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. 25

The Longitudinal Study of Australian Children

Wave 9C1 Weighting and non-response

Australian Bureau of Statistics LSAC processing team and the
Australian Bureau of Statistics Household Survey Methodology team

June 2021

Acknowledgements

Growing Up in Australia: The Longitudinal Study of Australian Children, is conducted in a partnership between the Australian Government Department of Social Services, the Australian Institute of Family Studies and the Australian Bureau of Statistics, with advice provided by a consortium of leading researchers.

The Australian Institute of Family Studies is committed to the creation and dissemination of research-based information on family functioning and wellbeing. Views expressed in its publications are those of individual authors and may not reflect those of the Australian Institute of Family Studies.

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.

Contents

List of figures

List of tables

Introduction

The Longitudinal Study of Australian Children (LSAC) began in 2004 with a sample of Australian children of two different age cohorts. The study collects data every two years from this sample, subject to attrition from non-response or non-contact.

The sample in the first year was intended to be representative of Australian children in each of the two selected age cohorts, allowing the assessment of developmental outcomes from infancy until middle childhood. Australian children include citizens, permanent residents and applicants for permanent residency (Soloff, Lawrence, & Johnstone, 2005).

The two cohorts of children included in the study were:

- the B cohort, who were aged 0–1 years at the beginning of the study (born between March 2003 and February 2004)
- the K cohort, who were aged 4–5 years at the beginning of the study (born between March 1999 and February 2000).

The first wave of data collection took place in 2004, with subsequent main waves conducted every two years.

Wave 9C1 of the Longitudinal Study of Australian Children was conducted in 2020 with B cohort children at age 16–17 years and K cohort children at age 20–21 years. The number of active participants continues to decrease from wave to wave. This is a result of failure to maintain contact, participants opting out of the study (including some instances where the study child is actually deceased), or children moving out of scope (e.g. moving overseas). Some children have come back into the sample after missing a wave if contact can be re-established (e.g. if they return from overseas). There were 18,814 families in the original mailout sample, of which 16,342 were contacted and 10,090 successfully recruited to participate in the study. Of these 10,090 children recruited in the Wave 1 sample, 2,958 children responded in Wave 9C1, and 3,259 children responded to all waves.

The use of weighting in analysis

Surveys often use probability samples to allow inferences about the population to be drawn. The Longitudinal Study of Australian Children tracks two child cohorts across time, and these were recruited using a probability sample design. Population inference from longitudinal cohorts over time is enabled using two main strategies: retaining a strong proportion of the original selected cohort through effective tracking and follow-up procedures, and performing missing data analysis to diagnose and correct for inevitable sample attrition.

The composition of the sample, and thus how well it represents the original population from which the sample was drawn, can be affected by non-participation of those chosen in the original random selection. The two main mechanisms of non-participation occur during the initial recruitment stage, when persons in the randomly selected sample cannot be contacted or do not agree to participate, and during subsequent waves, through attrition by loss of contact (non-contact), opting out (refusal), or otherwise moving beyond the scope of collection.

This can result in the composition of the active sample being skewed towards or against some demographics, affecting the ability to make inferences from the responding sample to the population of interest. If skewed demographics are related to study variables of interest, this can lead to bias when making population inferences. Adjusting unit weights to account for attrition can improve the reliability of population inferences.

Survey weights are most commonly defined for calculating descriptive statistics, and are essential in making accurate inferences from sample frequencies, particularly when missing data are not missing at random (Little & Rubin, 1987). Examples of descriptive statistics in a longitudinal study include the proportion of the children achieving a certain level of educational success or the proportion of the cohort improving on their educational success in the time span between waves.

Longitudinal analytic statistics; for example, the strength of correlations of modelled predictors for children improving on their educational success over time, can also be biased if missing participants behave differently to those remaining in the study. Some longitudinal analysis methods reduce bias by applying survey weights, while other methods reduce bias by including variables related to response propensity in the modelling process (Pfeffermann, 1993). Here, we highlight that the responsibility lies with the analyst to ensure that their methods are robust against the possible presence of bias due to missing data (Fairclough, 2010).

With this in mind, this paper describes the process of calculating weights for Wave 9C1 of the Longitudinal Study of Australian Children, with a focus on the treatment of bias. We encourage data users to either make use of survey weights or incorporate into their models those variables we have identified in the weighting process as being related to response propensity. We also offer a timely reminder to users that LSAC is based on a clustered sample design using a primary sampling unit of postcode, and that this variable should be used when conducting statistical tests to avoid overstating significance.

Summary of sample design properties

Full details about the LSAC sample design can be found in Soloff, Lawrence, & Johnstone (2005).

Table 1 provides a summary for reference.

Table 1: LSAC sample design properties

Property	Description
Scope (the population about which inference is to be made)	Two cohorts of children (the B cohort who were 0–1 year and the K cohort who were 4–5 years old during 2004, the Wave 1 recruitment year) The scope excluded very remote areas of Australia.
Coverage (the population represented by the active participating sample)	For Wave 1 recruitment: The subset of the Wave 1 scope who had contact records available through Medicare, who could be contacted and who agreed to participate in LSAC. For subsequent waves: The subset of the Wave 1 coverage who could be contacted. This included tracking address changes and re-recruitment after missing waves where possible, including cases of temporarily moving overseas.
Stratification (division of population into cells from which sample was drawn)	Cells of state x capital city/balance of state x large/small postcode
Selection frame (from which children were selected and contact details obtained)	List frame of Medicare records for children in scope
Sample design	Multi-stage cluster sampling
Selection unit(s)	Stage 1 Unit: Postcode Stage 2 Unit: One cluster of dwellings within postcode Stage 3 Unit: Children in dwellings in cluster
Reporting unit(s)	Parent 1, Parent 2, Child (when old enough), Interviewer, Child care worker, Teacher, Parent Living Elsewhere
Tabulation unit	Child
Selected sample size and fraction	Approximately 10,000 per cohort; approximately 4% of each cohort population
Recruited sample size and fraction at Wave 1	Approximately 5,000 per cohort approximately 2% of each cohort population
Design effects (factors by which variance is higher under cluster sampling as compared to simple random sampling)	Approximately 90% of LSAC variables have a design effect below 1.5 as stated in Wave 1 Weighting Paper.

Responding units

Table 2 reflects information, also provided in the Data User Guide, about what has been included as a responding record for each wave and cohort and, therefore, in scope for weighting.

In the earlier waves the Parent 1 was the main respondent, whereas over the waves the Study Child started to become a key respondent on their own. Therefore, records were considered responding if we were able to conduct an interview with either the Parent 1 or the Study Child. In Waves 1–7 confirmation of a Parent 2 or a Parent Living Elsewhere was contingent on a Parent 1 interview occurring. However, in Wave 8, parent records were generated if the Study Child did not object to a particular parent being interviewed, regardless of whether the Study Child was interviewed or not. In Wave 8, records were considered responding if we were able to

conduct an interview with the young person or any of the parental figures. In Wave 9C1, records were considered responding if the young person or any of the parental figures completed a CAWI.

Table 2: Responding units by wave

Cohort	Waves								
	1	2	3	4	5	6	7	8	9C1
B	P1	P1	P1	P1	P1 or SC	P1 or SC	P1 or SC	P1 or SC	SC or W8P1 or W8P2 or W8PLE
K	P1	P1	P1	P1 or SC	P1 or SC	P1 or SC	P1 or SC	SC or W7P1 or W7P2 or W7PLE	SC or W7P1 or W7P2 or W7PLE

Summary of weighting in Waves 1–8

Weights for Wave 1 were calculated beginning with the inverse probability of selection for each child and then adjusting these weights to align to known population benchmarks (Soloff, Lawrence, Misson & Johnstone, 2006). A complex variant on the method of post-stratification was used whereby alignment was achieved for row-and-column totals of key benchmark demographics but not all cross-classified cells. This method has variously been termed incomplete post-stratification or calibration to marginal benchmarks, and is useful when complete post-stratification would subdivide the sample too finely and lead to model overfitting and large weight changes (Akaike, 1974). Benchmarks for children in the B and K cohorts for each state by capital city/rest of state area were drawn from the ABS Estimated Resident Population as at March 2004, and benchmarks for households by language spoken at home and mother's education level within each region were generated using proportions taken from the 2001 Census.

Weights for Waves 2–8 were calculated by adjusting previous wave weights for differential sample attrition in two stages (Cusack & Defina, 2014; Sipthorp & Daraganova, 2011; Sipthorp & Misson, 2007, 2009; Usback, 2018) At the first stage a modelled response propensity factor was applied; at the second, the weights were adjusted to preserve stratum totals. Extreme weights were capped as a form of outlier treatment to avoid any particular child contributing much more than other children in the sample to a weighted estimate, because this can potentially lead to volatile statistics if any such child has unusual characteristics.

In each wave, a population weight is calculated that adds up to the number of children in the population and in the corresponding age group and year, and a sample weight is calculated that adds up to the number of children in the sample for that wave. The population weight conceptually represents the number of children in the population represented by each child in the sample when creating weighted estimates. The sample weight can be used as a measure of the representativeness of each child compared to the others in the sample. The sample weights are equal to the population weights multiplied by the sampling fraction.

In Waves 2–4, weights were produced for every combination of response to individual waves. In Wave 5 this was simplified to a concise set of eight weights: each cohort has a longitudinal weight (both sample and population weights), and a cross-sectional weight (both sample and population weights). The longitudinal and cross-sectional weights are produced for different combinations of response:

- The **longitudinal weights** are defined for the sample responding to all waves up to and including the current wave, and involve an adjustment made for each new wave response. Longitudinal weights are most suitable for analysis that makes use of data from many time periods.
- The **cross-sectional weights** are defined for the sample responding only to the most recent wave, irrespective of responses to all or some of the intervening waves since Wave 1. Cross-sectional weights are most suitable for analysis that makes use only of the current data.

For more information on weighting methods from previous waves, please refer to the technical papers available on the *Growing Up in Australia* website.

Wave 9C1 weighting method

This section contains a brief description of the method used to create weights for Wave 9C1 data. The method is largely unchanged from Wave 8.

The weighting process for LSAC is in two stages. First, the response propensity modelling adjustment is applied to correct for attrition between waves. Second, the stratum adjustment is applied to re-align weight totals with known totals from the original sample. Both stages contribute to non-response bias reduction.

Longitudinal weights are calculated by taking the longitudinal weight from the previous wave of the study and adjusting for any additional non-response in the current wave.

For calculation of cross-sectional weights, the final weights used in Wave 1 are adjusted for all additional non-response in the current wave – regardless of whether a unit responded to Waves 2–8.

