aged 3 to 4 years and aged 3 to 9 years, and HILDA Wave 4, families with children aged 4 to 5 years and aged 4 to 9 years

	<u>LSAC</u> 4-5 yrs sample	<u>SIH 2003–04</u> 3-4 yrs sample 3-9 yrs sample		<u>HILDA Wave 4</u> 4-5 yrs sample 4-9 yrs sample	
Family type					
Couple family	85.0	90.6*	86.9	89.9*	87.2
Lone-parent family	15.0	9.4*	13.1	10.1*	12.8
Age men					
Age < 30	7.8	11.8*	6.3	7.9	5.3*
Age 30 - 39 years	58.6	61.0	47.8*	53.7	43.4*
Age 40 + years	33.6	27.2*	45.9*	38.4	51.3*
Age women					
Age < 30	17.2	25.1*	15.8	19.4	13.7*
Age 30 - 39 years	64.6	62.1	55.8*	61.8	55.2*
Age 40 + years	18.2	12.8*	28.4*	18.9	31.1*
Proportion in Employment					
Men	92.3	90.9	88.6*	88.2	89.1
Women	55.2	51.5	54.2	50.7	57.1
Proportion with Degree					
Men	26.4	23.7	22.2*	26.5	23.3
Women	24.3	22.6	20.6*	21.8	20.8*
Proportion managers or professionals¹					
Men	33.2	31.5	32.4	36.7	33.0
Women	31.5	28.3	30.9	36.1	32.7
Sample N					
Men	4,320	661	948	376	797
Women	4,946	789	1,226	436	1,051

Employed N

Men	4,016	603	845	342	725
Women	2,833	407	712	270	644

* P < 0 .05; ¹ Base includes only those in employment. Weights applied.

Compared with LSAC, the SIH age 3 to 4 years sample has a significantly lower proportion of lone-parent families (and a significantly higher proportion of two-parent families). This is the case also for the HILDA age 4 to 5 years sample. However, there is no significant difference in family type between LSAC and the broader SIH and HILDA samples (age 3 to 9 and 4 to 9 years respectively). It is not unexpected that the share of children living in lone-parent families might be larger in a sample that includes older children in comparison with a sample that only includes 3 to 4 or 4 to 5 year olds. But this does not explain why the proportion of children in lone-parent families is higher in LSAC than in the more restricted SIH and HILDA sub-samples. This may be because LSAC is the only survey of the three that solely targets families with young children.

Mothers and fathers in the SIH age 3 to 4 years sub-sample are younger than parents in the LSAC sample. This may simply be an issue of timing. As Figure 2 shows, SIH respondents were for the most part interviewed earlier than LSAC respondents. The opposite is the case if we compare ages of parents in LSAC with ages of parents in the SIH age 3 to 9 years sub-sample: parents in the latter are significantly older than LSAC parents. For example, 28 per cent of mothers in the SIH age 3 to 9 years sub-sample are aged 40 years or over, compared with 18 per cent in the LSAC sample. The age profile of parents in the HILDA age 4 to 5 years sub-sample is not significantly different to that of parents in LSAC, but parents in the HILDA age 4 to 9 years sub-sample are again significantly older than LSAC parents. For example, just over half of fathers in the HILDA age 4 to 9 years sub-sample are 40 years or over compared with a third in the LSAC sample.

LSAC families are similar to both the SIH age 3 to 4 years and the HILDA age 4 to 5 years sub-samples with respect to the proportion of mothers and fathers in employment, with a degree or in a managerial or professional occupation. A significantly lower proportion of fathers in the SIH age 3 to 9 years sub-sample is employed (89 per cent, compared with 92 per cent in LSAC). Relatively large differences are also apparent between the proportions of men employed in both HILDA sub-samples (age 4 to 5 years and 4 to 9 years) compared with LSAC fathers, but these are not statistically significant. Comparatively small sample sizes may be a factor here. Finally, a significantly lower proportion of mothers and fathers in the SIH age 3 to 9 years sub-sample possess a degree or higher qualification than LSAC

parents. This is the case also for mothers in the HILDA age 4 to 9 years sub-sample.

As noted above, with respect to HILDA, one option would have been to draw a sample of families with a child 3 to 4 years from HILDA Wave 3. However, as Appendix Table A1 shows, the HILDA Wave 4 age 4 to 5 years sub-sample compares better with LSAC on demographic characteristics than the HILDA Wave 3 age 3 to 4 years sub-sample. Also notable are differences between the LSAC sample and the larger SIH (age 3 to 9) and HILDA (age 4 to 9) sub-samples. In particular, both men’s and women’s ages are significantly greater in the larger samples than in the smaller samples or in LSAC. Proportions of women in employment are also significantly larger. Since both age and employment are related to income, this suggests that comparisons between LSAC and the larger SIH and HILDA sub-samples may be problematic. Therefore, in the remainder of the analysis, we focus on comparing LSAC with the SIH age 3 to 4 years subsample and the HILDA Wave 4 age 4 to 5 years sub-sample. Figure 3 and Figure 4 chart the age distribution of parents in LSAC, the SIH age 3 to 4 years sub-sample and the HILDA age 4 to 5 years sub-sample for fathers and mothers respectively.

Figure 3: Age distribution of fathers in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample

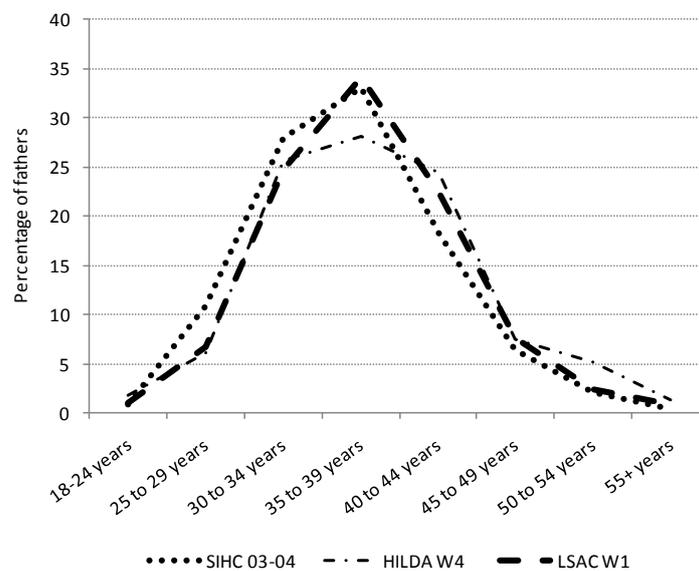
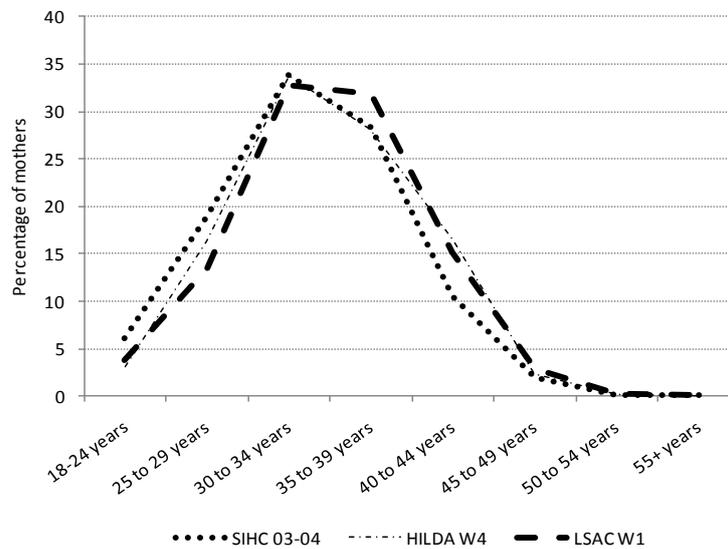


Figure 4: Age distribution of mothers in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample



Recall that both parents were younger in the SIH age 3 to 4 years sub-sample than in LSAC. Figure 3 and Figure 4 confirm that a higher proportion of parents in the former are aged 25 to 29 years, and a lower proportion are aged 40 to 44 years. The differences are not large, however. For example, 15 per cent of LSAC mothers are aged 40 to 44 years, compared with 11 per cent of mothers in the SIH age 3 to 4 years sub-sample. Parents in the HILDA Wave 4 age 4 to 5 sub-sample within these age groups are quite similar to LSAC parents. However, a higher proportion of LSAC parents, especially fathers, are aged 35 to 39 years. The overall picture is one of comparative uniformity in the age distribution of parents across the three surveys, when the more restricted SIH and HILDA sub-samples are included in the comparison.

To summarise: while the sample sizes of the SIH age 3 to 4 years sub-sample and the HILDA Wave 4 age 4 to 5 years subsample are relatively small, these are more comparable, in terms of demographic characteristics of parents, with the LSAC sample than with alternative sub-samples drawn from SIH or HILDA. However, the share of lone parents in the total in SIH and HILDA is lower than that in LSAC. The age profile of mothers in SIH is also somewhat lower than that in LSAC.

4.3 LSAC income data: item non-response

Respondents to LSAC provide information about their individual income, their partner's income (where appropriate) and banded income unit income (defined as the sum of both partners' incomes). At Wave 1, respondents are asked to give their own and their partner's incomes to the nearest dollar and to select from a list of 15 specified income bands for the income unit income. At Waves 2 and 3, they are not asked to give a banded estimate income unit income; that is, they are only asked about individual incomes (again to the nearest dollar).

At Wave 1, therefore, there are three points where item non-response can arise with respect to income data. Two points are worth noting in this respect. Firstly, data on individual and banded income unit incomes are not systematically related; information about individual income is not used at the coding stage to impute banded income unit income or vice versa. Therefore, respondents can provide dollar amounts for their own and their partner's incomes but provide no information on the band into which income unit income falls, or vice versa. Secondly, there may be item non-response for own income, partner's income, or both incomes. Our analysis explicitly addresses these unique features of individual and income unit income data in LSAC.

We begin this Section with an examination of item non-response to the banded income unit income question. Then we examine item non-response to individual income questions. We look at item non-response for own income and partner's income and at combinations of item non-response incorporating own income, partner's income, or both.

