Australian Government
Department of Social Services
Australian Institute of Family Studies

Growing Up in Australia:
The Longitudinal Study of Australian Children (LSAC)
LSAC Technical Paper No. 27

The Longitudinal Study of Australian Children

Missing values in household income over time

Kei Sakata, Neha Swami, Mabel Andalon,
Diana Warren and Chris Schilling

May 2023

Acknowledgements

This report makes use of data from *Growing Up in Australia*: The Longitudinal Study of Australian Children (LSAC). LSAC is conducted in a partnership between the Department of Social Services (DSS), the Australian Institute of Family Studies (AIFS) and fieldwork providers (the Australian Bureau of Statistics (ABS), up to and including Wave 9, and Roy Morgan, from Wave 10 onwards), with advice provided by a consortium of leading researchers.

The authors wish to thank Jennifer Prattley, Lisa Mundy, Amanda Vittiglia and the AIFS LSAC team for their contributions to the paper and project. The authors would especially like to thank LSAC participants for their ongoing contribution to the study.

Growing Up in Australia: The Longitudinal Study of Australian Children is conducted in partnership between the Australian Government Department of Social Services, the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS), with advice provided by a consortium of leading researchers from research institutions and universities throughout Australia growingupinaustralia.gov.au.

Technical paper

The Longitudinal Study of Australian Children: LSAC Technical paper No. 27. Missing values in household income over time.

Authors: Kei Sakata, Neha Swami, Mabel Andalon, Diana Warren and Chris Schilling

For more information, write to:
Longitudinal Studies – Research and Methods
Data Strategy and Development| Data and Evaluation
Australian Government Department of Social Services
PO Box 7576
Canberra Business Centre ACT 2610

Email: LongitudinalStudiesDataAccess@dss.gov.au

ISBN 978-1-76016-277-1 (Online)
ISBN 978-1-76016-278-8 (PDF)

LSAC_Household income_paper27

Contents

List of figures

List of tables

Glossary

Term	Description
*fn09a	Parent 1 reported usual weekly income
*fn09ai	Post-imputation Parent 1 usual weekly income
*fn09b	Parent 2 usual weekly income, as reported by Parent 1
*fn09bi	Post-imputation Parent 2 usual weekly income
*fn09o	Total of other adults' usual weekly income, reported by Parent 1. Other adults are defined as non-parent residents of the home aged 15+ (*oth15)
*hinc	Household income; calculated as ***fn09a** + ***fn09b** + ***fn09o**
*hinci	Post-imputation household income; calculated as ***fn09ai** + ***fn09bi** + ***fn09o**
*oth15	Resident in home other than parent aged 15+
*parentinci	Total parent income derived from post-imputation Parent 1 and Parent 2 income; calculated as ***fn09ai** + ***fn09bi**
*prel	Study child relationship to parents
LSAC	*Growing Up in Australia*: The Longitudinal Study of Australian Children
MAR	Missing at random
MCAR	Missing completely at random
MNAR	Missing not at random
Parent 1 (P1)	The resident parent of the Study Child/Young Person with legal or actual custody of, and who knows the most about the Study Child/Young Person.
Parent 2 (P2)	The Study Child's/Young Person's other resident parent/guardian, or the legal (married) or de-facto partner of Parent 1, including same-sex couples.
Study Child (SC)	The Study Child is the sampling unit of interest. Once an SC reaches the age of 18, they are referred to as a YP (Young Person).

Overview

Growing Up in Australia: The Longitudinal Study of Australian Children (LSAC) provides measures of household income for the Birth (B) cohort to Wave 8 and Kindergarten (K) cohort to Wave 7.[1] These measures are derived from separate income values of household members: the usual weekly individual income for each of Parent 1 and Parent 2 (where present in the household), and a total amount for all other resident adults aged 15 and over (where present in the household). Parent 1 has responsibility for reporting each amount. Imputed individual income variables for Parent 1 and Parent 2 are also available where amounts are unreported, but insufficient information is collected about other adults in the household to be able to impute their income where values are missing.

The household income measure is calculated as the sum of reported Parent 1, Parent 2 and other adult incomes.[2] A second measure, referred to as 'imputed household income', is the total of post-imputation Parent 1 and Parent 2 variables, with values of other adult income included where these are reported. However, missingness exists within both household level income variables where income of other adults is not reported. There is an increasing number of households with 'other adults' in successive waves of LSAC, mainly due to study children and their siblings reaching the age of 15, and reported values for 'other adult' incomes tend to be more often missing than those for Parent 1 and Parent 2. Consequently, missingness persists and increases in household income measures across waves.

Data analysts can derive an alternative post-imputation measure of income, total parental income, that does not depend on income of other adults. Parental income is available for over 97% of families in LSAC, reducing the bias that occurs if the sample is restricted to those with non-missing household income. Parental income can also be adjusted by an equivalence factor that accounts for the size and composition of the household, should the analyst wish. This provides users with an alternative measure of income at the family level that maximises sample size and facilitates comparison across sampling units.

Key messages

- The proportion of missing values in household income measures increases across LSAC Waves 1 to 8. When LSAC study children (and their siblings) reach the age of 15 they are counted as 'other adults'. Some contribute to household income by engaging in paid work, but details of their income are often unreported, leading to missingness in the household measures.
- Limited information is available on other, non-parent adult members of households, meaning imputation of their unreported income is not viable.
- Removing cases of unreported other adult income may bias samples in favour of higher income families, particularly in later waves of LSAC where older siblings of the study child are considered 'other adults'.
- In Wave 7 of LSAC, there are significant differences in the socio-demographic characteristics of the subsample with household income information, compared to the subsample where household income is missing. The subsample with missing household income is more likely to have lower socio-economic positions in terms of parental employment and education.
- This report describes a measure of equivalised parental income that analysts can derive as an alternative household income, based on the total of Parent 1 and Parent 2 incomes and adjusted using an equivalence factor that accounts for the size and composition of the nuclear family unit.

1 In 2003–04, children aged 0–1 years were recruited to the Birth (B) cohort, and children aged 4–5 years were recruited to the Kindergarten (K) cohort. More information about the LSAC study can be found at growingupinaustralia.gov.au. No imputed income values were included in Wave 8 (K cohort) and Wave 9 (B cohort) due to the issues with 'other adult' income as detailed in this report.

2 A household may contain multiple families or unrelated other individuals. Where individuals are aged 15 or above, their income is included in the household income calculation.