The main changes made to the Wave 9C1 weighting include:

- The logistic regression non-response model was updated for each of the four sets of weights, including updating variables for all models, and using Wave 8 data items for longitudinal models.
- Sample weight bounds were revised from [0.33, 3.5] to [0.28, 4.5] in light of a low response rate (due to COVID-19). Expanding the weights cap allows the proportion of capped units in previous waves to be consistent.

Initial weights

The final weights of a previous wave are carried forward to become the initial weights for the next wave.

- For Wave 9C1 longitudinal weights (which apply to those who have responded to all Waves 1, 2, 3, 4, 5, 6, 7, 8 and 9C1), the initial weight for children in Wave 9C1 is the final longitudinal corrected weight from Wave 8.
- For Wave 9C1 cross-sectional weights (which apply to all of those who responded in Wave 9C1), the initial weight for children in Wave 9C1 is the final weight from Wave 1.

Response propensity modelling

The purpose of this step is to adjust for differential non-response by particular demographic groups or individual characteristics that may have higher or lower sample attrition than average. This is done by modelling the response propensity using logistic regression (Little, 1986), using the dataset of respondents and non-respondents together, and using past wave survey responses as regressors. The modelled propensity is then used as a weight adjustment factor. For example, if a unit's response propensity is modelled at 90% then its response propensity adjusted weight is calculated as its initial weight divided by 0.9.

Selection of covariates for logistic regression non-response adjustment

The method for selection of covariates to use in the response propensity model is largely unchanged from Wave 8. A stepwise model selection process is used that considers all possible covariates for the response propensity model (list of covariates provided in Appendix E).

This stepwise process calculates the score chi-square statistics of covariates not in the model and adds the largest covariate not yet in the model. If any covariates are no longer found to be significant ($p < 0.05$) then they are removed from the model. This model selection process resulted in a shortlist of variables to consider adding to the Wave 9C1 model.

The response propensity model for Wave 9C1 was then re-run on the shortlisted variables together with the variables used for the derivation of Wave 8 weights. Variables that were no longer useful or significant ($p > 0.1$) were removed from the model. Variables used in the Wave 8 derivation that were still useful predictors in the Wave 9C1 model were maintained where possible to achieve consistency over time.

Wave 1 variables used in the B cohort cross-sectional weight model

- Parent 1 age
- Parent 2 age
- Mother's highest level of high school completed
- Mother's proficiency in spoken English
- Parent 1 self-completed questionnaire returned
- Parent 2 self-completed questionnaire returned
- Parent 1 renting home indicator
- Parent 1: Born in Australia
- SEIFA Education & Occupation (new)
- Number of siblings in household (new)
- Study child: Sex (new)

Wave 1 variables used in the K cohort cross-sectional weight model

- Parent 1 age
- Parent 2 age
- Parent 1 self-completed questionnaire returned
- Parent 2 self-completed questionnaire returned
- Parent 1 renting home indicator
- Mother: School completion (new)
- Parent 1: English as main language at home
- SEIFA Index of Education and Occupation
- Stratum (new)
- Study child: Sex (new)

Wave 8 variables used in the B cohort longitudinal weight model

- Parent 1: Study child overall school achievement
- Interviewed in November–February (Derived from fdatint)
- Mother: School completion (new)
- Parent 2 self-completed questionnaire returned
- SEIFA Index of Relative Socio-economic Advantage and Disadvantage
- Parent 1: Home ownership (new)
- Parent 1: Indigenous status (new)

Wave 8 variables used in the K cohort longitudinal weight model

- Parent 2 self-completed questionnaire returned
- SEIFA Index of Relative Socio-economic Advantage and Disadvantage
- Number of people in household (new)
- Number of siblings in household (new)
- SEIFA Economic Resources (new)
- Parent 1 Housing Tenure – (derived from jho06a, jho04a3a, jho04a3b, jho04a3c, jho04a3d, jho04a3e, jho04a3f) (new)
- Study child has 2 parents (new)
- Study child: Sex (new)

Model significance tests of the data items used in the above models can be found in Appendix C.

Odds ratio estimates for the data items used in the above models can be found in Appendix D.

A list of the variables considered in the selection of covariates for the response propensity models can be found in Appendix E.

Stratum weight adjustment

The purpose of this step is to use weighting to realign the sample composition within each stratum as at Wave 1, and to realign the sum of sample weights to be equal to the number of original participants in the first wave. The original selections were done by dividing each state into a capital city statistical division versus rest of state and then into groups of large or small postcodes. These are the original strata.

This adjustment accounts for some non-responses not already adjusted in the model, and ensures consistent estimates at the stratum level over time.

This stratum weight adjustment is also known as post-stratification or calibration to benchmarks. There is a separate adjustment factor calculated for each stratum based on the sum of the response propensity adjusted weights compared to the benchmark of the count of children within that stratum, subject to individual sample weights not exceeding the lower weight cap of 0.28 or the upper weight cap of 4.5. This process of calculating the weight adjustment for each unit to satisfy the benchmark specified while simultaneously satisfying the weight caps specified is achieved iteratively through the ABS SAS implementation of the generalised regression estimator (GREGWT).

In order to avoid larger adjustments of weight in strata with a small number of responding children, several strata were collapsed with other strata within the same state for the stratum weight adjustment

Weight capping

Weight capping is the process of limiting extreme values of weights for records that would otherwise have a large influence on estimates and calculations. Extreme weights can result from the logistic regression response propensity modelling step if a respondent's predicted chance of responding is very low, leading to a large weight adjustment. Weight capping is a robust form of automatic treatment of extreme values for weights, improving the variance characteristics of any analysis performed, at the expense of a slight reduction in contribution for some respondent groups (i.e. a slight risk of bias).

The weight caps are applied during the stratum weight adjustment step to ensure that any large response propensity adjusted weights are adjusted back to a reasonable level.

The lower cap of 0.28 and the upper cap of 4.5 are new for Wave 9C1. The sample weight bounds were updated in Wave 7 and designed to last a few waves, but an earlier update has been made because of the low response rates in Wave 9C1 (mainly due to COVID-19).

More detail on the number of units now appearing at the caps can be seen in Tables 6 and 7 in the next section of this paper.

Further characteristics of response across waves

Reacquisition of sample from previous waves

In this context, the reacquisition of sample refers to gaining a full response from a participant who was not considered responding in a previous wave. For the B cohort, out of 903 that did not respond to Wave 8, 118 responded to Wave 9C1. Out of the 2,385 that did not respond to at least one of Waves 2, 3, 4, 5, 6,7 or 8, 280 responded to Wave 9C1.

For the K cohort, out of 1,259 that did not respond to Wave 8, 70 responded to Wave 9C1. Out of the 2,473 that did not respond to at least one of Waves 2, 3, 4, 5, 6,7 or 8, 267 responded to Wave 9C1.

Table 3 shows those who have responded after previously being a 'non-responder' in a previous wave (sample reacquisition).

Table 3: Sample reacquisition for Waves 3, 4, 5, 6, 7, 8 and 9C1

Cohort	Resp. Wave 3, not Wave 2	Resp. Wave 4, not Wave 3	Resp. Wave 5, not Wave 4	Resp. Wave 6, not Wave 5	Resp. Wave 7, not Wave 6	Resp. Wave 8, not Wave 7	Resp. Wave 9C1, not Wave 8
B	133	135	129	89	124	134	118
K	135	119	94	77	120	301	70

For the K cohort, there were 70 units that responded to Wave 9C1 that did not respond to Wave 8. Of these 70 units, there were 67 units where the study child was interviewed. Refer to Table 2 for a definition of what is a considered a responding unit.

Total responding sample for each wave

The fully responding sample at each wave drives the calibration and hence the weighting process. See Tables 4 and 5 below for updated counts.

Table 4: Sample counts for the B cohort

Wave	1	2	3	4	5	6	7	8	9C1
Cross-sectional response	5,107	4,606	4,386	4,242	4,085	3,764	3,381	3,127	2,017
Longitudinal response	–	4,606	4,253	3,997	3,758	3,441	3,028	2,722	1,737
Cross-sectional attrition rate (%)	–	9.8	14.1	16.9	20.0	26.3	33.8	38.8	60.5
Longitudinal attrition rate (%)	–	9.8	7.7	6.0	6.0	8.4	12.0	10.1	36.2

Table 5: Sample counts for the K cohort

Wave	1	2	3	4	5	6	7	8	9C1
Cross-sectional response	4,983	4,464	4,331	4,169	3,956	3,537	3,089	3037	1,789
Longitudinal response	–	4,464	4,196	3,940	3,682	3,276	2,792	2510	1,522
Cross-sectional attrition rate (%)	–	10.4	13.1	16.3	20.6	29.0	38.0	39.1	64.1
Longitudinal attrition rate (%)	–	10.4	6.0	6.1	6.5	11.0	14.8	10.1	39.4

- Cross-sectional response – number of children who responded to that particular wave
- Longitudinal response – number of children who have responded to all waves up to and including that particular wave; that is, fully responding to each wave since Wave 1
- Cross-sectional attrition rate (%) – those not responding to that particular wave as a percentage of the Wave 1 cross-sectional response.
- Longitudinal attrition rate (%) – those not responding to the current wave, but having responded to all waves beforehand, as a percentage of the previous wave's longitudinal response.

Number of children with weight at cap

Tables 6 and 7 below show the number of children with a sample weight at the lower cap of 0.28 and upper cap of 4.5 by cohort and by type of weight.

For the B cohort, the number of units at the upper cap has decreased from 43 in Wave 8 to 34 for the cross-sectional weight, and increased from 20 in Wave 8 to 27 for the longitudinal weight.

Table 6: Counts of capped sample weights for Wave 9C1 – B cohort

State	Cross-sectional		Longitudinal	
	Lower cap (0.28)	Upper cap (4.5)	Lower cap (0.28)	Upper cap (4.5)
NSW	0	13	0	12
Vic.	0	10	0	4
Qld	9	6	2	7
SA	0	1	1	1
WA	2	2	2	2
Tas.	1	2	0	1
NT	12	0	9	0
ACT	0	0	0	0
Australia	**24**	**34**	**14**	**27**

For the K cohort, the number of units at the upper cap decreased from 25 in Wave 8 to 11 for the cross-sectional weight, and decreased from 11 in Wave 8 to 10 for the longitudinal weight.