Income item non-response: a descriptive overview

Table 2 reports the item non-response rate for the banded income unit and individual income questions in LSAC for all families and for couple families only. Note that 'all families' includes lone-parent families, for whom the individual income question for Parent 2 is not applicable.

Looking at the income questions applicable to all families, item non-response for the banded income unit income question is relatively low at 6.4 per cent. This compares favourably with, for example, a figure of 14.1 per cent for a similar question in the BSA, a figure which is seen as relatively low by international comparisons (Micklewright & Schnepf 2010). Item non-response for the respondent's individual income (Parent 1) is 11.2 per cent. This is higher than a figure of 8.5 per cent in the Omnibus survey, reported by Micklewright and Schnepf (2010). One possible reason for this is that, in LSAC, respondents are asked to provide a unique dollar amount, whereas in the Omnibus survey, respondents are asked to choose from 39 income

bands. The lower item non-response rate for the banded income unit income question in LSAC adds some weight to the proposition that respondents may be more willing or able to report their income in bands rather than in exact dollars. It is important to note, however, that other differences in the design of these surveys could also result in differences in item response rates.

Table 2 also shows item non-response for questions applicable to both parents in couple families. In particular, we examine the rates of non-response for individual income for Parent 1, Parent 2, or some combination of these outcomes. Non-response for respondents' partners' income is higher than for respondents' own income or for banded income unit income. A total of 15.4 per cent of respondents in couple families did not provide information about their partners' income, including 8.7 per cent where there was non-response with respect to both the respondent and her partner, and 6.7 per cent where there was non-response with respect to the partner's income only. Item non-response for respondents' own income in couple families is 12.2 per cent of couple families (8.7 plus 3.5). This is slightly larger than the figure reported across all family types (11.2 per cent), suggesting that item non-response to the individual income question for Parent 1 is relatively low for lone parents.

Table 2: Item non-response rates for the banded income unit and the individual income questions in LSAC K Cohort Wave 1

	Item non-response rate (%)
All families (n=4,983)	
Banded income unit question	6.4
Parent 1 individual income missing	11.2
Couple families only (n=4,283)	
Parent 1 individual income missing only	3.5
Parent 2 individual income missing	15.4
<i>Of which:</i>	
Parent 1 and 2 individual incomes missing	8.7
Parent 2 individual income missing only	6.7

Note: unweighted

Multivariate analysis of income item non-response

Item non-response to income questions is not random (Micklewright & Schnepf 2010). We conduct multivariate regression analysis to examine the determinants of item non-response to income questions in LSAC. Our models broadly follow those set out by Micklewright and Schnepf (2010), but we adapt these to suit the features specific to LSAC. We examine four models of item non-response corresponding to the three income questions in LSAC:

1. Banded income unit income
2. Parent 1 individual income
3. Parent 2 individual income
4. Parent 1 and/or Parent 2 individual income

We examine each of these in turn below, using logistic regression, a common technique in the social sciences for investigating the association between several explanatory variables and a single binary dependent variable. With logistic regression, the dependent variable is transformed into a *logit*—the natural log of the odds ratio of the event in question occurring. The odds ratio is defined as

$$\frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}$$

where p_1 is the probability of the event occurring in population 1, and p_2 is the probability of the event occurring in a separate population 2. In this case, the dependent variable is the binary response/non-response indicator. Associations between explanatory variables and the dependent variable are expressed in the form of odds ratios, where an odds ratio of one suggests that there is no association between an explanatory variable and the dependent variable. An odds ratio of greater than one suggests that the explanatory variable is associated with an increase in the dependent variable, while an odds ratio of less than one suggests the opposite. Explanatory variables to predict the logit representing response on banded income unit income include the following dummies: Parent 1 (the respondent) is male (more than 95 of every 100 are female); Parent 1 is a lone parent; either Parent 1 or Parent 2 is self-employed; and either Parent 1 or Parent 2 has a degree. Explanatory variables also include the following continuous variables: Parent 1's age; and number of children in the family.

Odds ratios for predicting response on banded income unit income (and associated 95 per cent confidence intervals) are reported in Table 3. Where the confidence interval does not contain 1, the odds ratio is statistically significant. The only significant factor in the model for item non-response to the LSAC banded income unit income question relates to the dummy variable indicating families where at least one parent is self-employed. Respondents in these families are around 2.3 times more likely not to provide a response to the banded income unit income question in LSAC.

Table 3: Odds ratios for determinants of item non-response to banded income unit income question in LSAC K Cohort Wave 1

Independent variables	Odds Ratio	95 % confidence interval	
		Low	High
Parent 1 is male	0.949	0.408	2.206
Parent 1's age	0.996	0.971	1.022
Family contains self-employed person	2.322	1.805	2.986
Family is headed by lone parent	0.697	0.215	2.257
Family contains someone with a degree	0.952	0.733	1.237
Number of children in the family	1.100	0.977	1.239

Notes: An odds ratio < 1 indicates a negative association, and > 1 indicates a positive association. Results are significant at the 5 per cent level when the confidence interval does not contain 1.

Pseudo $R^2 = 0.02$

Table 4 shows the predictors of non-response to questions on the individual income of Parent 1. In this model, different explanatory variables are included to those used to model response to banded income unit income. As in the first model, variables for the gender and age of the respondent are included, as are the number of children in the family and whether the respondent is a lone parent. The model also includes four employment indicators (Parent 1 is self-employed, Parent 2 is self-employed, Parent 1 is not employed, Parent 2 is not employed); and education indicators (Parent 1 and/or Parent 2 has not completed Year 12 education, Parent 1 has a degree and Parent 2 has Year 12 education or less, Parent 2 has a degree and Parent 1 has Year 12 education or less, and both Parent 1 and 2 have degrees).

Table 4: Odds ratios for determinants of item non-response to Parent 1 individual income question in LSAC K Cohort Wave 1

Independent variables	Odds Ratio	95 % confidence interval	
		Low	High
Parent 1 is male	0.951	0.556	1.625
Parent 1 age	1.036	1.019	1.053
Parent 1 is self-employed	2.712	2.130	3.453
Parent 1 is not employed	1.469	1.189	1.815
Parent 1 is a lone parent	0.595	0.431	0.822
Parent 1 has a degree	1.058	0.859	1.303
Parent 1 has not completed Year 12	0.907	0.710	1.159
Number of children in the family	0.956	0.875	1.046

Notes: An odds ratio < 1 indicates a negative association, and > 1 indicates a positive association. Results are significant at the 5 per cent level when the confidence interval does not contain 1. Pseudo $R^2 = 0.03$

The Table shows that item non-response to the Parent 1 individual income question increases with Parent 1's age and is significantly higher among Parent 1s who are not employed or are self-employed compared with Parent 1s who are employed. As is the case with banded income unit income, self-employed Parent 1s are 2.7 times more likely not to respond to the individual income question, while non-employed Parent 1s are approximately 1.5 times more likely not to respond to this question. In contrast, lone parents (mostly mothers) are more likely to respond to the individual income question. As with banded income unit income, education and number of children have no impact on item non-response for Parent 1 income.

The model for non-response to the individual income question for Parent 2 is similar to that for Parent 1. Again, variables for the gender and age of Parent 1 are included in this model, as is the number of children in the family. The model also includes four employment indicators (Parent 1 is self-employed, Parent 2 is self-employed, Parent 1 is not employed, Parent 2 is not employed); and four education indicators (Parent 1 and/or Parent 2 has not completed Year 12 education and neither has a degree; Parent 1 has a degree and Parent 2 has no Year 12 education; Parent 2 has a degree and Parent 1 has no Year 12 education; and both Parent 1 and Parent 2 have degrees). The reference group for the education variable is households where both parents have Year 12 or other post-secondary qualifications below degree level. This model is estimated on a sample of couple families only.

Results are reported in Table 5. As with the model for Parent 1 individual income, non-response on Parent 2 income increases with the age of Parent 1. Moreover, non-

employed and self-employed Parent 1s are both less likely to provide information about the individual income of Parent 2 relative to Parent 1s who are working employees (about 1.3 times more likely in each case). If Parent 2 is self-employed, then Parent 1 is about 2.6 times more likely not to provide information about her partner's individual income. In contrast to the other models, Table 5 shows that education of parents is associated with non-response on Parent 2 income. However, results are a little counter-intuitive. In families where both parents have a degree, Parent 1 is more likely not to report the income of Parent 2, in comparison with families where both parents do not have degrees. Families where both parents have degrees are likely to have relatively high earnings, and perhaps the lack of an option to indicate income above a certain threshold (as opposed to giving an exact dollar amount to the interviewer) may have deterred some parents from responding on this question. It is notable that this effect is evident for the income of Parent 2 only. These are predominantly men, and most likely earning more than Parent 1. It may therefore be the case that sensitivity around extremely high incomes is impacting negatively on item response. It may, however, also be the case that these Parent 1s do not know their partner's income.

Table 5: Odds ratios for determinants of item non-response to Parent 2 individual income question in LSAC K Cohort Wave 1

Independent variables	Odds Ratio	95 % confidence interval	
		Low	High
Parent 1 is male	0.841	0.456	1.548
Parent 1's age	1.021	1.003	1.039
Parent 1 is self-employed	1.326	1.027	1.711
Parent 1 is not employed	1.339	1.091	1.642
Parent 2 is self-employed	2.618	2.141	3.201
Parent 2 is not employed	1.430	0.981	2.085
Parent 1 and/or Parent 2 has not completed Year 12 (neither has a degree)	0.827	0.645	1.061
Parent 1 has degree only (Parent 2 has no Year 12)	1.008	0.757	1.342
Parent 2 has degree only (Parent 1 has no Year 12)	1.158	0.865	1.551
Both Parent 1 and 2 have a degree	1.375	1.072	1.763
Number of children in the family	0.994	0.910	1.086

Notes: An odds ratio < 1 indicates a negative association, and > 1 indicates a positive association. Results are significant at the 5 per cent level when the confidence interval does not contain 1. Pseudo $R^2 = 0.04$

Individual income: Parent 1 and/or Parent 2

This final model is identical to the model for Parent 2 individual income, but the dependent variable now includes non-response on Parent 1 and/or Parent 2 individual

income. That is, information could be missing for Parent 1, for Parent 2, or for both. Results are shown in Table 6 and are substantively identical to those presented in Table 5, with one exception. Non-response on Parent 1 and/or Parent 2 income is significantly higher in families where only Parent 2 has a degree. This was not a significant factor in the model for item non-response to Parent 2 individual income. This suggests a somewhat peculiar finding: that when Parent 2 has a degree, Parent 1 is less likely to provide information about her own income.