Introduction

The first wave of LSAC data was collected in 2003/04. Wave 7 was carried out in 2015/16 when the study children of the B cohort were aged 12–13 and the K cohort were 16–17. Wave 8 took place around two years later. LSAC maintains high unit response rates across waves with 89% of Wave 6 participants responding in Wave 7, and 90.2% of Wave 7 respondents taking part in Wave 8. The survey collects a wide range of information on the study child, their parent(s), and their household and environment. Measures of individual parent income for both cohorts have been provided in LSAC since its inception[3] to Wave 8. Household income variables have been given for B cohort to Wave 8 and K cohort to Wave 7. This paper outlines missing income data across Waves 1–7 and outlines strategies for using these variables.

Income variables in LSAC

Reported income variables

The income variables in the LSAC released data that are relevant to this report are listed in Table 1. Further details on how the questions were asked in the survey, including response options, are in Appendix 1. The asterisk (*) at the beginning of a variable name represents the child age indicator, which takes value a (for a child aged 0–1) through to *i* (for a child aged 16–17). Income data in LSAC are collected in the face-to-face interview. Parent 1 is asked to state his/her usual weekly income (***fn09a**), the income of Parent 2 where present (***fn09b**), and a total income for all other household members aged 15 and over (***fn09o**). Responses to ***fn09o** are assumed to include income of the study child once they turn 15. Income for other members (***fn09o**) is added to total parental income (***fn09a + *fn09b**) to give a measure of income at the household level (***hinc**).

Table 1: Income variables in LSAC released dataset discussed in this report

Variable	Label	Source
*fn09a	Usual weekly income of Parent 1	Reported by Parent 1
*fn09b	Usual weekly income of Parent 2	Reported by Parent 1 (where present)
*fn09o	Other adults' usual weekly income	Reported by Parent 1 (where present)
*hinc	Household income	***fn09a** + ***fn09b** (where present) + ***fn09o** (where present)
*fn09ai	Usual weekly income of Parent 1 (imputed)	Reported by Parent 1 or imputed where unreported
*fn09bi	Usual weekly income of Parent 2 (imputed)	Reported by Parent 1 or imputed where unreported
*hinci	Household income (imputed)	***fn09ai** + ***fn09bi** + ***fn09o**

Note: * Denotes age indicator.

Imputed income variables

Missing values within each of the Parent 1 and Parent 2 income variables are imputed using a combination of the Nearest Neighbour and Little and Su methods, as detailed in Mullan, Daraganova, and Baker (2015). Imputed income variables are denoted ***fn09ai** for Parent 1 and ***fn09bi** for Parent 2. The same imputation strategy cannot be applied in the case of unknown income of all other adult household members. Key variables used to impute Parent 1 and Parent 2 income include occupation, working hours, education level, gender, language spoken at home, main source of income, and presence of an ongoing medical condition. These measures are at an individual level and are not available for adults in the household other than Parent 1, Parent 2, and the study child. Consequently, the models used to impute unknown income for Parent 1 and Parent 2 cannot be used to impute income for other adult household members (***fn09o**). Missing values therefore persist within this variable, and there is no associated imputed version.

A post-imputation measure of income at the household level (***hinci**) is calculated as the sum of each relevant imputed parental measure (***fn09ai**, ***fn09bi**) and, if applicable, known values of income for all other adult household members (***fn09o**). Missingness within both ***hinc** and ***hinci**, therefore, occurs in cases where information about income for all other adult members is missing.

3 There are differences in the questions asked on income between Wave 1 and the subsequent waves; details are provided in Mullen & Redmond (2009). In order to minimise the effect of different question formats in this paper, we exclude data from Wave 1.

Missingness across income variables

The number and percentage of missing values in household income (***hinc**), household income after imputation of parental income (*hinci), and income of other adult members (***fn09o**) are shown in Table 2.

Missingness in household income (***hinc**) has increased in successive waves. In the B cohort sample, 8.1% were missing in Wave 2 and 19.0% in Wave 7. The percentage of unknown values in the K cohort were similar at Wave 2 (8.5%) but notably higher in Wave 7 (24.2%). Imputing unknown parental income reduces, but does not eliminate, missingness in the household income variable (***hinci**). After imputing the income of each relevant parent within a household, 3.7% of households in Wave 2 of B cohort and 3.7% of K cohort remain without a value for income. After imputing parental income in Wave 7, 9.5% of B cohort households and 16.4% of K cohort households have residual missingness.

The majority of missingness remaining within the household income variable after imputation (***hinci**) is explained by unreported income of other adult members of the household, rather than unknown values in variables used in the imputation model for contributing Parent 1 and Parent 2 incomes. Missing values in ***fn09o** lead to missing data in ***hinci** irrespective of whether or not parental income has been imputed.

Table 2: Missingness in LSAC household and other adult income variables

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
B cohort						
Total number of responses	4,606	4,386	4,242	4,085	3,764	3,381
Number missing on household income ***hinc** (% of all respondent households)	374 (8.1)	382 (8.7)	437 (10.3)	545 (13.3)	572 (15.2)	640 (19.0)
Number missing on imputed household income ***hinci** (% of all respondent households)	170 (3.7)	143 (3.3)	186 (4.4)	216 (5.3)	196 (5.2)	321 (9.5)
Number of households with other adults aged 15+ missing on income of other adults ***fn09o** (% of all respondent households)	161 (3.5)	137 (3.1)	164 (3.9)	187 (4.6)	178 (4.7)	257 (7.6)
K cohort						
Total number of responses	4,464	4,431	4,169	3,956	3,537	3,089
Number missing on household income ***hinc** (% of all respondent households)	377 (8.5)	413 (9.3)	561 (13.4)	607 (15.3)	638 (18.0)	748 (24.2)
Number missing on imputed household income ***hinci** (% of all respondent households)	164 (3.7)	173 (3.9)	247 (5.9)	284 (7.2)	339 (9.6)	506 (16.4)
Number of households with other adults aged 15+ missing on income of other adults ***fn09oa** (% of all respondent households)	149 (3.3)	150 (3.4)	216 (5.2)	295 (7.5)	305 (8.6)	361 (11.7)

Notes: a) Excludes Parent 1, Parent 2 *and* the study child. Study children from the K cohort start turning 15 by Wave 6. If they are included, 71.8% of Wave 6 and 100% of Wave 7 households contain another adult aged 15+.

Source: LSAC Wave 2–7, B and K cohorts

Changing household structure over time

This section highlights that the changing structure and composition of LSAC households over time is the major driver of the missingness identified above. In Wave 7, K cohort study children have reached the age of 16–17, B cohort children were 12–13 years and siblings have also aged. As the study children and their siblings reach the age of 15, they are considered to be 'other adult members' of the household in subsequent waves.

Table 3 shows that the proportion of households that contain an adult aged 15+ who is not a parent of the study child or the study child themselves (in Waves 6 and 7 for K cohort children) increases across successive waves. These families comprise 44.8% of the Wave 7 B cohort and 56.7% of K cohort, up from 12.0% and 17.3% in Wave 2 for each cohort respectively.