Table 7: Counts of capped sample weights for Wave 8 – K cohort

State	Cross-sectional		Longitudinal	
	Lower cap (0.28)	Upper cap (4.5)	Lower cap (0.28)	Upper cap (4.5)
NSW	0	3	0	2
Vic.	0	3	0	4
Qld	8	5	2	3
SA	0	0	0	0
WA	0	0	2	1
Tas.	11	0	6	0
NT	4	0	7	0
ACT	1	0	2	0
Australia	**24**	**11**	**19**	**10**

Conclusion

Sample attrition has continued again in this wave; with the responding sample around 1,800 for the K cohort and around 2,000 for the B cohort. The longitudinal dataset presents a rich source of information about Australian children. The response propensity models identify which characteristics of the sample were related to their probability of response. The weights developed help to correct for different response patterns, allowing users to better analyse the data and draw more accurate conclusions about the population, being the two cohorts of children (the B cohort who were 0–1 year and the K cohort who were 4–5 years old during 2004, the Wave 1 recruitment year).

The weight capping ensures that no unit contributes too much or too little to any analysis using these data.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Australian Bureau of Statistics. (2013). *Australian Demographic Statistics*, Sep. 2012. Canberra: Australian Bureau of Statistics.

Australian Institute of Family Studies (Ed.). (2013). *The Longitudinal Study of Australian Children Annual Statistical Report 2012*. Melbourne: Australian Institute of Family Studies.

Bell, P. (2000). *Weighting and Standard Error Estimation for ABS Household Surveys*. Australian Bureau of Statistics Methodology Advisory Committee Paper. Canberra: Australian Bureau of Statistics.

Cusack, B., & Defina, R. (2014). *LSAC Technical Paper No. 10: Wave 5 weighting and non-response*. Melbourne: Australian Institute of Family Studies.

Engle, R. (1983). Wald, Likelihood Ratio, and Lagrange multiplier tests in Econometrics. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of Econometrics II* (pp. 796-801). Elsevier.

Fairclough, D. L. (2010). *Design and analysis of quality of life studies in clinical trials*. Boca Raton, FL: Chapman and Hall/CRC.

Holt, D. & Smith, T. M. F. (1979). Post-stratification. *Journal of the Royal Statistical Society Series A*, *142*, 33–46.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Little, R. J. A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, *54*, 139–157.

Pfeffermann, D. (1993). The role of sampling weights when modelling survey data. *International Statistical Review*, *61*, 317–337.

Sarndal, C. E., Swensson, B., & Wretman, J. H (1992). *Model assisted survey sampling*. New York: Springer-Verlag,

Sipthorp, M., & Misson, S. (2007). *LSAC Technical Paper No. 5: Wave 2 weighting and non-response*. Melbourne: Australian Institute of Family Studies.

Sipthorp, M., & Misson, S. (2009). *LSAC Technical Paper No. 6: Wave 3 weighting and non-response*. Melbourne: Australian Institute of Family Studies.

Sipthorp, S., & Daraganova, G. (2011). *LSAC Technical Paper No. 9: Wave 4 weights*. Melbourne: Australian Institute of Family Studies.

Soloff, C., Lawrence, D., & Johnstone, R. (2005). *LSAC Technical Paper No. 1: Sample design*. Melbourne: Australian Institute of Family Studies.

Soloff, C., Lawrence, D., Misson, S., & Johnstone, R. (2006). *LSAC Technical Paper No. 3: Wave 1 weighting and non-response*. Melbourne: Australian Institute of Family Studies.

Swets, J. A. (1973). The Relative Operating Characteristic in Psychology. *Science*, *182*, 990–1000.

Usback, S. (2018). *LSAC Technical Paper No. 20: Wave 7 weighting and non-response*. Melbourne: Australian Institute of Family Studies

Appendix A: Glossary of terms and abbreviations

This paper uses many technical terms, some of which are not consistently used across the fields of longitudinal studies and sample designs. We offer a brief glossary as a guide to how the terms are used in this paper.

Term	Definition
ABS	Australian Bureau of Statistics
Attrition	Process of sample size shrinking over time due to any mechanism
Cohort	Sample with a particular characteristic, e.g. B cohort ages 0–1 years in first wave
Coverage	Population represented by the remaining active participants
Cross-sectional	Pertaining to a statistic at one time point, typically broken down by characteristics at that time point
Design effect	Penalty factor to variance due to sample tending to be similar within selected postcode clusters
Estimation	Process of calculating a descriptive statistic from sample using weight, acknowledging the presence of sampling error
F2F	Face-to-face
Longitudinal	Pertaining to a statistic involving many time points, typically with a focus on evolution of participants over time
LSAC	Longitudinal Study of Australian Children
Missing data	Data absent either from non-response or partial response
Non-response	Failure to acquire survey response due to non-contact or refusal (opt-out)
P1	Parent 1, generally the child's mother
P2	Parent 2, the child's second parent
PLE	Parent Living Elsewhere
Partial response	Acquisition of data for some study modules but not others
Post-stratification	Process of dividing population into post-strata for weighting
Recruited sample	Subset of selected sample who agreed to participate in Wave 1
Response propensity	Chance that a particular individual or group will respond to a given wave
Respondent	or Participant or Active Participant: Any child (family) active in the study
Selected sample	Selection of children (families) approached at time of Wave 1 recruitment
Stratification	Process of dividing population into strata for selection
Stratum (Strata)	Cell(s) of population from which set number of children selected in sample
Study variable	Any variable collected in the study that data users wish to analyse
Weight	Value for a respondent to correct, up or down, for representativeness based on characteristics of responding sample

Appendix B: Description of Wave 9C1 weights

Table B1: Description of Wave 9C1 weights

SAS name	Cohort	Type	Waves cases responded to
i1weight	B	Population	1 & 9C1
l1weights	B	Sample	1 & 9C1
bcdefghi1wt	B	Population	1, 2, 3, 4, 5, 6, 7, 8 & 9C1
bcdefghi1wts	B	Sample	1, 2, 3, 4, 5, 6, 7, 8 & 9C1
k1weight	K	Population	1 & 9C1
k1weights	K	Sample	1 & 9C1
defghijk1wts	K	Population	1, 2, 3, 4, 5, 6, 7, 8 & 9C1
defghijk1wt	K	Sample	1, 2, 3, 4, 5, 6, 7, 8 & 9C1

Appendix C: Logistic regression models: type 3 analysis of effects

Note that where a response was not obtained to a variable, this was included in the model.

Table C1: B cohort – cross-sectional weights

Variable name	Description	DF[a]	Wald Chi-Square[b]	Pr > ChiSq
af03m2	Parent 1 age	1	17.5	<0.0001
af03m3	Parent 2 age	1	10.0	0.0016
aho04a3b	Parent 1 rents home	2	16.6	0.0003
afd08m1	Mother's highest level of schooling completed		95.4	<0.0001
afd11m2	Mother's proficiency in spoken English	5	17.3	0.0004
ap1scd	Parent 1 self-completed questionnaire returned	1	23.8	<0.0001
ap2scd	Parent 2 self-completed questionnaire returned	2	21.1	<0.0001
zf09m2	Parent 1 born in Australia	1	6.1	0.0137
acnfseo	SEIFA Education and Occupation	1	14.7	0.0001
ansib	Number of siblings in household	1	30.6	<0.0001
zf02m1	Study child: sex	1	12.2	0.0005

Notes: [a] Degrees of Freedom; [b] Wald Chi-Square is computed by squaring the ratio of the parameter estimate divided by its standard error estimate.

Table C2: B cohort – longitudinal weights

Variable name	Description	DF	Wald Chi-Square	Pr > ChiSq
hlc08a3a	Parent 1: Study Child overall school achievement	5	32.5	<0.0001
hfd08m1	Mother's highest level of schooling completed	5	15.0	0.0104
hcnfsad2	SEIFA Index of Relative Socio-Economic Advantage and Disadvantage	1	21.3	<0.0001
hp2scd	Parent 2 self-complete questionnaire returned	2	67.0	<0.0001
eoy	Interviewed in November–February (Derived from fdatint)	1	6.8	0.0092
hho04a1	Parent 1 home ownership	4	20.7	0.0004
zf12hp1	Parent 1 Indigenous Status	1	5.8	0.0162

Table C3: K cohort – cross-sectional weights

Variable name	Description	DF	Wald Chi-Square	Pr > ChiSq
cf03m2	Parent 1 age	1	10.3	0.0013
cf03m3	Parent 2 age	1	2.9	0.0898
cp1scd	Parent 1 self-completed questionnaire returned	1	13.5	0.0002
cp2scd	Parent 2 self-completed questionnaire returned	2	37.3	<0.0001
cho04a3b	Parent 1 rents home	2	16.9	0.0002
cfd08m1	Mother's highest level of schooling completed	7	77.5	<0.0001
cf11m2	Parent 1: English as main language at home	1	16.7	<0.0001
ccnfseo	SEIFA Index of Education and Occupation	1	25.2	<0.0001
stratum	Stratum	21	33.9	0.0371
zf02m1	Study child: sex	1	29.8	<0.0001

Table C4: K cohort – longitudinal weights

Variable name	Description	DF	Wald Chi-Square	Pr > ChiSq
jp2cati	Parent 2 self-completed questionnaire returned	1	29.9	<0.0001
jcnfsad2	SEIFA Index of Relative Socio-Economic Advantage and Disadvantage	1	12.3	0.0005
jnpeople	Number of people in household	1	25.9	<0.0001
jnsib	Number of siblings in household	1	19.1	<0,0001
jcnfser2	SEIFA Economic Resources	1	4.4	0.036
tenure	Parent 1 Housing tenure (derived from jho06a, jho04a3a, jho04a3b, jho04a3c, jho04a3d, jho04a3e, jho04a3f)	4	37.2	<0.0001
jp2	Study child has 2 parents	2	64.6	<0.0001
zf02m1	Study child: sex	1	22.8	<0.0001

Appendix D: Odds ratio estimates for variables in Wave 9C1 response propensity models

These odds ratios show different categories of variables included in the model.

Variable categories can be collapsed or re-parameterised as part of the weighting process. Small categories may have been collapsed during this process. Re-parameterisation is the re-labelling of modal categories. The odds ratios are calculated using the maximum category. By re-labelling the reference category to be the maximum makes the results easier to interpret. All the information needed about changes in categories is contained in the description column. Please do not compare these results against the Data Dictionary. These changes also apply to the tables in Appendix H.