Table 6: Odds ratios for determinants of item non-response to Parent 1 and/or Parent 2 individual income question in LSAC K Cohort Wave 1 (couple families only)

Independent variables	Odds Ratio	95 % confidence interval	
		Low	High
Parent 1 is male	0.880	0.507	1.527
Parent 1's age	1.021	1.005	1.038
Parent 1 is self-employed	1.646	1.297	2.091
Parent 1 is not employed	1.609	1.333	1.943
Parent 2 is self-employed	2.160	1.787	2.610
Parent 2 is not employed	1.289	0.911	1.825
Parent 1 and/or Parent 2 has not completed Year 12 (neither has a degree)	0.855	0.681	1.073
Parent 1 only has a degree (Parent 2 has no Year 12)	0.953	0.727	1.249
Parent 2 only has a degree (Parent 1 has no Year 12)	1.372	1.057	1.781
Both parents have a degree	1.374	1.090	1.732
Number of children in the family	0.964	0.888	1.046

Notes: An odds ratio < 1 indicates a negative association, and > 1 indicates a positive association. Results are significant at the 5 per cent level when the confidence interval does not contain 1. Pseudo $R^2 = 0.04$

This finding is confirmed by the results of a multinomial logit regression analysis of different combinations of item non-response to the individual income questions in two-parent families (i: income reported for both Parent 1 and Parent 2—this is the reference category; ii. income missing for both Parent 1 and Parent 2; iii. Parent 1 only missing; iv. Parent 2 only missing). The multinomial logit model specification is identical to the two previous models (Table 5 and Table 6). However, results relating only to parent education are shown in Table 7. Since odds ratios calculated from multinomial logit parameters do not lend themselves to easy interpretation, results are presented as Beta coefficients. These have much the same interpretation as Beta coefficients in Ordinary Least Squares regression models.

Table 7: Multinomial logit regression coefficients relating to Parent education for a model of item non-response to individual income questions in LSAC K Cohort Wave 1 (couple parents only)

Education variables	Both missing	Parent 1 missing	Parent 2 missing
Parent 1 and/or Parent 2 has not completed Year 12 (neither has a degree)	-0.02	-0.02	-0.41*
Parent 1 only has a degree (Parent 2 has no Year 12)	0.13	-0.40	-0.18
Parent 2 only has a degree (Parent 1 has no Year 12)	0.30	0.71**	0.04
Both parents have a degree	0.34*	0.24	0.32

* P < .05; ** P < .01; Pseudo R² = .04

The results again show that Parent 1 is significantly less likely to report her own income when Parent 2 has a degree. The results also reveal that, where both parents have a degree, Parent 1 is less likely to provide information about individual incomes, either for herself or for Parent 2.

To summarise: the analysis of non-response to income questions in LSAC shows consistently that, where parents are self-employed, the probability of non-response on their income is high. Non-response is also high with respect to parents with high levels of education. On the other hand, non-response is low for lone parents.

4.4 Comparing averages and distributions

We now move to comparing actual incomes in the three surveys. We begin with a description of average income for men and women across a range of key socioeconomic and demographic indicators. Table 8 reports the results for men's income, and Table 9 reports the results for women's income. We conduct t-tests to assess the statistical significance of differences between income in LSAC and SIH, and LSAC and HILDA.³ Unweighted sample sizes for the categories in Table 8 are reported in Appendix Table A3, and linearised standard errors associated with the weighted mean incomes are reported in Appendix Table A4. In absolute terms, the difference in average incomes between the LSAC and SIH samples is \$50 (or 5 per cent of LSAC mean), but this is statistically not significant. In general, men's income in LSAC is similar to men's income in the SIH sample (no statistically significant differences), although incomes in SIH are higher in every category except self-employment. The largest difference between men in LSAC and SIH is for those aged 40 years and over. For this group of men, there is a statistically significant difference of \$193 or 18 per cent in average weekly income (P < 0.1).

Table 8: Men's mean weekly income in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample across select socioeconomic and demographic characteristics (\$)

	LSAC A	SIH B	B/A	HILDA C	C/A
Current weekly income	991	1,042	1.05	1,214*	1.22
Age					
Age < 30	718	735	1.02	872	1.21
Age 30–39 years	991	994	1.00	1,200*	1.21
Age 40 + years	1,061	1,254#	1.18	1,304	1.23
Employment					
In Employment	1,052	1,116	1.06	1,316*	1.25
Not in Employment	214	226	1.06	453*	2.12
Education					
No Year 12 (No other qualifications)	702	756	1.08	811	1.16
Year 12 (No degree)	899	952	1.06	1,078*	1.20
Degree	1,400	1,468	1.05	1,837*	1.31
Occupation†					
Manager/Professional	1,397	1,481	1.06	1,660*	1.19
Other	887	945	1.07	1,116*	1.26
Employed†					
Self-employed	972	964	0.99	1,162	1.20
Employee/other	1,076	1,144	1.06	1,374*	1.28

† Employed persons only; * P < .05; # P < 0.1

Note: Weights applied;

In contrast, there are much larger differences when we compare men's income in LSAC and HILDA. Overall, we find a difference of \$223 (significant at P < 0.05) or 22 per cent between LSAC and HILDA incomes. This scale of difference is fairly consistent across different age groups and employment and education categories. In some cases, the differences between LSAC and HILDA means are not significant. However, this is probably because of very small sample sizes in some of HILDA categories (see Appendix Table A3). For example, in the case of men aged less than 30 years, the HILDA mean is calculated from a sample of 32 observations. The difference between men's income in LSAC and HILDA is higher among employees and among men with a degree. This suggests that men's incomes in these surveys become less comparable at the upper end of the income distribution. We examine distributions in more detail below.

Table 9 shows that differences in means for women's income are generally smaller than those for men across the three surveys. Overall, there is a difference of \$5 (or 1 per cent) between women's income in LSAC and SIH, and a difference of \$25 (or 6 per cent) between women's income in LSAC and HILDA. In both cases, the dollar amount is less in LSAC than in the other surveys and, in both cases, the difference is not statistically significant.

Table 9: Women's mean weekly income in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample across select socioeconomic and demographic characteristics (\$)

	LSAC	SIH		HILDA	
	A	B	B/A	C	C/A
Current weekly income	420	425	1.01	446	1.06
Family type					
Couple family	402	392	0.98	407	1.01
Lone-parent family	524	603#	1.15	628*	1.20
Age					
Age < 30	377	400	1.06	466#	1.24
Age 30–39 years	418	425	1.02	432	1.03
Age 40 + years	474	475	1.00	472	1.00
Employment					
In Employment	560	625*	1.12	578	1.03
Not in Employment	249	212*	0.85	310*	1.24
Education					
No Year 12 (No other qualifications)	358	367	1.03	417	1.17
Year 12 (No degree)	383	405	1.06	398	1.04
Degree	569	537	0.94	587	1.03
Occupation†					
Manager/Professional	748	805	1.08	733	0.98
Other	472	550*	1.16	491	1.04
Employed†					
Self-employed	495	531	1.07	601	1.21
Employee/other	577	640*	1.11	573	0.99

† Employed persons only; * P < 0.05; # P < 0.1. Weights applied.

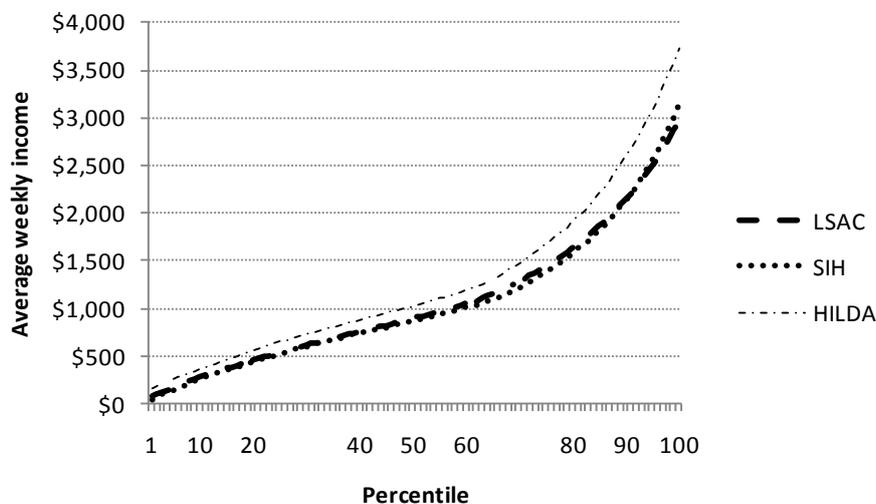
Among partnered women, differences between the three surveys are small and statistically not significant. Among lone parents, however, the differences between mean income in SIH and HILDA on the one hand, and LSAC on the other, are considerably larger (\$79 and \$104, or 15 per cent and 20 per cent respectively).

When comparing women's income in LSAC and SIH, we observe larger differences among employees who are not in professional/managerial occupation groups. This suggests that differences may arise from variation in the way women record employment-related income in these surveys. In contrast, when looking at women's income in LSAC and HILDA, we observe larger differences among women with lower qualifications and women not in employment, which is consistent with the difference between lone mothers' income in LSAC and HILDA. This suggests that there may be important variation in the way in which women report government benefits in each of these surveys; or it may be the result of imputation of benefits in HILDA.