These households can be categorised as one of two types, based on whether or not there is a sibling of the study child who is over the age of 15. That is, they can be categorised into either:

- households with at least one other adult that is a sibling of the study child

  or

- households with at least one other adult, but none are siblings of the study child.

Differentiating between households in this way can help show the extent to which siblings growing older might explain increasing levels of missingness in the household income variables, over successive waves. Sibling income could be difficult for Parent 1 to report as it may come from irregular, casual work. Income earned by other non-sibling adults – such as grandparents, for example – is more likely to include regular payments, which could include pension benefits paid at a set amount, making it easier for Parent 1 to include in their estimation of total other adult income.

Table 3 shows the number and percentage of households of each type, for each cohort in Waves 2 to 7, where adult siblings are aged 15 or over. In Wave 2, 8.9% of all B cohort respondent households had at least one adult aged 15+ who was a sibling, but this increased to 43.8% in Wave 7. The proportion of households with no adult siblings fell, over the same time frame, from 5.2% to 1.0%. For K cohort, 15.7% of all responding households had at least one sibling aged 15+ in Wave 2, increasing to 54.6% by Wave 7. However, the percentage of households with other adults who were not siblings decreased from 1.6% to 1.4%. The increase in the proportion of respondent households with other adults aged 15+ can therefore be explained by siblings growing older.

Table 3: LSAC households containing other adults aged 15+

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
B cohort						
Number of respondents	4,606	4,386	4,242	4,085	3,764	3,381
Age of study child (mode)	2–3 years	4–5 years	6–7 years	8–9 years	10–11 years	12–13 years
Number (%) of respondent households with other adults aged 15+[a]	551 (12.0)	540 (12.3)	648 (15.3)	772 (18.9)	1,013 (26.9)	1,513 (44.8)
Number of households with at least one sibling of study child aged 15+ (% of all respondent households)	410 (8.9)	478 (10.9)	600 (14.1)	733 (17.9)	977 (26.0)	1,480 (43.8)
Number of households with other adults aged 15+ and no siblings of study child aged 15+ (% of all respondent households)	141 (3.1)	62 (1.4)	48 (1.1)	39 (1.0)	36 (1.0)	33 (1.0)
K cohort						
Total number of responses	4,464	4,431	4,169	3,956	3,537	3,089
Age of study child (mode)	6–7 years	8–9 years	10–11 years	12–13 years	14–15 years	16–17 years
Number (%) of respondent households with other adults age 15+	771 (17.3)	884 (20.0)	1,197 (28.7)	1,756 (44.4)	1,876 (53.0)	1,752 (56.7)
Number of households with at least one sibling of study child aged 15+ (% of all respondent households)	699 (15.7)	842 (19.0)	1,158 (22.8)	1,712 (43.3)	1,833 (51.8)	1,686 (54.6)
Number of households with other adults aged 15+ and no siblings of study child aged 15+ (% of all respondent households)	72 (1.6)	42 (1.0)	39 (1.0)	44 (1.1)	34 (1.0)	42 (1.4)

Notes: a) Excludes Parent 1, Parent 2 *and* the study child. Study children from the K cohort start turning 15 by Wave 6. If they are included, 71.8% of Wave 6 and 100% of Wave 7 households contain another adult aged 15+.

Source: LSAC Wave 1–7, B and K cohorts

Total parental income

Measuring total parental income

Earlier sections of this report detail an increasing amount of missingness in both reported and imputed household income due to missingness. This section describes the construction of a total parental income variable (***parentinci**) that is independent of income of other resident adults.

In order to create a total parental income variable, it is necessary to distinguish between single parent and two parent families. This is indicated by the variable ***prel**, which describes the type of family in terms of the study child's relationship to his/her parents.

There are five categories to ***prel**:

1. Two biological parents
2. One biological parent and one non-biological parent
3. One biological parent only (no Parent 2)
4. Two non-biological parents
5. One non-biological parent only (no Parent 2).

Families of type 1, 2 or 4 are regarded as two parent families, and those in groups 3 and 5 are considered as single parents. A variable of total parental income (***parentinci**) can be calculated as:

- for two parent families, the sum of Parent 1 and Parent 2 post-imputation incomes (***fn09ai + *fn09bi**)
- for single parent families, the post-imputation income of Parent 1 (***fn09ai**).

Stata code for creating ***parentinci** for cohort B is given in Appendix II. Similar code can be used for cohort K.

Missingness in total parental income

The number and proportion of missing values in ***parentinci** are shown in Table 4. In Wave 7, ***parentinci** was not able to be calculated for 2.87% of B cohort and 2.56% of K cohort households, due to lack of reported information. In comparison, the amount of missingness in the household income variable (***hinci**), which included parental income and income of other adults (***fn09o**) where available, was 9.5% in cohort B and 16.4% in cohort K (Table 2). Rates of missingness in aggregate income variables are therefore notably higher when income of other adults is included in the formulation.

Table 4: Missing values for total post-imputation parental income (*parentinci)

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
B cohort						
Total number of responses	4,606	4,386	4,242	4,085	3,764	3,381
Number of ***parentinci** missing	0	8	7	37	68	97
% of all respondent households	0.00	0.18	0.17	0.91	1.81	2.87
K cohort						
Total number of responses	4,464	4,431	4,169	3,956	3,537	3,089
Number of ***parentinci** missing	2	8	18	47	73	79
% of all respondent households	0.04	0.18	0.43	1.19	2.06	2.56

Source: LSAC Wave 2–7

Distribution of total parental income

The distribution of total parental income (***parentinci**) is compared to that of imputed household income (***hinci**) in Table 5. Note that calculations are based on raw data and no adjustments for inflation have been made, to support reproducibility of figures by data users. Any differences between the two distributions are attributable to the inclusion of income of other adults (***fn09o**) in the household figures. The mean and median value of household income is higher than that of parental income in every wave, reflecting the contribution of other adult income. Furthermore, the difference tends to increase with successive waves. The difference in mean weekly incomes for the B cohort is $50.41 (median = $47.11) in Wave 2, increasing to $149.77 ($69.79) in Wave 7. The difference in means for the K cohort increases from $64.89 (median $49.49) in Wave 2 to $335.46 ($240.77) in Wave 7. These increases likely capture both inflationary effects and the consequence of some study children and siblings in the sampled households getting older and starting to earn income that is included in the household figures.

Patterns of note in variability in the ***hinci** and ***parentinci** measures are, firstly, that the standard deviation of each increases in successive waves, reflecting a greater range of incomes within more recent samples. Secondly, for the K cohort, in most waves there is greater variability in the household than parental income measure. This is likely due to variability across the income of other, non-parental adult members in the household.