Table D1: Odds ratio estimates for B cohort – cross-sectional weight

Effect	Description	Point estimate	95% Wald confidence interval	
acnfseo	SEIFA Education & Occupation	1.002	1.001	1.002
ansib	Number of siblings in household	0.833	0.781	0.889
af03m2	Parent 1 age	1.034	1.018	1.050
af03m3	Parent 2 age	1.023	1.009	1.037
afd08m1 6	**Mother: Highest level of schooling completed – Year 12 or equivalent – reference category**			
afd08m1 -2 vs 6	Mother completed – Don't know, not asked or refusal	0.525	0.091	3.034
afd08m1 2 vs 6	Mother completed Year 11 or equivalent	0.599	0.490	0.733
afd08m1 3 vs 6	Mother completed Year 10 or equivalent	0.523	0.438	0.625
afd08m1 4 vs 6	Mother completed Year 9 or equivalent	0.352	0.220	0.563
afd08m1 5 vs 6	Mother completed Year 8 or below or never attended school	0.090	0.032	0.256
aho04a3b -4 vs 2	Refusal or Don't know vs **Parent 1 does not rent home**	1.030	0.139	7.625
aho04a3b 1 vs 2	Parent 1 Rents home vs **Parent 1 does not rent home**	0.736	0.635	0.853
afd11m2 5	**Mother's proficiency in spoken English – Not asked – reference category**			
afd11m2 -2 vs 5	Mother's proficiency in spoken English – Don't know or refused	0.464	0.085	2.532
afd11m2 1 vs 5	Mother speaks English Very Well	0.665	0.528	0.837
afd11m2 2 vs 5	Mother speaks English Well	0.719	0.485	1.066
afd11m2 3 vs 5	Mother speaks English Not Well	1.380	0.819	2.327
afd11m2 4 vs 5	Mother speaks English Not at all	0.831	0.259	2.670
ap1scd 0 vs 1	Parent 1 did not return self-completed questionnaire vs **Parent 1 did return self-completed questionnaire**	0.501	0.380	0.662
ap2scd 1	**Parent 2 self-complete questionnaire returned – reference category**			
ap2scd -9 vs 1	No Parent 2 in household	0.862	0.513	1.449
ap2scd 0 vs 1	Parent 2 did not return self-completed questionnaire	0.575	0.454	0.728
zf02m1 1 vs 2	Study child sex male vs **female**	0.807	0.715	0.910
zf09m2 0 vs 1101	Parent 1: Born in Australia vs **Parent 1: Born elsewhere**	0.799	0.669	0.955

Table D2: Odds ratio estimates for B cohort – longitudinal weight

Effect	Description	Point estimate	95% Wald confidence interval	
hcnfsad2	SEIFA Index of Relative Socio-economic Advantage and Disadvantage	1.003	1.002	1.004
hlc08a3a 6	**Parent 1: Study Child overall school achievement average – reference category**			
hlc08a3a -9 vs 6	Parent 1: Study Child overall school achievement not asked or don't know	0.493	0.248	0.980
hlc08a3a 1 vs 6	Parent 1: Study Child overall school achievement excellent	1.603	1.260	2.037
hlc08a3a 2 vs 6	Parent 1: Study Child overall school achievement above average	1.319	1.084	1.605
hlc08a3a 4 vs 6	Parent 1: Study Child overall school achievement below average	0.738	0.532	1.026
hlc08a3a 5 vs 6	Parent 1: Study Child overall school achievement well below average	0.855	0.476	1.536
zf12hp1 1 vs 2	Parent 1 not Aboriginal or Torres Strait Islander or not asked **vs Parent 1 Aboriginal or Torres Strait Islander**	2.744	1.205	6.249
hp2scd 1	**Parent 2 self-complete questionnaire returned – reference category**			
hp2scd -9 vs 1	No parent 2 in household	0.800	0.625	1.024
hp2scd 0 vs 1	Parent 2 did not return self-completed questionnaire	0.433	0.354	0.529
hfd08m1 6	**Mother School completion Year 12 or equivalent – reference category**			
hfd08m1 -9 vs 6	Mother School completion Don't know or not asked	0.811	0.426	1.545
hfd08m1 2 vs 6	Mother School completion Year 11 or equivalent	0.931	0.695	1.247
hfd08m1 3 vs 6	Mother School completion Year 10 or equivalent	0.694	0.541	0.889
hfd08m1 4 vs 6	Mother School completion Year 9 or equivalent	0.616	0.317	1.196
hfd08m1 5 vs 6	Mother School completion Year 8 or below or never attended school	0.238	0.074	0.770
eoy 0 vs 1	Not interviewed in November–February vs **Interviewed in November–February** (derived from fdatint)	1.481	1.102	1.991
hho04a1 7	**Parent 1 home ownership – being paid off by you (and/or your partner) – reference category**			
hho04a1 -9 vs 7	Parent 1 home ownership not asked	1.253	0.704	2.230
hho04a1 2 vs 7	Parent 1 home ownership – owner outright by you (and/or your partner)	1.537	1.214	1.945
hho04a1 3 vs 7	Parent 1 home ownership – rented by you (and/or your partner)	0.793	0.627	1.004
hho04a1 6 vs 7	Parent 1 home ownership – being purchased under a rent/buy scheme by you (and/or your partner)	0.779	0.313	1.938

Table D3: Odds ratio estimates for K cohort – cross-sectional weight

Effect	Description	Point estimate	95% Wald confidence interval	
cf03m2	Parent 1 age	1.024	1.009	1.039
cf03m3	Parent 2 age	1.012	0.998	1.026
cf11m2 0 vs 1201	Parent 1: English as main language at home vs **Parent 1: Main language at home Non-English**	0.675	0.559	0.815
cp1scd 0 vs 1	Parent 1 did not return self-completed questionnaire vs **Parent 1 did return self-completed questionnaire**	0.599	0.455	0.788
cp2scd 1	**Parent 2 did return self-completed questionnaire – reference category**			
cp2scd -9 vs 1	No Parent 2 in household	0.716	0.415	1.237
cp2scd 0 vs 1	Parent 2 did not return self-completed questionnaire	0.473	0.372	0.603
cho04a3b 2	**Parent 1 does not rent home – reference category**			
cho04a3b -2 vs 2	Refusal or don't know if Parent 1 rents home	0.935	0.147	5.961
cho04a3b 1 vs 2	Parent 1 rents home	0.718	0.612	0.841
cfd08m1 7	**Mother school completion Year 12 or equivalent – reference category**			
cfd08m1 -9 vs 7	Mother's school completion not asked	0.616	0.260	1.463
cfd08m1 -4 vs 7	Mother's school completion refused or don't know	0.659	0.094	4.603
cfd08m1 2 vs 7	Mother's school completion Year 11 or equivalent	0.744	0.612	0.904
cfd08m1 3 vs 7	Mother's school completion Year 10 or equivalent	0.509	0.429	0.603
cfd08m1 4 vs 7	Mother's school completion Year 9 or equivalent	0.400	0.271	0.591
ccnfseo	SEIFA Index of Education and Occupation	1.002	1.001	1.003
stratum 81	**ACT – reference category**			
stratum 11 vs 81	NSW Met	0.658	0.433	1.000
stratum 13 vs 81	NSW Xmet large	0.875	0.559	1.372
stratum 14 vs 81	NSW Xmet small	0.460	0.219	0.965
stratum 21 vs 81	Vic. Met large	0.681	0.446	1.041
stratum 22 vs 81	Vic. Met small	0.473	0.158	1.419
stratum 23 vs 81	Vic. Xmet large	0.762	0.466	1.246
stratum 24 vs 81	Vic. Xmet small	0.640	0.345	1.187
stratum 31 vs 81	Qld Met	0.692	0.438	1.095
stratum 33 vs 81	Qld Xmet large	0.503	0.317	0.799
stratum 34 vs 81	Qld Xmet small	0.826	0.428	1.592
stratum 41 vs 81	SA Met large	0.813	0.501	1.319
stratum 43 vs 81	SA Xmet large	0.672	0.337	1.342
stratum 44 vs 81	SA Xmet small	0.540	0.182	1.603
stratum 51 vs 81	WA Met large	0.729	0.459	1.156
stratum 52 vs 81	WA Met small	0.646	0.208	2.001
stratum 53 vs 81	WA Xmet large	1.022	0.559	1.870
stratum 54 vs 81	WA Xmet small	0.573	0.266	1.234
stratum 61 vs 81	Tas. Met	0.972	0.487	1.943
stratum 63 vs 81	Tas. Xmet	1.147	0.614	2.143
stratum 71 vs 81	NT Met	1.039	0.485	2.227
stratum 73 vs 81	NT Xmet small and NT Xmet large	0.691	0.306	1.562
zf02m1 1 vs 2	Study child sex male vs **female**	0.709	0.626	0.802

Table D4: Odds ratio estimates for K cohort – longitudinal weight

Effect	Description	Point estimate	95% Wald confidence interval	
jcnfsad2	SEIFA Index of Relative Socio-Economic Advantage and Disadvantage	1.003	1.001	1.005
jcnfser2	SEIFA Economic Resources	0.998	0.996	1.000
jnpeople	Number of people in household	0.702	0.613	0.805
jnsib	Number of siblings in household	1.437	1.222	1.691
tenure 8	**Tenure – Parent 1: Owning home on a mortgage – reference category**			
tenure 1 vs 8	Parent 1: Owning home outright	1.342	1.082	1.664
tenure 3 vs 8	Parent 1: Renting	0.873	0.669	1.138
tenure 6 vs 8	Parent 1: Occupied rent free	1.600	0.793	3.227
tenure 7 vs 8	Parent 1: Other tenure types	0.515	0.385	0.689
jp2cati 0 vs 1	Parent 2: Self-complete questionnaire missing vs **returned**	0.573	0.469	0.700
jp2 1	**Study child has 2 parents in the home – reference category**			
jp2 -9 vs 1	Study child not asked if 2 parents in the home	0.122	0.071	0.207
jp2 0 vs 1	Study child does not have 2 parents in the home	0.756	0.583	0.982
zf02m1 1 vs 2	Study child sex: male vs **female**	0.662	0.559	0.784

Appendix E: Data items considered for response propensity models

Table E1: Wave 1 data items considered for B cohort – cross-sectional weight

Variable name	Variable label
acnfsad	0/1—Home—SEIFA Advantage/Disadvantage
acnfseo	0/1—Home—SEIFA Education & Occupation
acnfser	0/1—Home—SEIFA Economic Resources
af01am	0/1—M@0/1—Present for wave
af01m3	0/1—P2@W1—Present for wave
af03m2	0/1—P1@W1—F2F A4—Age
af03m3	0/1—P2@W1—F2F A4—Age
af11am	0/1—M@0/1—F2F A12—Main language spoken at home
af11m1	0/1—SC—F2F A12—Main language spoken at home
af11m2	0/1—P1@W1—F2F A12—Main language spoken at home
afd08a1	0/1—P1—F2F H3—School completion
afd08m1	0/1—M—F2F H3—School completion
afd11m2	0/1—M—F2F H10—Proficiency in spoken English
aho04a3b	0/1—P1—F2F L4—Rent home
aho04a5	0/1—P1—F2F L5—Housing tenure
aho09a1a1	0/1—P1—F2F L11—Safe neighbourhood
anpeople	0/1—No. people in household
ansib	0/1—No. siblings of SC in household
ap1scd	0/1—Parent 1 self-completed data present
ap2	0/1—SC has 2 parents in the home
ap2scd	0/1—Parent 2 self-completed data present
zf02m2	P1@W1—F2F A3—Sex
zf09m2	P1@W1—F2F A10—Country of birth
zf12m1	SC—F2F A13—Indigenous status
zf12m2	P1@W1—F2F A13—Indigenous status
zf02m1	SC - F2F A3 - Sex