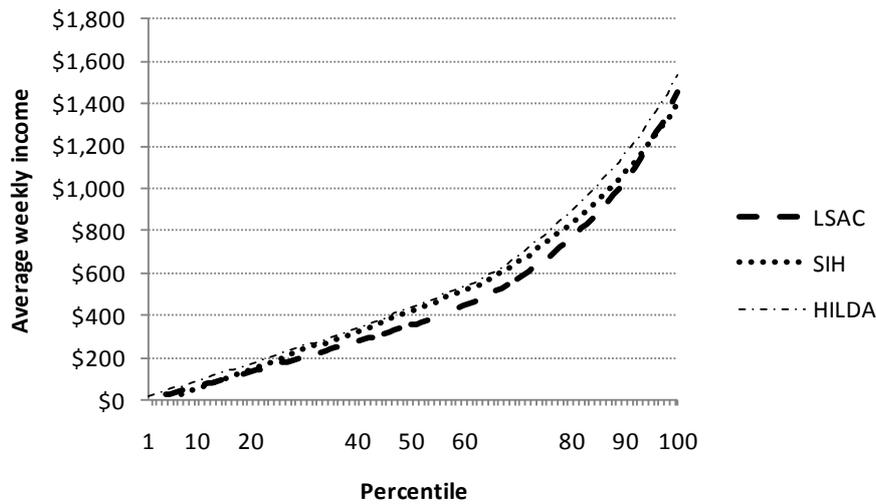
4.5 The distribution of men's and women's weekly income in LSAC, SIH and HILDA

In this Section, we examine income across the distribution of men's and women's incomes. To do this, we rank the men's and women's incomes (separately) into percentiles from 1 to 100. We then conduct non-parametric 'lowess' regression analyses for men and women separately, which yield a smoothed value of average income for each income percentile, which we then plot in Figure 5 for men and in Figure 6 for women.

Figure 5: Men's average weekly income across the income distribution



Source: LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample. Weights applied. Authors' calculations.

Figure 6: Women's average weekly income across the income distribution

Source: LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample. Weights applied. Authors' calculations.

Figure 5 shows that men's income in HILDA is higher than men's income in LSAC right across the income distribution, but the gap becomes larger as we move towards the upper portions of the income distribution. Men's incomes in LSAC and SIH are similar up to around the 95th percentile, where men's income in SIH begins to exceed men's income in LSAC. This result accords with the data presented in Table 8, which shows that men's incomes in LSAC and SIH were most different among higher socioeconomic groups. Recall, however, that differences in means were not statistically significant when comparing income in LSAC and SIH. As with men, women's income in HILDA is higher than women's income in LSAC across the income distribution. Women's income in SIH is somewhat higher than income in LSAC at the lower range of the income distribution. However, women's income in SIH moves closer to women's income in LSAC in the upper parts of the distribution, becoming very similar around the top decile.

We turn now to testing differences in mean income in different sections of the income distribution. Table 10 reports the mean income in the bottom, second, third and top quartiles of the income distribution for men and women in the three surveys. We conduct t-tests to compare differences in mean income in LSAC with SIH and HILDA for each quartile of the income distribution for men and women.

Table 10: Mean weekly income in the bottom, second, third and top quartiles of the income distribution for men and women in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample (\$)

	LSAC	SIH	HILDA
Men			
Bottom quartile	334	307	416*
Second quartile	730	721	851*
Third quartile	1,049	1,016*	1,183*
Top quartile	1,934	1,983	2,286*
Women			
Bottom quartile	85	77	113*
Second quartile	260	306*	299*
Third quartile	453	521*	544*
Top quartile	914	1,007*	1,052*

* P < 0.05; weights applied.

The results for men's income on Table 10 show that, within each of the four quartiles of the income distribution, the LSAC and SIH means are quite similar. Interestingly, men's income in the third quartile in LSAC is slightly greater than in SIH by the relatively small amount of \$33. Men's income in HILDA is consistently and significantly greater than LSAC across all four quartiles of the income distribution.

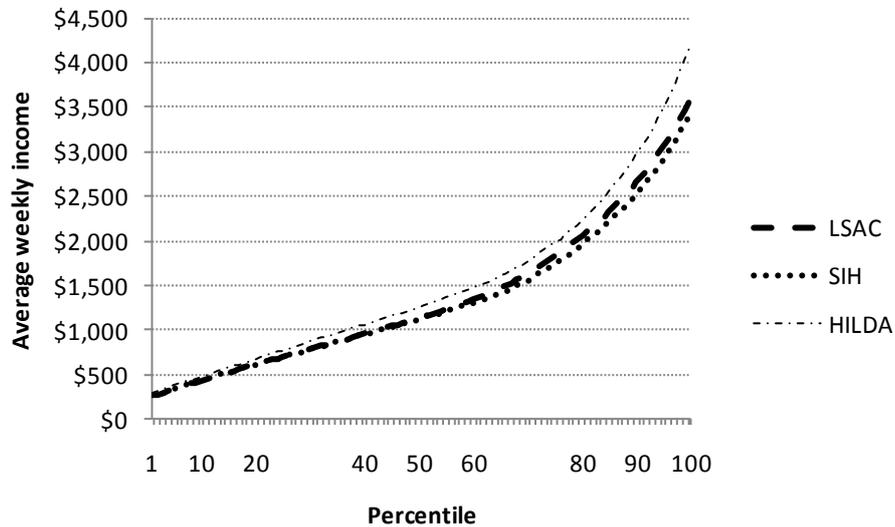
As indicated in Figure 6, the patterns for women's income are little different, particularly with respect to LSAC and SIH comparisons. In the bottom income quartile, women's income in LSAC is slightly greater than women's income in SIH but lower than women's income in HILDA. Differences are nominally small but, with the exception of the LSAC and SIH comparison in the bottom income quartile, they are statistically significant at the 5 per cent level. In the upper income quartiles, women's income in LSAC is significantly lower than women's average income in SIH and HILDA. However, LSAC–SIH and LSAC–HILDA differences, as a proportion of average income in LSAC, are largest in the third quartile.

Combining men's and women's weekly income

In this Section, we examine men's and women's combined incomes; that is, we construct a measure of income unit income based on the information from individual income questions for both parents. We repeat the smoothing procedure outlined above for men's and women's individual income. Smoothed values of average income unit income in LSAC, SIH and HILDA for each income percentile are shown in Figure 7. This measure of income is similar in LSAC and SIH (though LSAC is slightly higher

along the upper range of the income distribution). But the measure in HILDA is greater than that in LSAC, especially across the middle and upper ranges of the income distribution.

Figure 7: Average weekly income unit income across income percentiles



Source: LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample. Weights applied. Authors’ calculations.

Table 11 reports the mean household income in the bottom, second, third and top quartiles of the income distribution in LSAC, SIH and HILDA. Average household income amounts for each quartile in LSAC and SIH are similar. Differences between the second and third quartiles are statistically significant, even though the absolute amounts of difference are not large (\$18 and \$31, respectively). Average household income is greater in HILDA than in LSAC in all quartiles, and the difference is statistically significant in all quartiles except the top quartile. It is in the top quartile that the difference is most pronounced in absolute terms. However, this result is perhaps not very reliable because of large standard deviations and relatively small sample sizes.

Table 11: Mean household weekly income in the bottom, second, third and top quartiles of the income distribution in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample (\$)

	LSAC	SIH	HILDA

Bottom quartile	493	487	548*
Second quartile	920	938*	1043*
Third quartile	1,364	1,333*	1,492*
Top quartile	2,401	2,380	2,607

* $P < 0.05$; weights applied.

Grouped household income

The analysis above of income unit income suggests that family income calculated from individual parent incomes is similar in LSAC and SIH, but that it is greater in HILDA than in LSAC. As one might expect, this is consistent with results for men's and women's income examined separately. The major problem with this analysis, however, is that we must exclude cases where information about income is missing for either or both parents, which is approximately 20 per cent of households in LSAC.

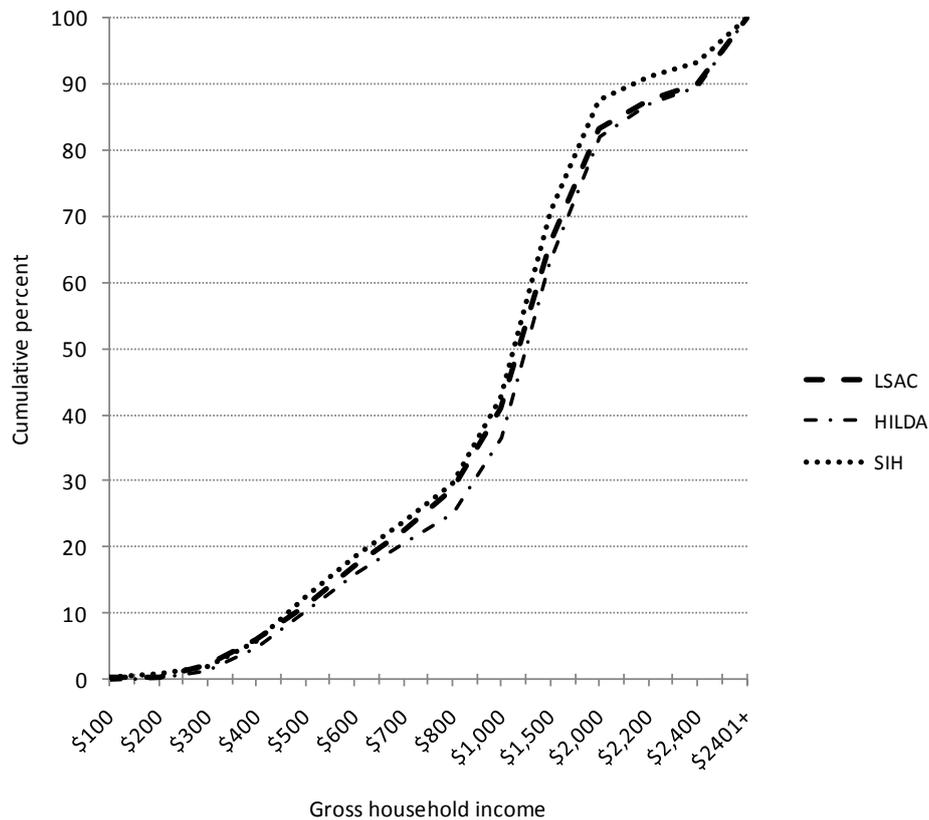
Recall that in LSAC, respondents at Wave 1 (but not at Waves 2 or 3) were asked to answer a question indicating the combined gross income of both parents in pre-defined income bands. Item non-response for this question was much lower (6.4 per cent). To compare these distributions, we grouped the measures of combined parental income in SIH and HILDA to correspond with the bands specified in the LSAC survey. The cumulative frequency distribution of banded household income in LSAC, SIH and HILDA is shown in Figure 8.