Figures 1 and 2 show the kernel density of the household (***hinci**) and parental income (***parentinci**) distributions at Wave 7 for the B and K cohorts respectively. The differences between the two measures, in terms of central tendency and spread, are more marked for K cohort, reflecting figures from Table 5 and the increasing contributions from study children and siblings as they reach the age of 15.

When comparing total parental income across sampling units, users may wish to apply an equivalence factor that adjusts for the different size and composition of families. Appendix C details the construction and application of an equivalence factor.

Table 5: Descriptive statistics for *hinci and *parentinci, LSAC Waves 2–7

	Sample size (*n*)		Mean ($)			Standard deviation ($)		
B cohort	***hinci**	***parentinci**	***hinci**	***parentinci**	**Difference**	***hinci**	***parentinci**	**Difference**
Wave 2	4,436	4,606	1,598.80	1,548.39	50.41	1,095.45	1,173.63	-78.18
Wave 3	4,243	4,378	1,890.29	1,831.00	59.29	1,349.51	1,327.60	21.90
Wave 4	4,056	4,235	2,032.59	1,977.99	54.60	1,468.81	1,516.95	-48.13
Wave 5	3,869	4,048	2,326.43	2,235.83	90.60	1,728.51	1,718.05	10.47
Wave 6	3,568	3,696	2,480.85	2,407.18	73.67	1,753.37	1,780.61	-27.24
Wave 7	3,060	3,284	2,717.76	2,568.00	149.77	1,996.96	1,819.04	177.93
K cohort	***hinci**	***parentinci**	***hinci**	***parentinci**	**Difference**	***hinci**	***parentinci**	**Difference**
Wave 2	4,300	4,462	1,654.38	1,589.49	64.89	1,188.03	1,162.32	25.72
Wave 3	4,158	4,323	1,938.40	1,867.82	70.57	1,332.27	1,314.12	18.15
Wave 4	3,922	4,151	2,099.11	2,026.45	72.66	1,449.73	1,566.38	-116.65
Wave 5	3,672	3,909	2,360.92	2,236.02	124.90	1,667.81	1,618.60	49.21
Wave 6	3,198	3,464	2,577.96	2,408.96	169.00	1,754.59	1,689.98	64.62
Wave 7	2,583	3,010	2,870.86	2,535.40	335.46	1,939.91	1,706.55	233.35

Figure 1: B cohort Wave 7 density plot of post-imputation household income (*hinci) and parental income (*parentinci)

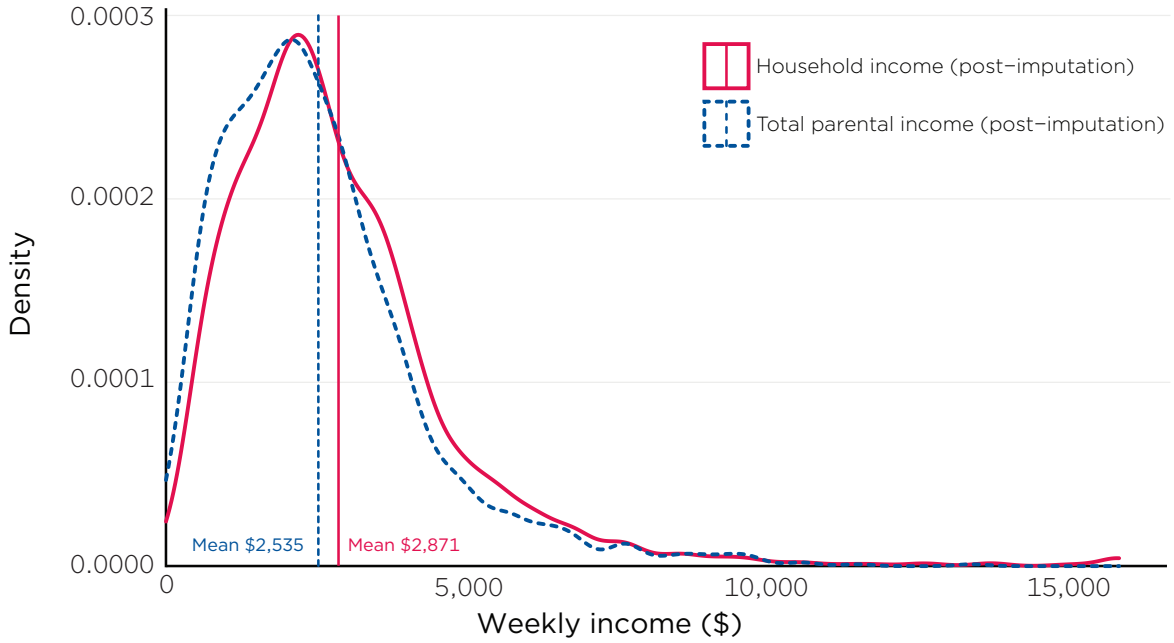
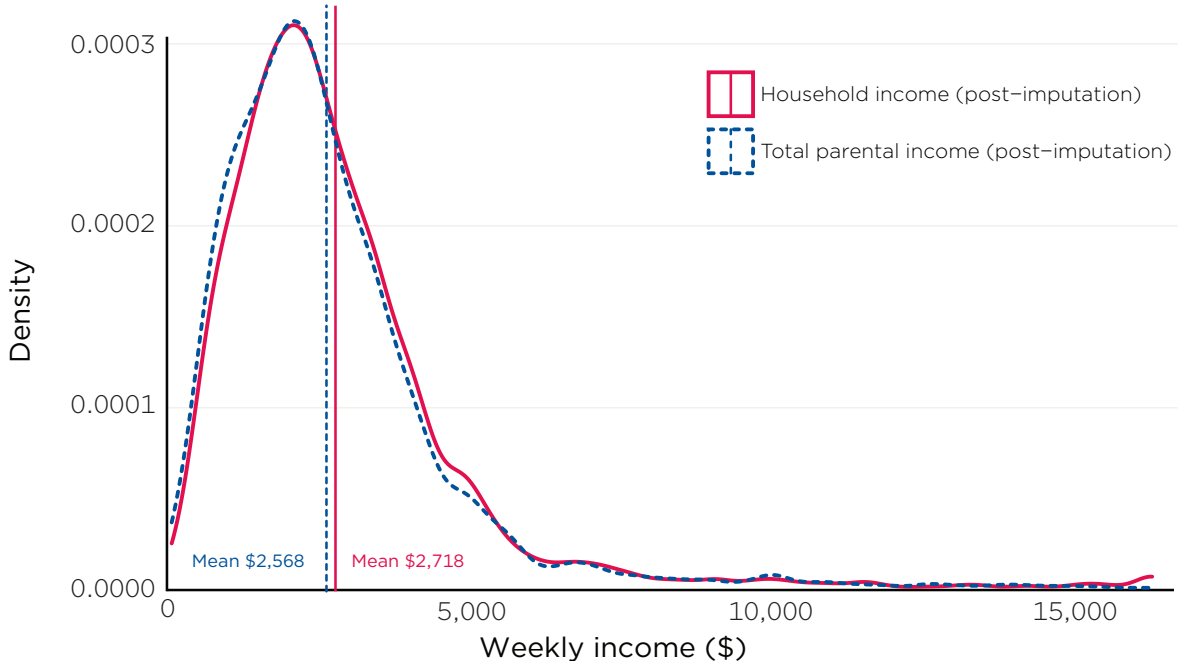


Figure 2: K cohort Wave 7 density plot of post-imputation household income (*hinci) and parental income (*parentinci)



Note: Due to confidentialisation of the household income data, *hinci is top-coded.