Table E2: Wave 7 data items considered for B cohort – longitudinal weight

Variable name	Variable label
hcnfsad2	14/15 - SEIFA - Index of Relative Socio-Economic Advantage and Disadvantage - 2016 - SA2 - Score
hcnfsad2d	14/15 - SEIFA - Index of Relative Socio-Economic Advantage and Disadvantage - 2016 - SA2 - Score - Deciles - National
hcnfser2	14/15 - Home - SEIFA Economic Resources - 2016 - SA2 - Score
hcnfser2d	14/15 - Home - SEIFA Economic Resources - 2016 - SA2 - Deciles - National
hf01hm	M@14/15 - Present for wave
hf03hp1	P1@14/15 - Age
hf03hp2	P2@14/15 - Age
hf11hm	M@14/15 - Language other than English spoken at home
hf11hp1	P1@14/15 - Language other than English spoken at home
hf11m1	14/15 - SC - Language other than English spoken at home
hfd08a1	14/15 - P1 - P1 CAI A1.1/A1.3+W1-6 - School completion
hfd08a2a	14/15 - P1 - P1 CAI A1.2/A1.3+W1-6 - Finished other post-secondary qualification
hfd08a3a	14/15 - P1 - P1 CAI A1.3+W1-6 - Highest qualification
hfd08m1	14/15 - M - P1CAI A1.1/A1.3+W1-6 - School completion
hfemp	14/15 - F - Employment status
hho04a1	14/15 - P1 CAI I1.6-1.8 - Home ownership
hho04a3b	14/15 - P1 CAI I1.6.2 - Rent home
hho04a5	14/15 - P1 CAI I1.6-1.8 - Housing tenure
hlc08a3a	14/15 - P1 - P1 CAI C7.5 - Overall school achievement
hmemp	14/15 - M - employment status
hnpeople	14/15 - No. people in household
hnsib	14/15 - No. siblings of SC in household
hp2	14/15 - SC has 2 parents in the home
hp2scd	14/15 - Parent 2 Self-complete data present
zf02hp1	P1@14/15 - Sex
zf09hp1	P1@14/15 - Country of birth
zf12hp1	P1@14/15 - Indigenous status
hhe11a3e	14/15 - P1 - P1 CAI C6.2 - How often help child with homework
hhe09a1	14/15 - P1 CAI F8.4 - Extra-curricular - Any
hdatint	14/15 - Date of interview
hf01hp2	P2@14/15 - Present for wave

Table E3: Wave 1 data items considered for K cohort – cross-sectional weight

Variable name	Variable label
caangb	4/5—P1—Angry parenting (v3)
cahact	4/5—P1—Home activities index
ccnfsad	4/5—Home—SEIFA Advantage/Disadvantage
ccnfseo	4/5—Home—SEIFA Education & Occupation
ccnfser	4/5—Home—SEIFA Economic Resources
cf01cm	4/5—M@4/5—Present for wave
cf01m3	4/5—P2@W1—Present for wave
cf03m2	4/5—P1@W1—F2F A4—Age
cf03m3	4/5—P2@W1—F2F A4—Age
cf11cm	4/5—M@4/5—F2F A12—Main language spoken at home
cf11m1	4/5—SC—F2F A12—Main language spoken at home
cf11m2	4/5—P1@W1—F2F A12—Main language spoken at home
cfd08a1	4/5—P1—F2F H3—School completion
cfd08m1	4/5—M—F2F H3—School completion
cfd11m2	4/5—M—F2F H10—Proficiency in spoken English
cho04a3b	4/5—P1—F2F L4—Rent home
cho04a5	4/5—P1—F2F L5—Housing tenure
cho09a1a1	4/5—P1—F2F L11—Safe neighbourhood
cnpeople	4/5—No. people in household
cnsib	4/5—No. siblings of SC in household
cp1scd	4/5—Parent 1 self-completed data present
cp2	4/5—SC has 2 parents in the home
cp2scd	4/5—Parent 2 self-complete data present
zf02m2	P1@W1—F2F A3—Sex
zf09m2	P1@W1—F2F A10—Country of birth
zf12m1	SC—F2F A13—Indigenous status
zf12m2	P1@W1—F2F A13—Indigenous status
Stratum	Stratum
zf02m1	SC - F2F A3 - Sex

Table E4: Wave 8 data items considered for K cohort – longitudinal weight

Variable name	Variable label
jcnfsad2	18/19 - SEIFA - Index of Relative Socio-Economic Advantage and Disadvantage - 2016 - SA2 - Score
jcnfsad2d	18/19 - SEIFA - Index of Relative Socio-Economic Advantage and Disadvantage - 2016 - SA2 - Score - Deciles - National
jcnfser2	18/19 - Home - SEIFA Economic Resources - 2016 - SA2 - Score
jcnfser2d	18/19 - Home - SEIFA Economic Resources - 2016 - SA2 - Deciles - National
jf01jm	M@18/19 - Present for wave
jf03jp1	P1@18/19 - Age
jf03jp2	P2@18/19 - Age
jf11jm	M@18/19 - Language other than English spoken at home
jf11jp1	P1@18/19 - Language other than English spoken at home
jf11m1	18/19 - SC - Language other than English spoken at home
jfd08a3a	P1@18/19 - Highest qualification
jfd08m3a	M@18/19 - Highest qualification
stratum	18/19 - Stratum
jfemp	18/19 - F - Employment status
jf01jp2	P2@18/19 - Present for wave
jho04a3a	P1@18/19 - Owns home
jho06a	P1@18/19 - Making payments on mortgages
jho04a3b	18/19 - P1 CAI I1.6.2 - Rent home
jho04a3e	P1@18/19 - Rent or board home
jho04a3c	P1@18/19 - Rent/buy scheme home
jho04a3d	P1@18/19 - Life tenure scheme home
jho04a3f	P1@18/19 - Rent free home
jmemp	18/19 - M - employment status
jnpeople	18/19 - No. people in household
jnsib	18/19 - No. siblings of SC in household
jp2	18/19 - SC has 2 parents in the home
jp2cati	18/19 - Parent 2 Self-complete data present
zf02jp1	P1@18/19 - Sex
zf09jp1	P1@18/19 - Country of birth
zf12jp1	P1@18/19 - Indigenous status
zf02m1	18/19 - SC - Sex

Appendix F: Distributional checks of non-response modelling

In order to validate the logistic regression non-response adjustment procedure, the estimated response propensities have been plotted below. There are also plots of the final sample weight under each model, where the approximate proportion of units at the caps can be observed.

B cohort – cross-sectional weight

Figure F1: Distribution of estimated response propensities – B cohort cross-sectional weight
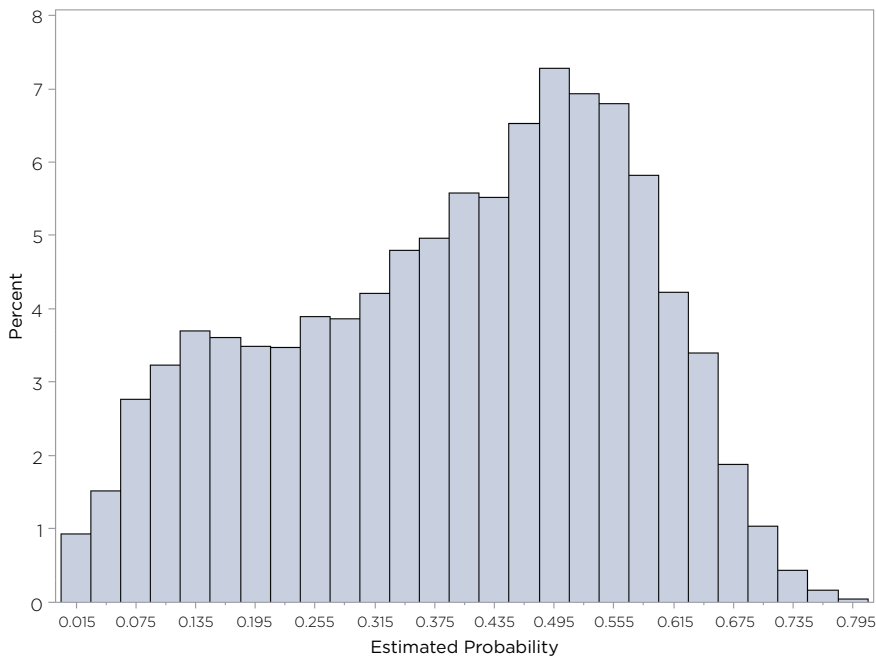
Table F1: Analysis variable: estimated probability – B cohort cross-sectional weight

Mean	Std Dev	Minimum	Maximum	Mode	Range	Sum	N
0.394948	0.17507	0.006151	0.798634	0.44285	0.79248	2,017	5,107

Figure F2: Distribution of final sample weight for Wave 9C1 – B cohort cross-sectional weight
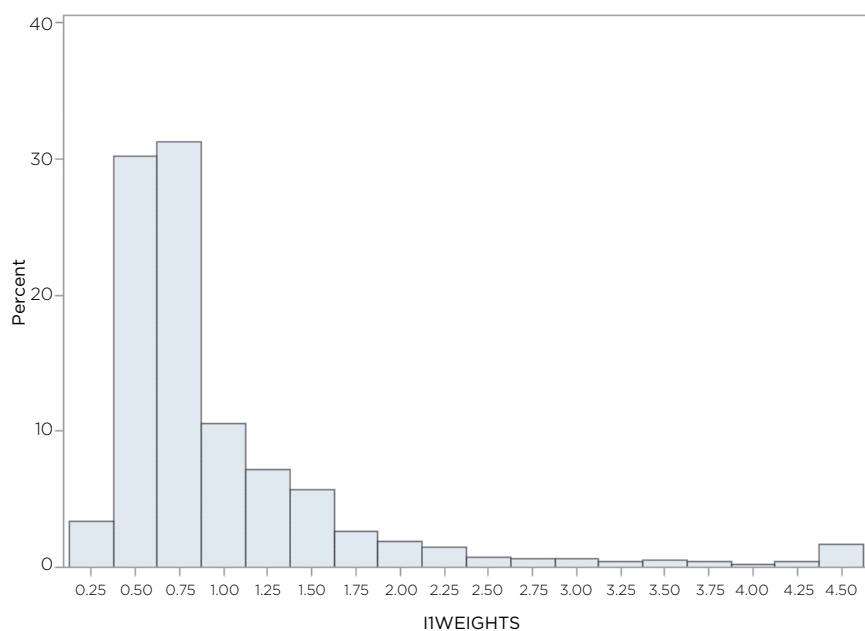
Table F2: Analysis variable: IWEIGHTS – B cohort cross-sectional weight