Figure 8 effectively shows the proportion of households below certain amounts of banded income. For example, about 40 per cent of households have income at or below \$1,000. As with the previous measure of parents' combined income, the proportion of LSAC and SIH households in each weekly income band is very similar. What is surprising, however, is that the proportion of LSAC and HILDA households is also very similar, which is in contrast to the findings set out above showing that estimates of income in HILDA are consistently higher than LSAC. This suggests that, where exact estimates of weekly income are compared across the three surveys, there is a large amount of error. However, if comparison of incomes is restricted to relatively broad bands, then estimates for the three surveys begin to look much more alike.

To summarise: in terms of overall means, and means for demographic subgroups for men, the LSAC sample compares well with the SIH sample but less well with the HILDA sample. For women, on the other hand, both SIH and HILDA compare well with LSAC, with one exception: in both SIH and HILDA, mean incomes of lone parents are significantly higher than those in LSAC. Comparisons of distributions of men's and women's incomes show growing differences between LSAC and HILDA towards the top of the distribution, but also notable differences between LSAC and

SIH for women in the middle of the distribution. Comparisons of banded income unit income in LSAC with similar bands in SIH and HILDA tend to obscure differences between the surveys and emphasise similarities in their distributions.

Figure 8: Cumulative distribution of banded income unit income in LSAC K Cohort Wave 1, SIH 2003–04 age 3 to 4 years sub-sample and HILDA Wave 4 age 4 to 5 years sub-sample



Note: Categories represent the upper end of each income band rounded up to the nearest dollar. Weights applied. Authors' calculations.

4.6 Income over time: LSAC and HILDA compared

In this final Section of the analysis, we examine weekly income measured in repeat waves of LSAC and HILDA. As noted above, there are differences in the income questions in Waves 2 and 3 of LSAC. A comprehensive analysis of the potential impact of these changes is beyond the scope of this report. Here, we assume that income data in each wave attempt to capture the same underlying concept. We therefore look at mean weekly income for men and women reported in the three waves of LSAC, and the five waves of HILDA, collected over the same period. Table 12 shows the sequencing of waves in both surveys and the corresponding age of the

study child in LSAC. Note that data are not available when the LSAC study child is aged 5 to 6 years and 7 to 8 years, which corresponds to HILDA Waves 5 and 7. For presentational purposes only, we conduct a straightforward linear imputation of income in LSAC for these points.

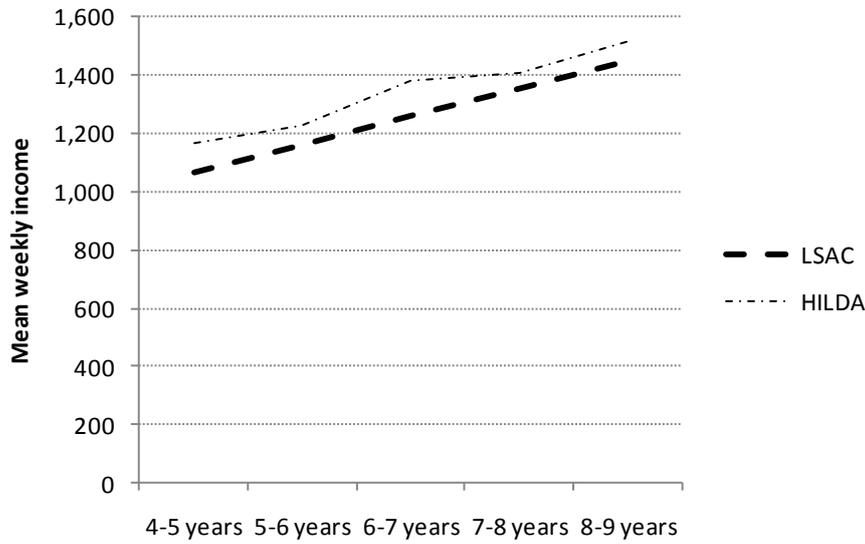
Table 12: Sequence of waves for LSAC and HILDA with the corresponding age of the LSAC study child

Age of LSAC study child	LSAC	HILDA
4 to 5 years	Wave 1	Wave 4
5 to 6 years	-	Wave 5
6 to 7 years	Wave 2	Wave 6
7 to 8 years	-	Wave 7
8 to 9 years	Wave 3	Wave 8

The analysis in this Section represents an important step in understanding more about the validity of LSAC income data across waves. We restrict our analysis to cases where information on individual income is available for all selected waves in each of the surveys. In LSAC, there are 2,688 observations for men and 3,471 observations for women (2,931 households). In HILDA, there are 624 observations for men and 827 observations for women (809 households). Therefore, although analyses are weighted to compensate for attrition, we are nonetheless likely to find that results are affected by longitudinal sample non-response and attrition effects.

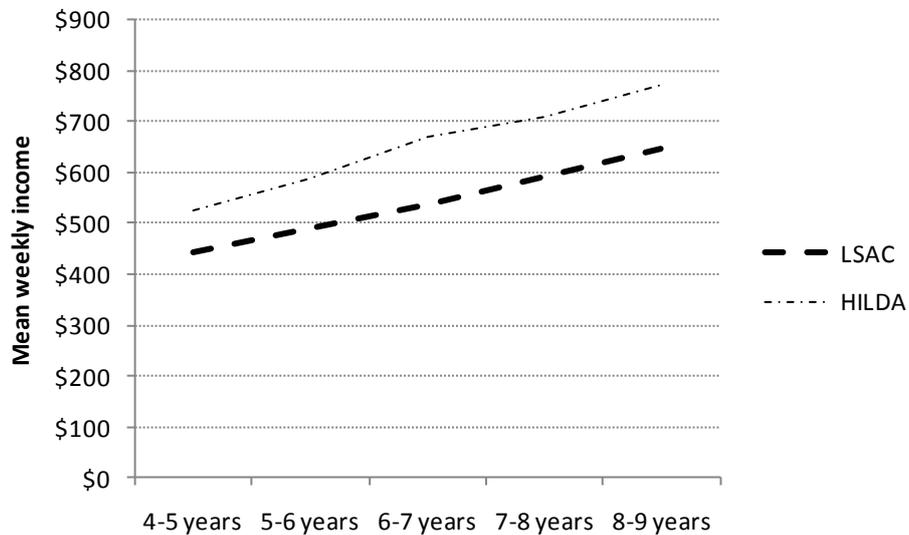
Mean weekly incomes for men and women across LSAC Waves 1 to 3 and HILDA Waves 4 to 8 are shown in Figure 9 and Figure 10 respectively. For men, the difference in average income between LSAC and HILDA at the first wave of data (4 to 5 years) is narrower than was reported in Section 4.5. In other words, the estimate of men's income in LSAC is greater when only those who respond and report income in all waves are included in the analysis. Generally, the picture for men shows that average income between the two surveys is converging over time. A slight increase in the gap is apparent at the second wave of LSAC and the corresponding sixth wave of HILDA (6 to 7 years), but the averages are closer at the last data point where LSAC study children are 8 to 9 years (LSAC Wave 3; HILDA Wave 8).

Figure 9: Men’s mean weekly income across LSAC Waves 1 to 3 and HILDA Waves 4 to 8



Note: LSAC income imputed for 5 to 6 years and 7 to 8 years (HILDA Waves 5 and 7 respectively). Weights applied.

Figure 10: Women’s mean weekly income across LSAC Waves 1 to 3 and HILDA Waves 4 to 8

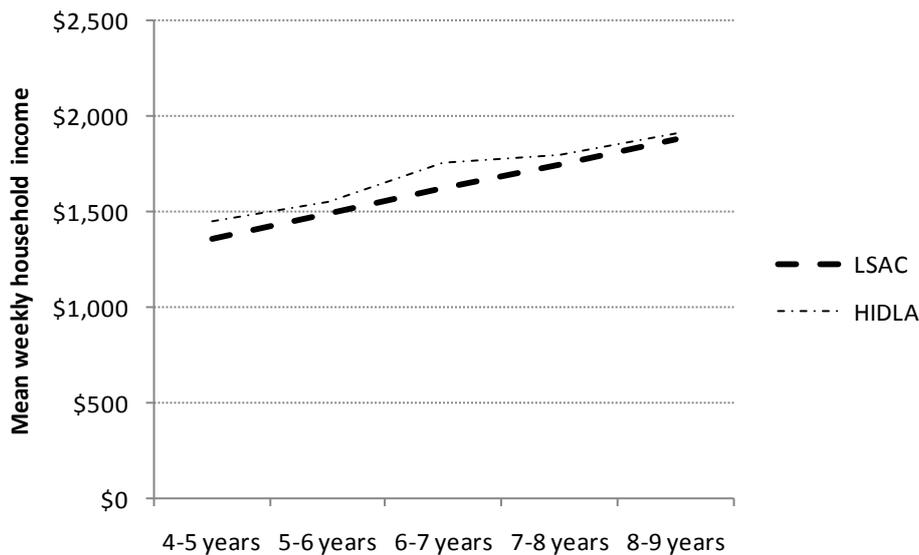


Note: LSAC income imputed for 5 to 6 years and 7 to 8 years (HILDA Waves 5 and 7 respectively). Weights applied.

In contrast to men, a large difference between women’s income in LSAC and HILDA is evident at the first wave of data (Figure 10). Recall that there was a small difference of \$25 reported above between women’s income in LSAC and HILDA, where the average was calculated for all women who reported income at Wave 1, whereas the difference is \$85 for women who responded and reported income over all waves considered here. Moreover, unlike men’s income, the difference in women’s income in LSAC and HILDA appears to be relatively constant across time.