Sample differences using household income versus parental income

Differences in income

The previous section described a measure of total parental income (***parentinci**), which is derived only from the post-imputation income of Parent 1 and Parent 2 (where present). Analysts have the choice of using this new variable or using the original imputed household income variable (***hinci**) that has more missingness. This section investigates the sample representativeness of using ***hinci** given that it typically involves removing households lacking information on other adults (***fn09o**) from the analytic sample. The pertinent question is: what kind of households would be lost from the sample if all observations without an income reported for other adults, or without the total household measure, are dropped?

First, we examine differences in parental income when the sample is restricted to households with an adult member aged 15 and over who is not a parent. For each wave and cohort, the mean income in ***parentinci** is calculated for households with known values of other adult income (***fn09o**), and also for those missing that information. Results from *t*-tests conducted on the difference in means of ***parentinci** for these two groups are presented in Table 6. With the exception of B cohort Wave 2 and K cohort Wave 6, the differences are statistically significant and positive, with higher means observed in the samples of households who have a reported ***fn09o** value. Income of other adults aged 15+, therefore, tends to be reported by parents in households with higher average parental income. These results indicate that using ***hinci** and removing households with missing information on ***fn09o** would result in a sample with higher average parental incomes.

Table 6: *t*-tests for difference in mean of *parentinci between households with known and unknown values in *fn09o, LSAC Waves 2–7

		B cohort				K cohort			
	Group	*n*	Mean *parentinci ($)	Std error. ($)	*t*-test	*n*	Mean *parentinci ($)	Std error ($)	*t*-test
Wave 2	Non-missing *fn09o	390	1,202.60	50.02		625	1,352.90	44.31	
	Missing *fn09o	161	1,052.20	75.12		148	1,145.00	72.39	
	Difference		150.5	91.61			207.9	97.64	**
Wave 3	Non-missing *fn09o	402	1,659.10	63.6		733	1,700.50	46.87	
	Missing *fn09o	132	1,091.60	76.39		147	1,349.60	83.9	
	Difference		567.5	119.35	***		350.9	111.23	***
Wave 4	Non-missing *fn09o	482	1,817.40	63.5		978	1,931.80	52.31	
	Missing *fn09o	162	1,344.50	76.28		216	1,504.40	84.46	
	Difference		473	118.16	***		427.4	118.19	***
Wave 5	Non-missing *fn09o	584	2,097.40	70.28		1,459	2,194.00	42.63	
	Missing *fn09o	181	1,604.60	79.52		272	1,801.50	79.61	
	Difference		492.7	133.81	***		392.5	104.54	***
Wave 6	Non-missing *fn09o	834	2,313.70	59.88		2,101	2,508.00	38.99	
	Missing *fn09o	158	1,984.70	146.86		379	2,107.00	75.82	
	Difference		329	151.69	**		401	97.28	
Wave 7	Non-missing *fn09o	1,254	2,619.30	51.06		2,572	2,596.80	33.92	
	Missing *fn09o	240	2,317.20	113.99		438	2,174.90	75.38	
	Difference		302.1	126.92	**		421.9	87.89	***

Notes: ***, ** denote significance at a 1% and 5% level respectively. The sample is confined to households with an adult member aged 15+ who is not a parent or study child.

Differences in socio-demographics

Using Wave 7 as an example, there are significant differences in some socio-demographic characteristics of the subsample with household income information, compared to the subsample where household income is missing but parental income can be calculated (Table 7). Compared to the subsample of households with household income information (***hinci**), the subsample with missing household income, but known parental income, is more likely to include:

- study children with older siblings
- for K cohort only, study children that are currently working
- higher average age of Parent 1 and Parent 2 (although this difference is relatively small)
- higher percentage of households in which no parent is employed
- higher percentage of single parent households
- fewer households in which at least one parent has a degree qualification.

These findings suggest that households with missing household income (***hinci**) but known parental income are more likely to have a lower socio-economic position in terms of parental employment and education. A part of the explanation for these findings could be that, in general, LSAC families from low socio-economic backgrounds are likely to have higher non-response rates (Bandara et al., 2018). It could also be the case that parents with lower levels of education and poor employment outcomes may be less aware of their children's income, potentially due to challenges relating to financial literacy or budgeting; or they may regard a child's income as belonging to the child, rather than as contributing to the wider household unit, and not report it as such.

These results also indicate that using ***hinci** and removing households with missing information on ***fn09o** would result in a sample with differences in socio-demographic characteristics, compared to the fuller sample available using the post-imputation total parent income measure.

Table 7: Differences in socio-demographic characteristics between households with missing and non-missing values in *hinci, LSAC Wave 7

B cohort				K cohort			
Characteristics	*hinci non-missing	*hinci missing but *parentinci non-missing	Pearson chi2/ *t*-test	Characteristics	*hinci non-missing	*hinci missing but *parentinci non-missing	Pearson chi2/ *t*-test
Birth order of Study Child (%)				**Birth order of Study Child (%)**			
Oldest child	37.3	15.1	***	Oldest child	44.5	24.6	***
Middle child/Twin	20.8	23.5		Middle child/Twin	14.5	20.5	
Youngest child	33.2	51.4		Youngest child	23.8	42.3	
Only child	8.8	10.0		Only child	17.2	12.6	
Total, *N*	3,060	251		Total, *N*	2,572	459	
Household structure (%)				**Household structure (%)**			
Single parent household	14.4	33.9	***	Single parent household	17.3	26.6	***
Two parent household	85.6	66.1		Two parent household	82.7	73.4	
Total, *N*	3,060	251		Total, *N*	2,572	459	
Parental employment (%)				**Parental employment (%)**			
At least one parent employed	95.3	85.4	***	At least one parent employed	94.6	90.0	***
No parent employed	4.8	14.6		No parent employed	5.4	10.0	
Total, *N*	3,029	247		Total, *N*	2,548	451	
Highest parental education (%)				**Highest parental education (%)**			
Year 12 or below	6.6	11.2	***	Year 12 or below	9.3	14.5	***
Certificate/Diploma	41.1	54.1		Certificate/Diploma	41.4	45.4	
Degree	52.3	34.7		Degree	49.4	40.1	
Total, *N*	3,027	242		Total, *N*	2,498	441	
Parent 1 age (years)				**Parent 1 age (years)**			
Avg. age of P1	44.0	44.9	***	Avg. age of P1	47.6	48.8	***
Total, *N*	3,060	251		Total, *N*	2,537	444	
Parent 2 age (years)				**Parent 2 age (years)**			
Avg. age of P2	46.0	48.2	***	Avg. age of P2	49.6	50.7	***
Total, *N*	2,619	166		Total, *N*	2,178	362	