Mean	Std Dev	Minimum	Maximum	Mode	Range	Sum	N
1.00000	0.776998	0.28	4.5	4.5	4.22	2,017	2,017

B cohort – longitudinal weight

Figure F3: Distribution of estimated response propensities – B cohort longitudinal weight
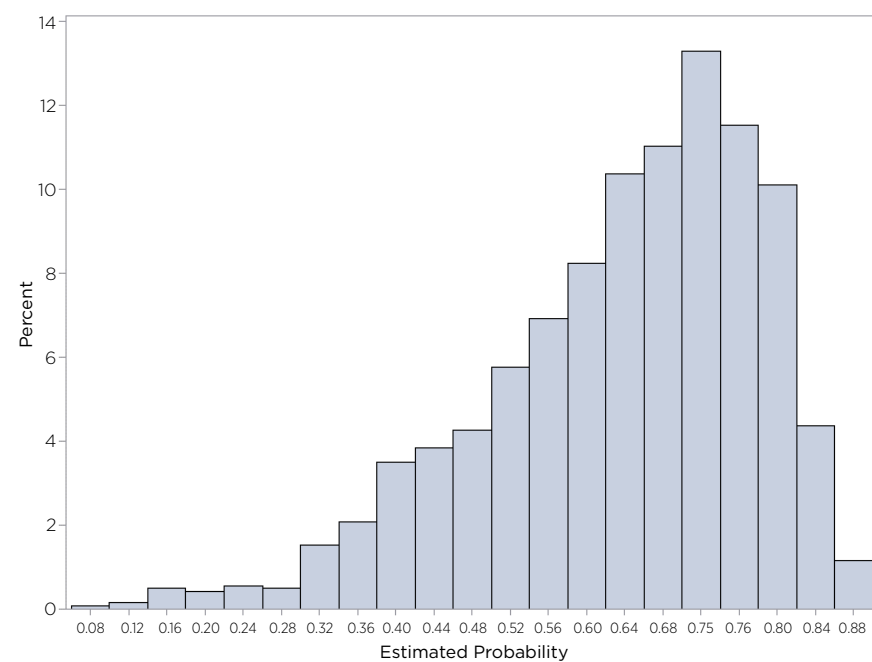
Table F3: Analysis variable: estimated probability – B cohort longitudinal weight

Mean	Std Dev	Minimum	Maximum	Mode	Range	Sum	N
0.638134	0.144022	0.060611	0.885169	0.783674	0.82456	1,737	2,722

Figure F4: Distribution of final sample weight for Wave 9C1 – B cohort longitudinal weight
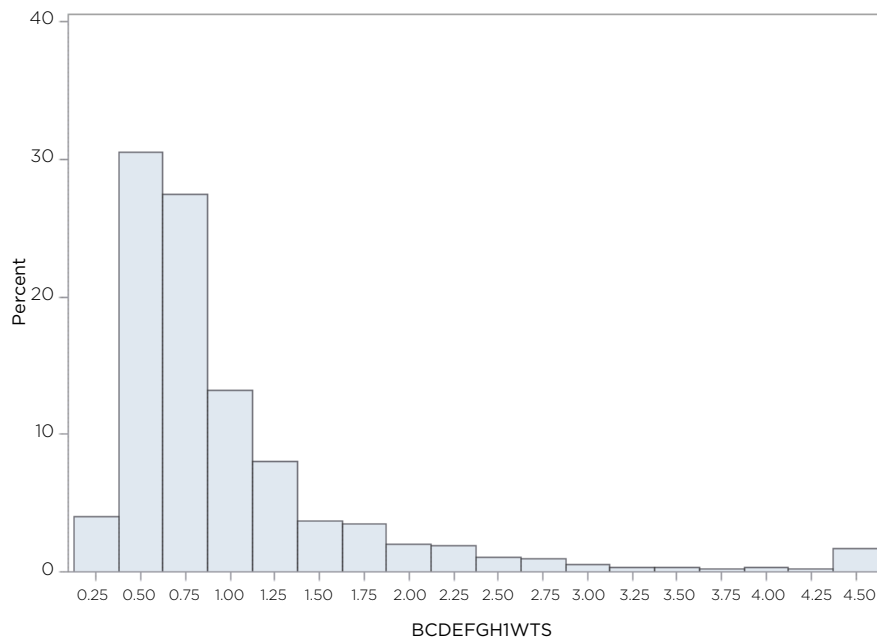
Table F4: Analysis variable: BCDEFGHI1WTS – B cohort longitudinal weight

Mean	Std Dev	Minimum	Maximum	Mode	Range	Sum	*N*
1.000000	0.760795	0.28	4.5	4.5	4.22	1,737	1,737

K cohort – cross-sectional weight

Figure F5: Distribution of estimated response propensities – K cohort cross-sectional weight
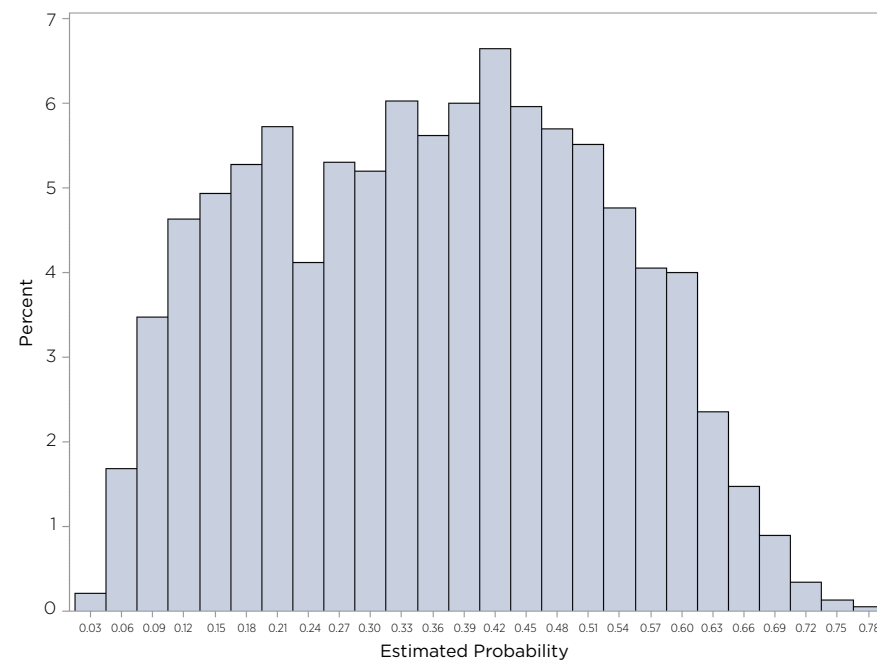
Table F5: Analysis variable: estimated probability – K cohort cross-sectional weight

Mean	Std Dev	Minimum	Maximum	Mode	Range	Sum	*N*
0.359028	0.162609	0.028552	0.783336	0.111396	0.75478	1,789	4,983

Figure F6: Distribution of final sample weight for Wave 9C1 – K cohort cross-sectional weight
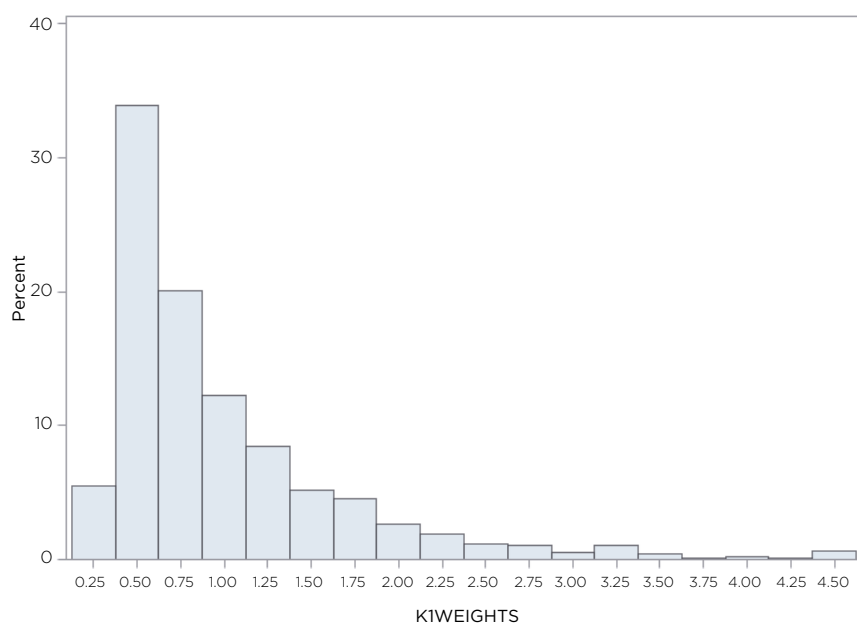
Table F6: Analysis variable: IWEIGHTS – K cohort cross-sectional weight

Mean	Std Dev	Minimum	Maximum	Mode	Range	Sum	N
1.000000	0.719115	0.28	4.5	0.28	4.22	1,789	1,789

K cohort – longitudinal weight

Figure F7: Distribution of estimated response propensities – K cohort longitudinal weight
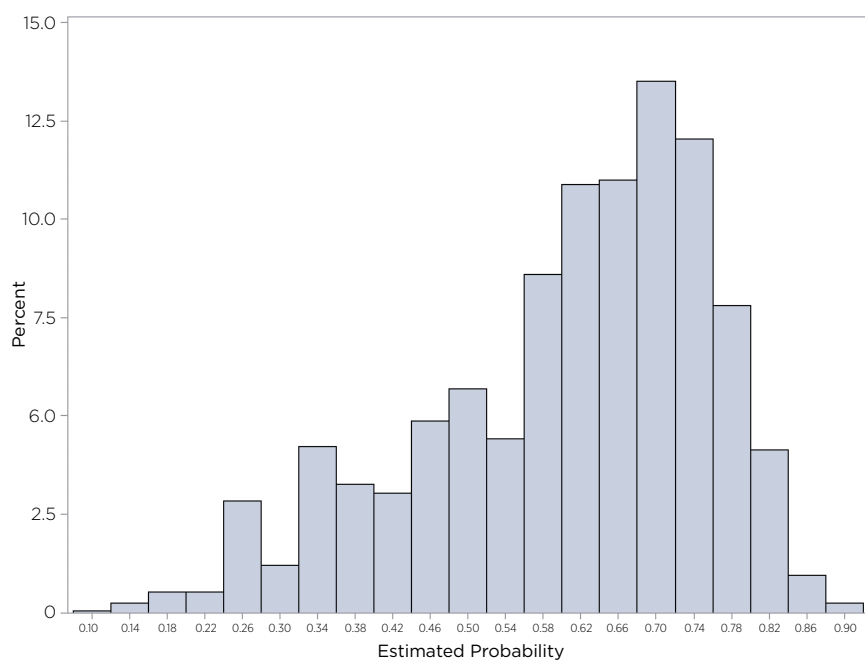
Table F7: Analysis variable: estimated probability – K cohort longitudinal weight