Figure 11 shows income unit income across all waves of LSAC and HILDA Waves 4 to 8. Generally, across waves, this measure of household income in LSAC and HILDA is very similar and almost equal in the third wave of LSAC (HILDA Wave 8). The sample is now restricted to households where all parents respond and report income over all waves considered here. The implication is that income is more similar across surveys when looked at for households who fully respond to repeat waves of data collection.

Figure 11: Men’s and women’s combined mean weekly income across LSAC Waves 1 to 3 and HILDA Waves 4 to 8



Note: LSAC income imputed for 5 to 6 years and 7 to 8 years (HILDA Waves 5 and 7 respectively).

Incomes can change over time, and so too can the relative ranking of individuals within the income distribution. To look at this, we rank men’s, women’s and household income into quartiles and examine changes in the relative ranking of men, women and households in LSAC and HILDA. Firstly, we compare changes in rankings in LSAC between Waves 1 and 2 with changes in rankings in HILDA

between Waves 4 and 6; secondly, we compare changes in rankings in LSAC between Waves 2 and 3 with changes in rankings in HILDA between Waves 6 and 8. For each comparison, we place outcomes across waves for men, women and households into one of three groups: i) remained in the same quartile; ii) moved to a lower quartile; and iii) moved to a higher quartile. Results are reported as percentages in Table 13. Detailed cross-tabulations are presented in Appendix Tables A5 to A10.

While there are small differences, the general picture here is one of consistency in the movements of men, women and households in the income distribution across waves in LSAC and HILDA. In the first comparison, 60 per cent of men in LSAC remain in the same income quartile between Waves 1 and 2, compared with 64 per cent of men in HILDA. More men in LSAC move either into a lower quartile or a higher quartile than in HILDA. In comparison with HILDA, a slightly higher proportion of men, women, and men and women combined in LSAC move into a higher income quartile between both Waves 1 and 2 and Waves 2 and 3. We do not observe this consistency when looking at the proportion of men, women or households who move to a lower income quartile across waves. These patterns notwithstanding, nearly all differences are small.

Table 13: Changes in ranking within the income distribution for men, women and households in LSAC and HILDA (per cent)

	Earlier Waves		Later Waves	
	LSAC Wave 1 to 2	HILDA Wave 4 to 6	LSAC Wave 2 to 3	HILDA Wave 6 to 8
Men				
Same	60	64	63	65
Lower	18	17	17	18
Higher	21	19	20	16
All	100	100	100	100
Women				
Same	55	57	56	60
Lower	22	22	22	19
Higher	23	22	22	21
All	100	100	100	100
Men and women combined				
Same	62	65	63	64
Lower	16	17	16	19
Higher	22	18	21	17
All	100	100	100	100

To summarise: in order to compare income dynamics across the three waves of LSAC with five waves of HILDA, it is necessary to restrict analysis to those observations that respond at all waves. The resulting analysis is therefore affected by sample non-response and attrition. The analysis shows that men's incomes in the three waves of LSAC compare well with those of men in the corresponding five waves of HILDA. This is also the case with income unit income. Women's incomes, on the other hand, do not compare so well, with differences between the two surveys larger, on average, than those found in the Wave 1 analysis. However, these differences are consistent over time.

5 Discussion

The literature on survey methodology reviewed in Section 3 highlights the many steps that the survey respondent has to take when asked to report her personal or family income—what to count as one’s own, as opposed to somebody else’s, income; what to include or exclude; over what period to report it (Collins & White 1996; Moore et al. 2000). The detailed discussion in Section 2 on how income is recorded in LSAC gives an impression of the amount of work that the respondent has to do in reporting her current income in dollars, and this can be contrasted with the amount of guidance that respondents to SIH and HILDA are given. In LSAC, the respondent is simply asked whether she receives income from a number of sources and then asked what her total current income before tax is from all sources. In the other two surveys, respondents are not only asked if they have income from a long list of possible sources, but also asked separately, for each source, how much they receive and over what period they receive it. When these two methods are compared, it seems inevitable that LSAC respondents will under-report their incomes. We thread this hypothesis through the entire analysis and generally find that it holds, whether we look at men’s and women’s personal incomes or at household incomes, and whether we focus only on comparisons with LSAC Wave 1 data or on comparisons with all three waves of LSAC data. However, we discover a number of twists and turns along the way that can be of some assistance in considering how to evaluate the utility in research of incomes in LSAC, and we now discuss these.

Changes in LSAC income questions

Inconsistencies in the way respondents are asked about their incomes in the three waves of LSAC are perhaps indicative of the relatively low priority given to data collection on income in this survey. The changes from wave to wave are small but could potentially have a significant impact in terms of how respondents report their incomes. Potentially, the most significant changes happen between Waves 1 and 2. Here we highlight two.

Firstly, at Wave 1, the very first question on income asks respondents if they receive income from any of nine sources. They are then asked about receipt of a long list of government transfer payments. After this, respondents are asked how much, in dollars, they usually receive from all sources. At Waves 2 and 3, respondents are first asked about receiving income from just four sources before being asked about their total personal income in dollars. At all waves, similar questions are then asked of respondents about their partners’ incomes. As Moore et al. (2000) emphasise, prompting is an important mechanism used by survey designers to ensure better income reporting. Considerably more prompting occurs at Wave 1 than occurs at

Waves 2 or 3. This should have the effect of increasing under-reporting of income in the later waves in comparison with Wave 1. It is difficult to conclude from the information actually given in the surveys whether under-reporting does actually increase. Certainly, cross-wave comparisons of incomes in LSAC and HILDA do not generally suggest greater under-reporting in the latter LSAC waves. However, further research could be usefully carried out on this issue.

Secondly, Wave 1 respondents are asked to give the present income unit income (Parents 1 and 2 combined) in one of fifteen bands. This should be a relatively weak measure of income. However, as Table 2 shows, either the income of the respondent or of her partner is not reported in about one-fifth of LSAC families, while information for the banded income unit income question is only missing in the case of 6 per cent of families. Therefore, analyses that only use person-level income information effectively exclude about 20 per cent of the Wave 1 sample, while analyses that use information on banded income unit income exclude only 6 per cent. This difference suggests that the banded income unit income data deserve serious consideration, and that the effect of omitting this question from Waves 2 and 3 should be the subject of more analysis.

Non-response

In comparison with other studies that look at non-response to income questions in non-income-oriented surveys (see, for example, Micklewright & Schnepf 2010), it appears that non-response in LSAC is relatively low. However, researchers will need to examine the significance of non-response to income questions in the context of analyses that they propose to carry out. Our analysis of non-response to questions on personal and income unit income at Wave 1 of LSAC revealed, perhaps unsurprisingly, that non-response was particularly high where respondents reported income from self-employment for themselves or their partners. This finding is consistent with findings from more comprehensive income surveys, where information on self-employment incomes is generally seen as difficult to collect (Micklewright & Schnepf 2010; Moore et al. 2000). On the other hand, again perhaps unsurprisingly, respondents who were lone parents were most likely to report on income. (For this group, individual and income unit incomes should be the same.) This is consistent with Micklewright and Schnepf's (2010) finding that the reliability of income reporting decreases with the number of adults in the household. On the other hand, among couple families, respondents with partners who were not employed had higher than average non-response rates.

Differences in means and distributions

The analysis of non-response at Wave 1 in LSAC revealed some factors that may

point towards lower mean incomes in LSAC in comparison with SIH and HILDA. If it is assumed that the two latter surveys perfectly capture the incomes of self-employed people and people with high levels of education, then it might also be assumed that non-response by these groups in LSAC should give lower average incomes in comparison with these surveys. On the other hand, higher rates of non-response among respondents whose partners were not in employment might be expected to increase average incomes in LSAC relative to the two other surveys.

Other factors point towards an expectation of lower incomes in LSAC than in the other two surveys. Chief among these is the effort made to collect information on incomes in SIH and HILDA. Also important with respect to HILDA, in particular, is the timing of interviews at Wave 4, which for the most part occurred later than interviews in LSAC (Figure 2). (On the other hand, most SIH interviews took place before LSAC interviews.)

Indeed, differences between men's mean incomes in LSAC and in SIH are generally quite small—5 per cent overall, while differences between LSAC and HILDA Wave 4 are large—22 per cent overall (Table 8). Differences between the LSAC and SIH means are proportionally largest (18 per cent) in the case of respondents who are aged 40 years or more (although this difference is only statistically significant at 0.1). Differences between LSAC and HILDA are proportionally largest (112 per cent) in the case of respondents who are not in employment. However, these comprise a small group in both surveys. To summarise, it appears that LSAC best captures the incomes of men who are employees (92 per cent of men and 55 per cent of women; see Table 1).

Among women, differences in mean incomes between LSAC and SIH are very small indeed—a statistically insignificant 1 per cent overall, with the largest difference occurring for lone parents. Differences between LSAC and HILDA are somewhat larger overall—6 per cent—but still statistically not significant. However, mean differences among lone parents, mothers aged under 30 and mothers not in employment are all 20 per cent or higher.

The finding of large differences for lone parents between LSAC and the other two surveys is particularly interesting, given that non-response on income questions is low among lone parents. We believe that the reason for this difference lies in the approach taken in SIH and, we believe, even more so in HILDA, to impute the value of a wide range of government payments, in particular Family Tax Benefit. Such payments are likely to comprise a large share of total incomes for lone parents. However, total payments can be difficult for individuals to calculate, since they may receive several payments each fortnight, and these may be supplemented by lump sum payments (or

reduced by repayments of overpayments) at different points in the tax year. Therefore, imputed amounts for these payments may, on average, be higher than reported amounts.