Table continued over page →

B cohort				K cohort			
Characteristics	*hinci non-missing	*hinci missing but *parentinci non-missing	Pearson chi2/ *t*-test	Characteristics	*hinci non-missing	*hinci missing but *parentinci non-missing	Pearson chi2/ *t*-test
Study child worked in last 12 months (%)				**Study child's current employment status (%)**			
Yes	14.6	18.2		Currently employed	48.0	53.00	**
No	85.4	81.8		Looking for work	9.7	11.0	
				Not in labour force	42.3	36.0	
	3,041	248			2,459	417	

Note: ***, ** denote significance at 1% and 5% level respectively

Conclusion

This paper has examined issues associated with missingness within household income data in LSAC across Waves 1 to 7. Household income (***hinc**) is constructed by adding incomes of Parent 1 (***fn09a**), Parent 2 where present (***fn09b**), and a total amount for other household members aged 15+ where relevant (***fn09o**). Imputation is used to address missingness within parental income variables (***fn09ai** and ***fn09bi**) but a lack of information on other household members prevents the same methodology being used to impute values of other adult income. Missingness therefore persists within post-imputation household income (***hinci**).

The proportion of missing values in household income variables (***hinc**, ***hinci**) increases in successive waves. This is largely explained by increases in non-reporting of income for adult members of the household other than parents. As the study child and their siblings reach their mid-to-late teens, this becomes more of an issue because the percentage of households with other adult members aged 15+ increases. In such households, the income of those other adult members is less likely to be reported among low parental income families. Parental income is, on average, higher in the subsample for whom household income is available, compared to those with missing household income. Further, there are significant differences in socio-demographic characteristics of these two groups, which may result in biased estimates if the sample used for analysis is restricted to those with non-missing household income. A measure of total parental income is an alternative income variable that is comparable across sampling units and maximises the available sample size.

Bibliography

Australian Bureau of Statistics (ABS). (2014). *Family, Household and Income Unit Variables*, 1286.0. Canberra: ABS.

Australian Bureau of Statistics. (2016). *Census of Population and Housing: Census Dictionary*, 2901.0. Canberra: ABS.

Australian Institute of Family Studies. (2018). *Data User Guide: The Longitudinal Study of Australian Children.* Melbourne: Australian Institute of Family Studies.

Bandara, D., Gasser, C., Jessup, K., Renda, J., Warren, D., & Daraganova, G. (2018). *Factors associated with non-response in* Growing Up in Australia: *The Longitudinal Study of Australian Children* (LSAC Technical paper No. 23, June 2019) (unpublished). Melbourne: Australian Institute of Family Studies.

Cobb-Clark, D. A. (2008), Leaving home: What economics has to say about the living arrangements of young Australians. *Australian Economic Review*, *41*(2), 160–176.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). Household surveys in crisis. *Journal of Economic Perspectives*, Fall, 199–226.

Mullan, K., Daraganova, G., & Baker, K. (2015). *Imputing income in the Longitudinal Study of Australian Children* (LSAC Technical paper No. 14, January 2015). Melbourne: Australian Institute of Family Studies.

Mullan, K., & Redmond, G. (2009) *Validating income in the Longitudinal Study of Australian Children* (LSAC Technical paper No. 7, October 2011). Melbourne: Australian Institute of Family Studies.

Qu, L. (2020). Families Then and Now: Households and families. Melbourne: Australian Institute of Family Studies.

Appendix A: Income questions in LSAC

LSAC collects information on income from the main respondent (Parent 1) by face-to-face interview. During the interview, Parent 1 is asked about his/her income, his/her partner's income (if applicable) and the other household members' income (if applicable). The answers are recorded in open-ended numerical format. For example, in Wave 7, Parent 1 is first asked about where s/he receives his/her income from with options of: 1) wages and salary; 2) profit or loss from own unincorporated business or share in partnership; 3) any government pension, benefit or allowance; and 4) any other regular source. Parent 1 is then asked:

> Before income tax, salary sacrifice or anything else is taken out, how much do you usually receive from this/these source(s) in total?
>
> *Interviewer: If respondent unable to answer, prompt for their best estimate.*
>
> 1. Amount
> 2. Nil
> -99. Loss
> -2. Don't know

***fn09a** is created from responses to this question. To create ***fn09b**, a similar question is asked of Parent 2 if s/he is present. Parent 1 answers the income question on behalf of Parent 2 if s/he is not present:

> Before income tax, salary sacrifice or anything else is taken out, how much does P2 usually receive from this/these source(s) in total?

LSAC also asks Parent 1 in relation to the remaining members of households who are aged 15 years or older:

> The next question is about the income of members of your household aged 15 years or over, excluding yourself and P2.
>
> Before income tax, salary sacrifice or anything else is taken out, how much income in total do these people usually receive from all sources?

This income information is the basis for ***fn09o**. In Waves 6 and 7, the K cohort study children fall into the category of 'members of the household aged 15 years or over, excluding Parent 1 and Parent 2'. From Wave 7, LSAC has started directly asking the study child the income questions for the K cohort.

> After the amount of income is asked, in all waves, the following question is asked:
>
> What period does that cover?
>
> 1) Week
>
> 2) Fortnight Monthly
>
> 3) Month
>
> 4) Annual
>
> 5) Other (Specify)

Based on the answer to this question, the income information for Parent 1, Parent 2 and other adult members is converted into weekly income.