Mean	Std Dev	Minimum	Maximum	Mode	Range	Sum	N
0.606374	0.151908	0.103798	0.905156	0.267204	0.80136	1,522	2,510

Figure F8: Distribution of final sample weight for Wave 9C1 – K cohort longitudinal weight
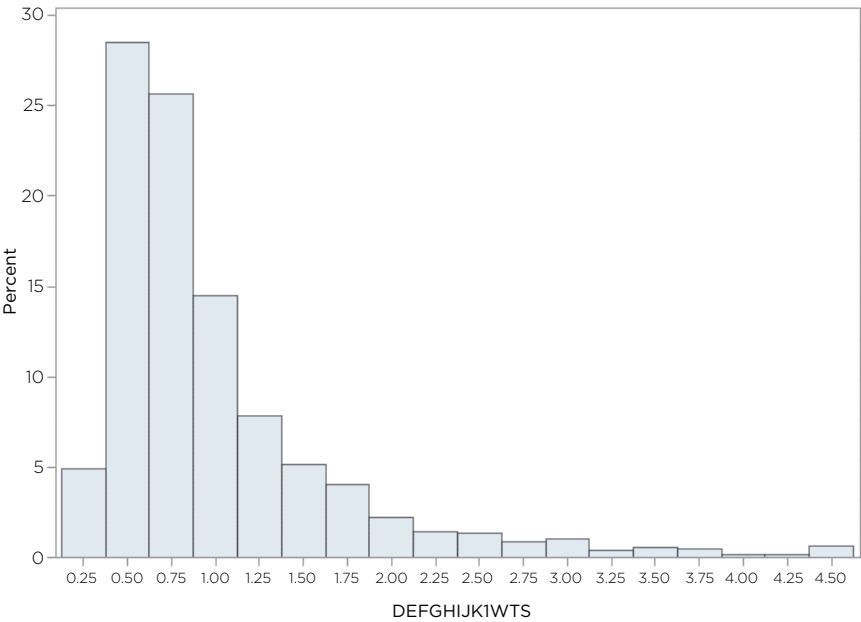
Table F8: Analysis variable: DEFGHIWTS – K cohort longitudinal weight

Mean	Std Dev	Minimum	Maximum	Mode	Range	Sum	*N*
1.0000000	0.707069	0.28	4.5	0.28	4.22	1,522	1,522

Appendix G: Non-response to instruments

Table G1: Non-response to instruments

	Eligible	Responding	% Wave 1	Response rate %
		B cohort		
Wave 8 (issued sample = 3,904)				
Interview	3,127	3,127	61.2	100.0
P1CASI	3,123	3,086	60.4	98.8
P2SC	2,575	1,854	na	72.0
PLECATI	521	319	na	61.2
TEACH	3,060	2,317	na	75.7
ACASI	3,036	3,011	59.0	99.2
TUD	3,036	2,827	55.4	93.1
Wave 9C1 (issued sample = 3,849)				
SC CAWI	3,849	1,595	31.2	41.4
P1 CAWI	3,844	1,296	25.4	33.7
P2 CAWI	2,542	770	na	30.3
PLE CAWI	442	130	na	29.4
		K cohort		
Wave 8 (issued sample = 3,943)				
YP or P Interview	3,037	3,037	60.9	100.0
YP Interview	3,037	2,708	54.3	89.2
CASI	2,708	2,656	53.3	98.1
CAWI	2,708	1,908	38.3	70.5
CAWSI	800	596	na	74.5
PCATI - W7P1	3,207	2,635	na	82.2
PCATI - W7P2	2,057	1,681	na	81.7
PCATI - W7PLE	473	317	na	67.0
Wave 9C1 (issued sample = 3,809)				
SC CAWI	3,809	1,361	27.3	35.7
P1 CAWI (W7P1)	3,110	975	19.6	31.4
P2 CAWI (W7P2)	2,020	541	na	26.8
PLE CAWI (W7PLE)	413	110	na	26.6

For Appendix G, the issued sample numbers are for representations only, not for direct comparisons. This is due to a change in methodology for Wave 9C1.

Instrument	Description
CAWI	Computer Assisted Web Interview
P1/P2/PLE CATI	Parent Computer Assisted Telephone Interview
CAWSI	Computer Assisted Web Self Interview
EHC	Event History Calendar
CAI	Computer Assisted Interview
P1CASI	Parent 1 Computer Assisted Self Interview
P2SC	Parent 2 Self-Complete Questionnaire
PLECATI	Parent Living Elsewhere Computer Assisted Telephone Interview
Teach	Teacher Questionnaire
ACASI	Audio-Computer Assisted Self Interview
CSR	Child Self Report
TUD	Time Use Diary
MR	Matrix Reasoning
EXEC	Executive Functioning (CogState)
GJA	Rice Test of Grammatical Judgement
na	Not appropriate to compare with Wave 1

Appendix H: Joint distributions of response status and each predictor

These tables show how each variable in the response propensity model look within the responding sample versus the non-responding households. The tables show how each categorical variable is distributed between the responding and non-responding households. Each tab finishes with a single table for all the continuous variables showing the mean and standard deviation of each variable in the responding and non-responding households.

Variable categories can be collapsed or re-parameterised as part of the weighting process. Small categories may have been collapsed during this process. Re-parameterisation is the re-labelling of modal categories. The odds ratios are calculated using the maximum category. By re-labelling the reference category to be the maximum makes the results easier to interpret. All the information needed about changes in categories is contained in the tables. Please do not compare these results against the Data Dictionary.

B cross-sectional

		Highest year of primary or secondary school completed by mother – afd08m1						
		Don't know, not asked, Refusal	Yr 11 or equivalent	Yr 10 or equivalent	Yr 9 or equivalent	Yr 8 or below, never attended school	Year 12 or equivalent	Total
Non-responding	Freq	7	400	628	143	84	1,828	3,090
	%	0.14	7.83	12.30	2.80	1.64	35.79	60.51
Responding	Freq	2	175	231	23	4	1,582	2,017
	%	0.04	3.43	4.52	0.45	0.08	30.98	39.49
Total	Freq	9	575	859	166	88	3,410	5,107
	%	0.18	11.26	16.82	3.25	1.72	66.77	100.00

		(Is this house/flat/unit, etc.) rented by you (and/or your partner) – aho04a3b			
		Refusal or don't know	Yes	No	Total
Non-responding	Freq	5	1,072	2,013	3,090
	%	0.10	20.99	39.42	60.51
Responding	Freq	2	401	1,614	2,017
	%	0.04	7.85	31.60	39.49
Total	Freq	7	1,473	3,627	5,107
	%	0.14	28.84	71.02	100.00

		Parent 1 Self-complete data added to file – ap1scd		
		No	Yes	Total
Non-responding	Freq	638	2,452	3,090
	%	12.49	48.01	60.51
Responding	Freq	128	1,889	2,017
	%	2.51	36.99	39.49
Total	Freq	766	4,341	5,107
	%	15.0	85.0	100.00

		Parent 2 Self-complete data added to file – ap2scd			
		Not asked	No	Yes	Total
Non-responding	Freq	401	730	1,959	3,090
	%	7.85	14.29	38.36	60.51
Responding	Freq	76	204	1,737	2,017
	%	1.49	3.99	34.01	39.49
Total	Freq	477	934	3,696	5,107
	%	9.34	18.29	72.37	100.00

		Parent 1 Country of birth – zf09m2		
		Elsewhere	Australia	Total
Non-responding	Freq	718	2,372	3,090
	%	14.06	46.45	60.51
Responding	Freq	393	1,624	2,017
	%	7.70	31.80	39.49
Total	Freq	1111	3,996	5,107
	%	21.75	78.25	100.00

		Study child sex – zf02m1		
		Male	Female	Total
Non-responding	Freq	1,630	1,460	3,090
	%	31.92	28.59	60.51
Responding	Freq	978	1,039	2,017
	%	19.15	20.34	39.49
Total	Freq	2,608	2,499	5,107
	%	51.07	48.93	100.00

		Mother's proficiency in spoken English – afd11m2						
		Don't know or refused	Very well	Well	Not well	Not at all	Not asked	Total
Non-responding	Freq	7	327	114	66	21	2,555	3,090
	%	0.14	6.40	2.23	1.29	0.41	50.03	60.51
Responding	Freq	2	161	48	30	4	1,772	2,017
	%	0.04	3.15	0.94	0.59	0.08	34.70	39.49
Total	Freq	9	488	162	96	25	4,327	5,107
	%	0.18	9.56	3.17	1.88	0.49	84.73	100.00

	Variable	Label	Observations	Mean	SD	Minimum	Maximum
Non-responding	acnfseo	0/1 – Home - SEIFA Education & Occupation	3,090	987.91	76.44	783.20	1,222.48
	af03m2	0/1 - P1@W1 - F2F A4 - Age	3,090	30.30	5.76	0	63
	af03m3	0/1 - P2@W1 - F2F A4 - Age	3,090	28.84	12.56	0	65
	ansib	0/1 - No. siblings of SC in household	3,090	1.03	1.13	0	10
Responding	acnfseo	0/1 - Home - SEIFA Education & Occupation	2,017	1,009.28	82.24	690.24	1,221.12
	af03m2	0/1 - P1@W1 - F2F A4 - Age	2,017	32.08	4.95	18	55
	af03m3	0/1 - P2@W1 - F2F A4 - Age	2,017	33.20	8.63	0	62
	ansib	0/1 - No. siblings of SC in household	2,017	0.86	0.95	0	10

B longitudinal

		Highest year of primary or secondary school completed by mother - hfd08m1						
		Don't know or not asked	Yr 11 or equivalent	Yr 10 or equivalent	Yr 9 or equivalent	Yr 8 or equivalent or never attended school	Year 12 or equivalent	Total
Non-responding	Freq	20	93	158	22	14	678	985
	%	0.73	3.42	5.8	0.81	0.51	24.91	36.19
Responding	Freq	24	141	176	18	4	1,374	1,737
	%	0.88	5.18	6.47	0.66	0.15	50.48	63.81
Total	Freq	44	234	334	40	18	2,052	2,722
	%	1.62	8.6	12.27	1.47	0.66	75.39	100.00