In terms of distributions of men's incomes, LSAC and SIH distributions appear very similar. The distributions of LSAC and HILDA are less similar, especially at the top of the distribution. For women, differences between all three distributions are considerably smaller. However, differences between LSAC and the two other surveys are most evident at the middle of the distribution. In terms of income unit incomes calculated from reports of individual incomes, differences between LSAC and SIH are again small, but differences between LSAC and HILDA are notable at the top of the distribution. Yet when income unit income is reported in bands by LSAC respondents, it compares much more favourably with similarly banded income unit incomes in SIH and HILDA. This suggests an important finding—that once expectations of pinpoint accuracy are removed for LSAC respondents in the way they report their incomes, not only do response rates improve greatly, but comparability with other surveys also improves. Unfortunately, the banded income question was only asked at Wave 1 of LSAC, so we are unable to ascertain whether this trend would continue through other waves.

Differences across waves

However, our findings with respect to individual incomes of men and women in LSAC over the three waves are at the same time reassuring and worrying. Our focus in this part of the analysis is on respondents who were present in three waves of LSAC, in comparison with HILDA respondents who were present during five waves of that survey carried out over the same period. Our findings are very reassuring in that the differences between the two surveys for men and for income unit incomes are narrower than what we find in the Wave 1 analysis and do not increase over time. Less reassuring is the finding that differences between LSAC and HILDA for women's incomes are wider over the three waves than in the Wave 1 analysis. This suggests that sample attrition in LSAC and HILDA may be driving this increased similarity among men and increased dissimilarity among women. This result needs further analysis. In the meantime, we propose that analyses which use income dynamics in LSAC need to be carried out with caution and with due attention to the effects of attrition on sample representativeness.

6 Conclusion

This report has examined the comparability of measures of income in LSAC and in SIH and HILDA. The basic finding of the report is positive, in that measures of income in LSAC are broadly comparable with measures derived from both SIH and HILDA. Broadly, we find that both men's and women's average incomes in LSAC compare well with those in SIH, but somewhat less well with those in HILDA. In the case of women, it is especially notable that income for lone mothers in LSAC is lower than income for lone mothers in the other surveys. The comparisons give some confidence that income as reported in LSAC is a reasonable approximation of income among the population from which the LSAC K Cohort is drawn. However, levels of non-response on income questions are also significant, though very much in line with rates observed in other surveys. Researchers should examine the impact of this non-response on their analyses.

Some differences in incomes between the surveys are noteworthy. The finding relating to lone mothers is potentially of some concern and suggests an issue around how well LSAC is measuring income from benefits, or at least around how this data is imputed in the other surveys. There may be a case for experimenting with imputation for some benefit income, or even sourcing information directly from Centrelink to validate responses provided by lone mothers in LSAC.

Our report has a number of weaknesses. Notably, our analysis is based on relatively small sample sizes in both our comparator surveys. Though we have used weights when conducting any statistical tests, it would be preferable to be able to carry out comparisons on significantly larger samples. This notwithstanding, comparisons of average income for the larger 3 to 9 years and 4 to 9 years samples, in SIH and HILDA respectively, are broadly in line with those for the more restricted (and more comparable) samples used for our analyses.

There are a number of outstanding issues from this report which point to potential avenues for future research. To begin with, there are many instances where the wording of questions between surveys is different and, perhaps more worryingly, instances where the wording of questions within LSAC changes across waves. The extent to which this impacts upon comparisons across surveys or across waves within LSAC was not considered in this report. Future research on the manner in which respondents understand the wording of the questions should be conducted. Another limitation of this report is that it did not consider income from the infant cohort, which could perhaps be addressed in further work. The analysis of income measured across waves in this report is only a first step, and future research should build upon this, considering the importance of income and longitudinal data for social science

research.

Appendixes

Table A1: Comparison of characteristics of LSAC with HILDA Wave 3 and Wave 4 sub-samples (per cent, unless otherwise indicated)

	LSAC	HILDA Wave 3		HILDA Wave 4	
	4-5 years	3-4 years	3-9 years	4-5 years	4-9 years
Family type					
Couple family	85.0	89.0	86.9	89.9	87.2
Lone-parent family	15.0	11.0	13.1	10.1	12.8
Age men					
Age < 30	7.8	10.1	6.3	7.9	5.3
Age 30 - 39 years	58.6	59.4	47.8	53.7	43.4
Age 40 + years	33.6	30.5	45.9	38.4	51.3
Age women					
Age < 30	17.2	24.8	15.8	19.4	13.7
Age 30 - 39 years	64.6	60.9	55.8	61.8	55.2
Age 40 + years	18.2	14.3	28.4	18.9	31.1
Proportion in Employment					
Men	92.3	88.9	88.6	88.2	89.1
Women	55.2	48.3	54.2	50.7	57.1
Proportion with Degree					
Men	26.4	25.3	22.2	26.5	23.3
Women	24.3	22.0	20.6	21.8	20.8
Current weekly income (\$)					
Men	987	1,196	1,116	1,214	1,181
Women	420	411	451	446	507
Sample N					
Men	4,318	379	948	372	776
Women	4,946	488	1,226	484	1,044

Source: LSAC K Cohort Wave 1; SIH 2003-04; HILDA Waves 3 and 4. Authors' calculations.

Table A2: Comparison of incomes of LSAC with SIH age 3 to 9 years and HILDA Wave 4 age 4 to 9 years sub-samples (\$)

	LSAC 4 to 5 years		SIH 3 to 9 years		HILDA (W4) 4 to 9 years	
	Men	Women	Men	Women	Men	Women
Current weekly income	985	407	1,050	466	1,181	507
Family type						
Couple family	985	402	1,050	429	1,181	464
Lone-parent family	-	524	-	624	-	669
Age						
Age < 30	718	377	707	412	869	471
Age 30 - 39 years	991	418	977	463	1,142	492
Age 40 + years	1,061	474	1,197	503	1,249	551
Employment						
In Employment	1,052	560	1,133	635	1,267	652
Not in Employment	214	249	232	235	424	315
Education						
No Yea r12 (No other qualifications)	702	358	736	422	832	415
Yea r12 (No degree)	899	383	938	420	1,114	472
Degree	1,400	569	1,658	635	1,696	739
Occupation*						
Manager/Professional	1,397	748	1,569	829	1,604	894
Other	887	472	932	564	1,103	536
Self-employed*						
Self-employed	972	495	896	472	1,332	869
Employee/other	1,076	577	1,174	659	1,243	615

*Employed only

Table A3: Men's and women's characteristics in LSAC, SIH age 3 to 4 years sub-sample and HILDA age 4 to 5 years sub-sample (unweighted N)

	Fathers			Mothers		
	LSAC	SIH	HILDA	LSAC	SIH	HILDA
Total	3,610	659	372	4,370	785	484
Family type						
Couple family	3,610	659	372	3,755	644	400
Lone-parent family	-	-	-	615	141	84
Age						
Age < 30	287	79	32	747	205	96
Age 30 - 39 years	2,149	390	201	2,855	476	292
Age 40 + years	1,174	190	139	767	104	96
Employment						
In Employment	3,382	601	339	2,499	406	269
Not in Employment	222	58	33	1,871	379	215
Education						
No Year 12 (No other qualifications)	548	140	84	947	226	159
Year 12 (No degree)	1,970	368	186	2,191	391	214
Degree	1,019	151	102	1,224	168	111
Occupation*						
Manager/Professional	1,180	192	132	904	114	97
Other	2,200	405	206	1,593	287	172
Self-employed*						
Self-employed	789	94	95	531	59	51
Employee/other	2,592	507	244	1,966	347	218

* Employed persons only

Table A4: Linearised standard errors associated with men's and women's mean weekly incomes in LSAC, SIH age 3 to 4 years sub-sample and HILDA age 4 to 5 years sub-sample (\$)

	Fathers			Mothers		
	LSAC	SIH	HILDA	LSAC	SIH	HILDA
Total	22.50	39.36	54.98	7.57	16.37	19.40
Family type						
Couple family	22.50	39.36	54.98	7.57	17.69	21.71
Lone-parent family	-	-	-	11.24	38.92	29.61
Age						
Age < 30	23.01	51.81	96.73	9.51	22.16	44.74
Age 30 - 39 years	21.62	37.43	45.56	7.65	22.97	24.30
Age 40 + years	38.52	108.40	127.54	16.45	45.97	45.30
Employment						
In Employment	21.80	40.56	56.56	9.84	25.41	27.30
Not in Employment	14.42	27.57	58.18	5.47	10.11	22.44
Education						
No Year 12 (No other qualifications)	20.90	43.79	63.72	8.01	21.20	33.42
Year 12 (No degree)	18.05	42.66	46.88	6.49	18.29	23.12
Degree	43.76	112.76	156.76	16.09	52.89	53.20
Occupation*						
Manager/Professional	36.09	105.47	92.02	20.48	65.75	58.04
Other	15.47	29.26	68.69	8.12	23.05	25.61
Self-employed*						
Self-employed	38.62	105.07	101.65	22.10	92.02	67.33
Employee/other	21.35	43.98	67.72	10.31	25.40	29.87

* Employed persons only

Table A5: Changes in men's ranking in the income distribution from LSAC Waves 1 to 2 and Waves 2 to 3

		Wave 2				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 1	Bottom quartile	355	135	46	21	557
	Second quartile	166	359	135	37	697
	Third quartile	56	172	380	119	727
	Top quartile	23	31	126	527	707
	Total	600	697	687	704	2,688
		Wave 3				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 2	Bottom quartile	400	149	35	16	600
	Second quartile	155	364	146	32	697
	Third quartile	41	150	407	89	687
	Top quartile	20	27	140	517	704
	Total	616	690	728	654	2,688

Note: Figures in bold represent no change; figures above and to the right of those in bold represent a decline; and figures below and to the left of those in bold represent an increase in the ranking.

Table A6: Changes in women's ranking in the income distribution from LSAC Waves 1 to 2 and Waves 2 to 3

		Wave 2				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 1	Bottom quartile	516	208	103	59	886
	Second quartile	214	381	199	60	854
	Third quartile	72	185	392	151	800
	Top quartile	59	64	187	621	931
	Total	861	838	881	891	3,471
		Wave 3				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 2	Bottom quartile	495	207	114	45	861
	Second quartile	199	394	187	58	838
	Third quartile	84	223	418	156	881
	Top quartile	29	55	173	634	891
	Total	807	879	892	893	3,471

Note: Figures in bold represent no change; figures above and to the right of those in bold represent a decline; and figures below and to the left of those in bold represent an increase in the ranking.