Appendix B: Stata code for imputed parental income for the B cohort (*parentinc)

```
// create single parent household indicator
foreach z in b c d e f g {
recode `z'prel (-9/-1 = .)(1 2 4 = 0 "No")(3 5 = 1 "Yes"), gen(`z'solepar)
}

// converting negative values in income variables to missing values
replace `z'fn09ai = . if `z'fn09ai<0
replace `z'fn09bi = . if `z'fn09bi<0

// assemble the P1 and P2 income variables
gen `z'parentinc=`z'fn09ai+`z'fn09bi if `z'solepar==0
replace `z'parentinc=`z'fn09ai if `z'solepar==1

}
```

Appendix C: Equivalised total parental income

The measure of total parental income (***parentinci**) is not impacted by missingness in income of other adults (***fn09o**) and thus maximises sample size. When comparing total parental income across sampling units, users may wish to apply an equivalence factor that accounts for the different size and composition of households. A larger household requires a higher income to achieve a similar standard to living as a smaller household, and while a larger family may have more income, that does not necessarily mean that each member receives the same level of financial benefits as members of smaller households. A per capita household income is one adjustment that can be made for household size, but such an approach assumes an equal weight for every member of the family, adult or child. Moreover, the strategy ignores economies of scale. The Australian Bureau of Statistics (ABS, 2016) describes the construction of a household equivalence factor based on a different points value allocated to each person in a household. One point is allocated to the first adult, 0.5 points to each additional person aged 15 years or over, and 0.3 to each child under the age of 15. The equivalence factor is the sum of these points. Household income is then divided by the equivalence factor to create an equivalised household income variable.

The ABS methodology can be adapted for LSAC, where the focus is on the nuclear rather than extended family. Nuclear families, comprised of parents, the study child and siblings, are more prevalent in Australia and extended families are uncommon (Qu, 2020). Furthermore, extended family members may share the household, but they may not necessarily share responsibility for finances. For the purposes of LSAC parents and siblings (biological or non-biological) are considered to be part of the same nuclear family if they are living in the same household. This is in line with the ABS definition of an income unit as 'a group of two or more people who are usually resident in the same household and are related to each other through a couple relationship and/or parent/dependent child relationship' (ABS, 2014).

To this end, the ABS household equivalence factor is adopted for LSAC with modifications that account for nuclear family members only: 1 point is given to the first parent, 0.5 to the second parent; resident children of either of the study child's parents are given 0.5 points if they are aged 15 or older and 0.3 points if they are under 15. There are a small number of LSAC households where Parent 1 is a grandparent. In such cases, s/he is considered as part of the nuclear family. Equivalence points are summed to create an equivalence factor. Equivalised parental income is calculated by dividing post-imputation total parent income by the equivalence factor:

Equivalised post-imputation total parent income = ***parentinci** / equivalence factor

Descriptive statistics for the equivalence factor for Waves 2 to 7 are presented in Table 8. Users need to be aware of cases (around 30) that arise in later waves, particularly K cohort Waves 6 and 7, where study children move out of their parental home. The minimum value of the equivalence factor in these waves is 1, and this is observed where the study child has formed a single person household. When the study child is under the age of 15, the minimum value for the equivalence factor is 1.3; this happens where the household is comprised of Parent 1, as a single parent, and the study child. Stata code for creating the LSAC equivalence factor are provided in Appendices D and E. Note that the codes for B cohort and K cohort are different. The final lines in the code reflect differences between the two cohorts when converting unrealistic cases (no adult member and/or child member) to missing values.

Table 8: Descriptive statistics for LSAC equivalence factor, Waves 2–7

	B cohort					K cohort				
	N	Mean	Std dev.	Min.	Max.	*N*	Mean	Std dev.	Min.	Max.
Wave 2	4,604	2.13	0.38	1.3	5.6	4,460	2.22	0.41	1.3	5.6
Wave 3	4,382	2.20	0.38	1.3	6.1	4,322	2.25	0.43	1.3	5.3
Wave 4	4,237	2.23	0.40	1.3	5.3	4,149	2.26	0.45	1.3	5.5
Wave 5	4,072	2.24	0.42	1.3	5.1	3,940	2.29	0.48	1.3	5.4
Wave 6	3,755	2.26	0.44	1.3	5.4	3,523	2.36	0.48	1	5
Wave 7	3,315	2.30	0.46	1.3	5.8	3,061	2.43	0.50	1	5.3

Source: LSAC Waves 2–7, B cohort

Figures 3 and 4 compare the mean of unequivalised imputed parental income (***parentinci**) with the equivalised measure, for Wave 7 of each cohort. For both cohorts, the highest mean income is observed among families with four people, and lowest among two people families. Without equivalisation, the difference between the mean incomes of two person families and for four person households is $1,850.20 for B cohort and $1,650.80 for K cohort. After equivalisation, this difference is reduced to $580.30 for B cohort and $418.20 for K cohort. By using the equivalised parental income, the larger households of four or more people are not necessarily considerably better off in terms of income compared to smaller households. Comparison of unequivalised income, therefore, may be misleading in understanding household financial circumstances.

Figure 3: Parental income and equivalised income (B cohort, Wave 7)