		Home ownership – hho04a1					
		Not asked	Owner outright by you (and/or your partner)	Rented by you (and/or your partner)	Being purchased under a rent/ buy scheme by you (and/or your partner)	Being paid off by you (and/or your partner)	Total
Non-responding	Freq	33	121	203	11	617	985
	%	1.21	4.45	7.46	0.40	22.67	36.19
Responding	Freq	38	368	228	11	1,092	1,737
	%	1.40	13.52	8.38	0.40	40.12	63.81
Total	Freq	71	489	431	22	1,709	2,722
	%	2.61	17.96	15.82	0.81	62.78	100.00

		Parent 1: Overall school achievement of study child – hlc08a3a						
		Not asked or don't know	Excellent	Above average	Below average	Well below average	Average	Total
Non-responding	Freq	33	142	287	92	24	407	985
	%	1.21	5.22	10.54	3.38	0.88	14.95	36.19
Responding	Freq	19	370	622	90	28	608	1,737
	%	0.70	13.59	22.85	3.31	1.03	22.34	63.81
Total	Freq	52	512	909	182	52	1,015	2,722
	%	1.91	18.81	33.39	6.69	1.91	37.29	100.00

		Parent 2 Self-complete data added to file – hp2scd			
		Not asked	No	Yes	Total
Non-responding	Freq	181	305	499	985
	%	6.65	11.20	18.33	36.19
Responding	Freq	257	275	1,205	1,737
	%	9.44	10.10	44.27	63.81
Total	Freq	438	580	1,704	2,722
	%	16.09	21.31	62.60	100.00

		Parent 1 Indigenous status – zf12hp1		
		Not asked	Yes, Torres Strait Islander	Total
Non-responding	Freq	963	22	985
	%	35.38	0.81	36.19
Responding	Freq	1,728	9	1,737
	%	63.48	0.33	63.81
Total	Freq	2,691	31	2,722
	%	98.86	1.14	100.00

		Interviewed in November–February (Derived from fdatint) – eoy		
		No	Yes	Total
Non-responding	Freq	882	103	985
	%	32.40	3.78	36.19
Responding	Freq	1,623	114	1,737
	%	59.63	4.19	63.81
Total	Freq	2,505	217	2,722
	%	92.03	7.97	100.00

	Variable	Label	Observations	Mean	SD	Minimum	Maximum
Non-responding	hcnfsad2	14/15 - SEIFA -Inx Rel Soc-eco Adv and Disad - 2016 - SA2 - Score	985	1,001.87	76.85	777	1,178
Responding	hcnfsad2	14/15 - SEIFA -Inx Rel Soc-eco Adv and Disad - 2016 - SA2 - Score	1,737	1,024.47	73.53	730	1,178

K cross-sectional

		Parent 1 Main language spoken at home – cf11m2		
		English	Non-English	Total
Non-responding	Freq	550	2,644	3,194
	%	11.04	53.06	64.10
Responding	Freq	227	1,562	1,789
	%	4.56	31.35	35.90
Total	Freq	777	4,206	4,983
	%	15.59	84.41	100.00

		Highest year of primary or secondary school completed by mother – cfd08m1								
		Not asked	Don't know or refusal	Year 11 or equivalent	Year 10 or equivalent	Year 9 or equivalent	Year 8 or below	Never attended school	Year 12 or equivalent	Total
Non-responding	Freq	30	5	465	781	167	95	5	1,646	3,194
	%	0.60	0.10	9.33	15.67	3.35	1.91	0.10	33.03	64.10
Responding	Freq	7	2	203	264	35	26	3	1,249	1,789
	%	0.14	0.04	4.07	5.30	0.70	0.52	0.06	25.07	35.90
Total	Freq	37	7	668	1,045	202	121	8	2,895	4,983
	%	0.74	0.14	13.41	20.97	4.05	2.43	0.16	58.10	100.00

		Renting home – cho04a3b			
		Refusal or don't know	Yes	No	Total
Non-responding	Freq	7	1,002	2,185	3,194
	%	0.14	20.11	43.85	64.10
Responding	Freq	2	327	1,460	1,789
	%	0.04	6.56	29.30	35.90
Total	Freq	9	1,329	3,645	4,983
	%	0.18	26.67	73.15	100.00

		Parent 1 Self-complete data added to file – cp1scd		
		No	Yes	Total
Non-responding	Freq	626	2,568	3,194
	%	12.56	51.54	64.10
Responding	Freq	128	1,661	1,789
	%	2.57	33.33	35.90
Total	Freq	754	4,229	4,983
	%	15.13	84.87	100.00

		Parent 2 Self-complete data added to file – cp2scd			
		Not asked	No	Yes	Total
Non-responding	Freq	566	719	1,909	3,194
	%	11.36	14.43	38.31	64.10
Responding	Freq	131	179	1,479	1,789
	%	2.63	3.59	29.68	35.90
Total	Freq	697	898	3,388	4,983
	%	13.99	18.02	67.99	100.00

		Stratum							
		NSW Met	NSW Xmet large	NSW Xmet small	Vic. Met large	Vic. Met small	Vic. Xmet large	Vic. Xmet small	Qld Met
Non-responding	Freq	607	359	42	543	13	171	64	261
	%	12.18	7.20	0.84	10.90	0.26	3.43	1.28	5.24
Responding	Freq	344	206	15	325	6	92	30	147
	%	6.90	4,13	0.30	6.52	0.12	1.85	0.60	2.95
Total	Freq	951	565	57	868	19	263	94	408
	%	19.08	11.34	1.14	17.42	0.38	5.28	1.89	8.19

		Stratum							
		Qld Xmet large	Qld Xmet small	SA Met large	SA Xmet large	SA Xmet small	WA Met large	WA Met small	WA Xmet large
Non-responding	Freq	355	50	169	47	13	212	9	60
	%	7.12	1.00	3.39	0.94	0.26	4.25	0.18	1.20
Responding	Freq	129	26	104	20	6	133	6	37
	%	2.59	0.52	2.09	0.40	0.12	2.67	0.12	0.74
Total	Freq	484	76	273	67	19	345	15	97
	%	9.71	1.53	5,48	1.34	0.38	6.92	0.30	1.95

		Stratum						
		WA Xmet small	Tas. Met	Tas. Xmet	NT Met	NT Xmet small and NT Xmet large	ACT	Total
Non-responding	Freq	36	29	48	24	28	54	3,194
	%	0.72	0.58	0.96	0.48	0.56	1.08	64.10
Responding	Freq	14	25	34	18	13	59	1,789
	%	0.28	0.50	0.68	0.36	0.26	1.18	35.90
Total	Freq	50	54	82	42	41	113	4,983
	%	1.00	1.08	1.65	0.84	0.82	2.27	100.00

		Study child: sex – zf02m1		
		Male	Female	Total
Non-responding	Freq	1,711	1,483	3,194
	%	34.34	29.76	64.10
Responding	Freq	825	964	1,789
	%	16.56	19.35	35.90
Total	Freq	2,536	2,447	4,983
	%	50.89	49.11	100.00

	Variable	Label	Observations	Mean	SD	Minimum	Maximum
Non-responding	ccnfseo	4/5 - Home - SEIFA Education & Occupation	3,194	989.77	76.14	783.20	1,221.12
	cf03m2	4/5 - P1@W1 - F2F A4 - Age	3,194	34.19	5.79	0	73
	cf03m3	4/5 - P2@W1 - F2F A4 - Age	3,,194	30.43	15.21	0	70
Responding	ccnfseo	4/5 - Home - SEIFA Education & Occupation	1789	1,013.15	82.10	839.28	1,209.20
	cf03m2	4/5 - P1@W1 - F2F A4 - Age	1,789	35.70	4.88	0	65
	cf03m3	4/5 - P2@W1 - F2F A4 - Age	1,789	35.30	11.38	0	69

K longitudinal

		Housing tenure (derived from jho06a, jho04a3a, jho04a3b, jho04a3c, jho04a3d, jho04a3e, jho04a3f)					
		Owning home outright	Renting	Occupied rent free	Other tenure types	Owning home on a mortgage	Total
Non-responding	Freq	186	152	13	161	476	988
	%	7.41	6.06	0.52	6.41	18.96	39.36
Responding	Freq	413	183	27	105	794	1,522
	%	16.45	7.29	1.08	4.18	31.63	60.64
Total	Freq	599	335	40	266	1,270	2,510
	%	23.86	13.35	1.59	10.60	50.60	100.00

		Study child has 2 parents at home – jp2			
		Not asked	No	Yes	Total
Non-responding	Freq	155	275	558	988
	%	6.18	10.96	22.23	39.36
Responding	Freq	87	378	1,057	1,522
	%	3.47	15.06	42.11	60.64
Total	Freq	242	653	1,615	2,510
	%	9.64	26.02	64.34	100.00

		Parent 2 CATI data present – jp2cati		
		Missing	Yes	Total
Non-responding	Freq	509	479	988
	%	20.28	19.08	39.36
Responding	Freq	503	1,019	1,522
	%	20.04	40.60	60.64
Total	Freq	1,012	1,498	2,510
	%	40.32	59.68	100.00

		Study child sex – zf02m1		
		Male	Female	Total
Non-responding	Freq	559	429	988
	%	22.27	17.09	39.36
Responding	Freq	715	807	1,522
	%	28.49	32.15	60.64
Total	Freq	1,274	1,236	2,510
	%	50.76	49.24	100,00

	Variable	Label	Observations	Mean	SD	Minimum	Maximum
Non-responding	jcnfsad2	18/19 - SC - SEIFA - Inx Rel Soc-eco Adv and Disad - 2016 - SA2 - Score	988	1,007.84	71.84	796	1,178
	jcnfser2	18/19 - SC - Home - SEIFA Economic Resources - 2016 - SA2 - Score	988	1,008.84	62.13	733	1,171
	jnpeople	18/19 - No. people in household	988	3.42	1.94	0	11
	jnsib	18/19 - No. siblings of SC in household	988	0.93	1.01	0	7
Responding	jcnfsad2	18/19 - SC - SEIFA - Inx Rel Soc-eco Adv and Disad - 2016 - SA2 - Score	1,522	1,024.40	74.26	730	1,184
	jcnfser2	18/19 - SC - Home - SEIFA Economic Resources - 2016 - SA2 - Score	1,522	1,014.37	64.08	732	1,171
	jnpeople	18/19 - No. people in household	1,522	3.74	1.58	0	11
	jnsib	18/19 - No. siblings of SC in household	1,522	1.07	1.01	0	8