Table A7: Changes in men’s ranking in the income distribution from HILDA Waves 4 to 6 and Waves 6 to 8

		Wave 6				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 4	Bottom quartile	108	32	10	5	155
	Second quartile	29	81	29	10	149
	Third quartile	9	35	94	21	159
	Top quartile	11	4	29	117	161
	Total	157	152	162	153	624

		Wave 8				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 6	Bottom quartile	110	29	14	4	157
	Second quartile	25	90	31	6	152
	Third quartile	8	33	91	30	162
	Top quartile	6	5	25	117	153
	Total	149	157	161	157	624

Note: Figures in bold represent no change; figures above and to the right of those in bold represent a decline; and figures below and to the left of those in bold represent an increase in the ranking.

Table A8: Changes in women’s ranking in the income distribution from HILDA Waves 4 to 6 and Waves 6 to 8

		Wave 6				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 4	Bottom quartile	133	37	25	10	205
	Second quartile	46	98	50	12	206
	Third quartile	13	57	92	45	207
	Top quartile	8	12	44	145	209
	Total	200	204	211	212	827

		Wave 8				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 6	Bottom quartile	136	42	16	6	200
	Second quartile	47	101	38	18	204
	Third quartile	13	54	105	39	211

Top quartile	7	7	46	152	212
Total	203	204	205	215	827

Note: Figures in bold represent no change; figures above and to the right of those in bold represent a decline; and figures below and to the left of those in bold represent an increase in the ranking.

Table A9: Changes in ranking of households in the income distribution from LSAC Waves 1 to 2 and Waves 2 to 3 (unweighted N)

		Wave 2				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 1	Bottom quartile	481	96	24	10	611
	Second quartile	163	381	154	39	737
	Third quartile	35	237	381	133	786
	Top quartile	20	32	170	575	797
	Total	699	746	729	757	2,931

		Wave 3				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 2	Bottom quartile	534	127	27	11	699
	Second quartile	147	412	159	28	746
	Third quartile	46	185	376	122	729
	Top quartile	6	30	187	534	757
	Total	733	754	749	695	2,931

Note: Figures in bold represent no change; figures above and to the right of those in bold represent a decline; and figures below and to the left of those in bold represent an increase in the ranking.

Table A10: Changes in ranking of households in the income distribution from HILDA Waves 4 to 6 and Waves 6 to 8 (unweighted N)

		Wave 6				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 4	Bottom quartile	144	31	8	6	189
	Second quartile	38	111	46	10	205
	Third quartile	8	47	113	37	205
	Top quartile	6	6	39	159	210
	Total	196	195	206	212	809

		Wave 8				
		Bottom quartile	Second quartile	Third quartile	Top quartile	Total
Wave 6	Bottom quartile	143	43	7	3	196

Second quartile	32	104	54	5	195
Third quartile	9	43	113	41	206
Top quartile	8	10	33	161	212
Total	192	200	207	210	809

Note: Figures in bold represent no change; figures above and to the right of those in bold represent a decline; and figures below and to the left of those in bold represent an increase in the ranking.

References

- AIFS (2010), *Growing Up in Australia: The Longitudinal Study of Australian Children Data User Guide*, April 2010, Australian Institute of Family Studies, Melbourne.
- Austen, S. and Redmond, G. (2008). 'Women's Incomes', in *Australian Social Trends 2008* (pp. 153-157) ABS Cat. No.4102.0, Australian Bureau of Statistics, Canberra.
- Australian Bureau of Statistics (2005), *Household Expenditure Survey, Australia: Summary of Results 2003-04*, Cat. No. 6530.0, Australian Bureau of Statistics, Canberra.
- Bradbury, B. (2007), *Child Outcomes and Family Socio-Economic Characteristics: Final Report of the Project LSAC Outcomes and the Family Environment*, May 2007, Social Policy Research Centre, University of NSW, <http://www.sprc.unsw.edu.au/reports/LSACOutcomesSPRCReport.pdf>.
- Collins, D. and White, A. (1996), 'In Search of an Income Question for the 2001 Census' *Survey Methodology Bulletin*(39), 2-10.
- Davern, M., Rodin, H., Beebe, T. J. and Call, K. T. (2005), 'The Effect of Income Question Design in Health Surveys on Family Income, Poverty and Eligibility Estimates' *Health Services Research*, vol.40(5 Pt .1), 1434-1552.
- Duncan, G. J., Yeung, W. J., Brooks-Gunn, J. and Smith, J. R. (1998), 'How Much Does Childhood Poverty Affect the Life Chances of Children?' *American Sociological Review*, v.63(3), 406-423.
- Harding, A. (1997), 'The Suffering Middle: Trends in Income Inequality in Australia, 1982 to 1993-94' *The Australian Economic Review*, vol.30(4), 341-358.
- Harding, A. and Szukalska, A. (1999), *Trends in Child Poverty in Australia: 1982 to 1995-96*, Discussion Paper no. 42, April 1999, NATSEM, University of Canberra, http://www.canberra.edu.au/centres/natsem/publications?sq_content_src=%2BdXJsPWh0dHAIM0EIMkYIMkZ6aWJvLndpbi5jYW5iZXJyYS5lZHUuYXUIMkZuYXRzZW0lMkZpbmRleC5waHAIM0Ztb2RIJTNEcHVibGljYXRpb24lMjZwdWJsaWNhdGlvbiUzRDYyMCZhbGw9MQ%3D%3D.
- Haveman, R. and Wolfe, B. (1994) *Succeeding Generations: On the Effects of Investments in Children*, Russell Sage Foundation, New York.
- Katz, I. and Redmond, G. (2009). 'Family Income as a Protective Factor for Child Outcomes', in K. Rummery, I. Greener and C. Holden (eds.), *Social Policy Review 21* (pp. 167-197)Policy Press, Bristol.
- Khanam, R., Nghiem, H. S. and Connelly, L. B. (forthcoming), 'Child Health and the Income Gradient: Evidence from Australia' *Journal of Health Economics*, (available online).

- Mayer, S. (1997) *What Money Can't Buy: Family Income and Children's Life Chances*, Harvard University Press, Cambridge, Mass.
- Micklewright, J. and Schnepf, S. V. (2010), 'How Reliable are Income Data Collected With a Single Question?' *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol.173(2), 409-429.
- Moore, J. C., Stinson, L. L. and Welniak, E. J. J. (2000), 'Income Measurement Error in Surveys: A Review' *Journal of Official Statistics*, vol.16(4), 331-361.
- Redmond, G. (1999). 'Income Sharing and Income Distribution', in Australian Bureau of Statistics (ed.), *Australian Social Trends 1999 Cat. No. 4102.0* (pp. 129-133) Australian Bureau of Statistics, Canberra.
- Sanson, A., Nicholson, J., Ungerer, J., Zubrick, S., Wilson, K., Ainley, J., Berthelsen, D., Bittman, M., Broom, D., Harrison, L., Rodgers, B., Sawyer, M., Silburn, S., Strazdins, L., Vimpani, G. and Wake, M. (2002), *Introducing the Longitudinal Study of Australian Children*, LSAC Discussion Paper no.1, June 2002, Australian Institute of Family Studies
<http://www.aifs.gov.au/growingup/pubs/discussion/dp1/>.
- Saunders, P. G. and Bradbury, B. (2006), 'Monitoring Trends in Poverty and Income Distribution: Data, Methodology and Measurement' *Economic Record*, vol.82(258), 341-364.
- Siminski, P., Saunders, P. G. and Bradbury, B. (2003), 'Reviewing the Intertemporal Consistency of ABS Household Income Data through Comparisons with External Aggregates' *The Australian Economic Review*, vol.36(3), 333-349.
- Smeeding, T. M. (2002). *Globalisation, Inequality and the Rich Countries of the G-20: Evidence from the Luxembourg Income Study (LIS)*. Paper presented at the Conference on Globalisation, Living Standards and Inequality: Recent Progress and Continuing Challenges, Sydney, 27-28 May 2002,
- Soloff, C., Lawrence, D., Misson, S. and Johnstone, R. (2006), *The Longitudinal Study of Australian Children: Wave 1 Weighting and Non-Response*, LSAC Technical Paper No. 3, May 2006, Australian Institute of Family Studies, Melbourne. <http://www.aifs.gov.au/growingup/pubs/technical/tp3.pdf>.
- Soloff, C., Millward, C. and Sanson, A. (2003), *Growing Up In Australia: The Longitudinal Study of Australian Children: Proposed Study Design and Wave 1 Data Collection*, Discussion Paper No. 2, Australian Institute of Family Studies, Melbourne. <http://www.aifs.gov.au/growingup/pubs/dp2/dp2.html>.
- UNICEF (2005), *Child Poverty in Rich Countries 2005*, Innocenti Report Cards No.6, UNICEF Innocenti Research Centre, Florence.
- Watson, N. (2010) *HILDA User Manual – Release 8*, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

List of shortened forms

ABS	Australian Bureau of Statistics
AIFS	Australian Institute of Family Studies
BSA	British Social Attitudes Survey
FaHCSIA	Australian Government Department of Families, Housing, Community Services and Indigenous Affairs
FRS	Family Resources Survey
HILDA	Household Income and Labour Dynamics Australia survey
LSAC	Longitudinal Study of Australian Children
Omnibus	Office for National Statistics Omnibus Survey (UK)
SIH	Survey of Incomes and Housing

Endnotes

-
- ¹ The supplementary questions on incomes of other household members aged over 15 years besides the respondent and her partner do allow an approximation of household income to be calculated at LSAC Waves 2 and 3.
 - ² Such an approach would still only achieve partial comparability, since, as noted above, the LSAC sample is drawn from Medicare records, and the SIH and HILDA samples are drawn from households. Moreover, structures of non-response are likely to differ across the surveys.
 - ³ Note that in this part of the analysis we restrict the scope of the analysis to parents in heterosexual two-parent households and lone mother families. That is, we exclude families headed by same sex couples and families headed by lone fathers. Both groups are small and do not influence overall results.