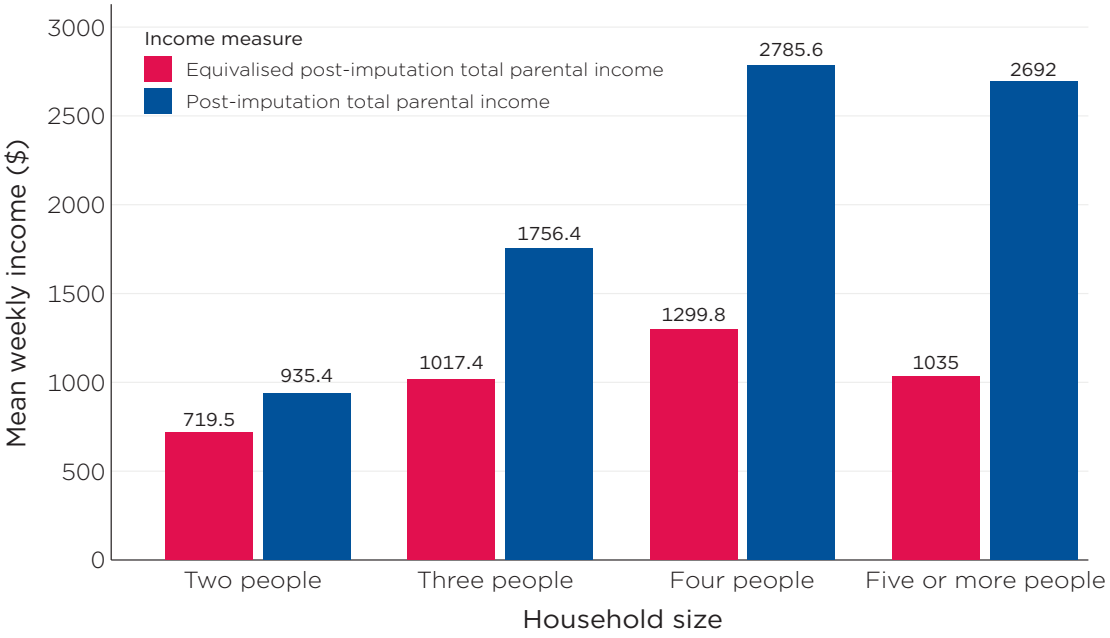
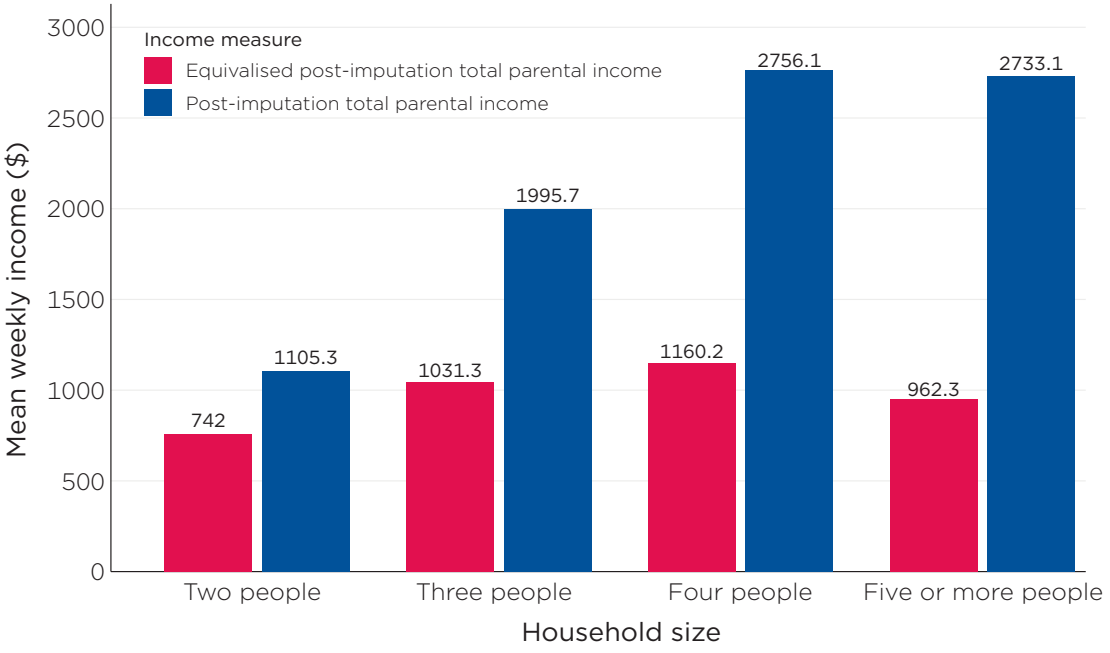


Figure 4: Parental income and equivalised income (K cohort, Wave 7)



Appendix D: Stata code for the equivalence factor for the B cohort

```
// mark a nuclear family member who is present in the household
foreach z in a b c d e f g {
    forval x = 1/21 {
    recode `z'f08m`x' (-9/-1 = 0)(0=1)(1/10=0)(12/17 = 1)(18/22=0)(23/26 =1), gen(`z'nuke`x')
    replace `z'nuke`x' =0 if `z'f01m`x' == 0
    }
}
// create an adult member indicator
foreach z in a b c d e f g {
    forval x = 1/21 {
        recode `z'f03m`x' (-9/14 = 0)(15/110 = 1), gen(`z'ad`x')
        replace `z'ad`x' = 0 if `z'f01m`x' == 0
// replace with 0 for a non-nuclear family member
        replace `z'ad`x' = 0 if `z'nuke`x' ==0
        }
// create a child member indicator
    forval x = 1/21 {
        recode `z'f03m`x' (-9/-1 = 0)(0/14 = 1)(15/110 = 0), gen(`z'cd`x')
        replace `z'cd`x' = 0 if `z'f01m`x' == 0
// replace with 0 for a non-nuclear family member
        replace `z'cd`x' = 0 if `z'nuke`x' ==0
        }
// sum up adult nuclear family indicator and child nuclear family indicator
    egen `z'nadnuke = rsum(`z'ad1-`z'ad21)
    egen `z'ncdnuke = rsum(`z'cd1-`z'cd21)
// add 1 if `z'nadnuke = 0 and P1 is present (P1 is a grandparent for example.)
    replace `z'nadnuke = `z'nadnuke + 1 if `z'nadnuke == 0 & `z'f01m2 == 1
// assemble the equivalence factor.
gen `z'nukeqfactor = 1 + 0.5*(`z'nadnuke-1) + 0.3*(`z'ncdnuke)
//replace the value to . if z'nadnuke == 0 or `z'ncdnuke == 0
replace `z'nukeqfactor = . if `z'nadnuke == 0 | `z'ncdnuke == 0
}
```

Appendix E: Stata code for the equivalence factor for the K cohort

```
// mark a nuclear family member who is present in the household
foreach z in c d e f g h i {
    forval x = 1/21 {
    recode `z'f08m`x' (-9/-1 = 0)(0=1)(1/10=0)(12/17 = 1)(18/22=0)(23/26 =1), gen(`z'nuke`x')
    replace `z'nuke`x' =0 if `z'f01m`x' == 0
    }
}
// create an adult member indicator
foreach z in c d e f g h i {
    forval x = 1/21 {
        recode `z'f03m`x' (-9/14 = 0)(15/110 = 1), gen(`z'ad`x')
        replace `z'ad`x' = 0 if `z'f01m`x' == 0
// replace with 0 for a non-nuclear family member
        replace `z'ad`x' = 0 if `z'nuke`x' ==0
        }
// create a child member indicator
    forval x = 1/21 {
        recode `z'f03m`x' (-9/-1 = 0)(0/14 = 1)(15/110 = 0), gen(`z'cd`x')
        replace `z'cd`x' = 0 if `z'f01m`x' == 0
// replace with 0 for a non-nuclear family member
        replace `z'cd`x' = 0 if `z'nuke`x' ==0
        }
// sum up adult nuclear family indicator and child nuclear family indicator
    egen `z'nadnuke = rsum(`z'ad1-`z'ad21)
    egen `z'ncdnuke = rsum(`z'cd1-`z'cd21)
// add 1 if `z'nadnuke = 0 and P1 is present (P1 is a grandparent for example.)
    replace `z'nadnuke = `z'nadnuke + 1 if `z'nadnuke == 0 & `z'f01m2 == 1
// assemble the equivalence factor.
gen `z'nukeqfactor = 1 + 0.5*(`z'nadnuke-1) + 0.3*(`z'ncdnuke)
}
// replace the value to . if z'nadnuke == 0 or `z'ncdnuke == 0 for c d e f g
foreach z in c d e f g {
    replace `z'nukeqfactor = . if `z'nadnuke == 0 | `z'ncdnuke == 0
}
// replace the value to . if z'nadnuke == 0 for h i ***
foreach z in h i {
    replace `z'nukeqfactor = . if `z'nadnuke == 0
}
```