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. 20

Wave 7 weighting and non-response

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

March 2018

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>.

Technical paper

The Longitudinal Study of Australian Children: LSAC Technical paper No.20, Wave 7 Weighting and Non-Response

Authors: Simon Usback, Australian Bureau of Statistics LSAC processing team and the Australian Bureau of Statistics Household Survey Methodology team.

Contents

Acknowledgements

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. The Australian children included 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); and
- 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. Parents were also sent a mail survey or link to confirm their contact details via a webform between each main wave.

Wave 7 of the Longitudinal Study of Australian Children was conducted in 2016 with B-cohort children at age 12–13 years and K-cohort children at age 16–17 years. The number of active participants continues to decrease from wave to wave, as a result of failure to maintain contact, participants opting out or children moving out of scope (e.g., moving overseas). Some children are brought 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 mail-out 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, 6,470 children responded in Wave 7, and 5,820 children responded to all waves.

In undertaking the Wave 7 weighting process two issues were encountered that needed investigation, which resulted in a decision to change the components of the weighting process. The first of these was the discovery of an error in the longitudinal propensity models. The model was not accounting correctly for non-response in previous Wave 5, which resulted in the need for a change to the model. Please see the "Wave 5 re-weighting" section for a full explanation. The other issue encountered was a continuing increase in the number of units with sample weights appearing at the top weight cap of 2.5. Investigation was also done into this issue and the result was an increase to the top weight cap to 3.5 for Wave 7. This is described in more detail under the "Weight capping" section of the paper. Despite the correction to the propensity model and the increase in the top weight cap, the overall method for producing the weights is still unchanged from Wave 6.

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 population, 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 toward or against some demographics, affecting the ability to make inference 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 inference. Adjusting unit weights to account for attrition can improve the reliability of population inference.

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 7 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 postcodes, and that this variable should be used when conducting statistical tests to avoid overstating significance.

Summary of sample design properties

Full details of the LSAC sample design can be found in Soloff, Lawrence, and Johnstone (2005). We provide a summary here for your 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 years 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 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 the Wave 1 Weighting Paper.

Summary of weighting in Waves 1–5

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, Mission, & 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 and 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 2–5 were calculated by adjusting previous wave weights for differential sample attrition in two stages (Cusack & Defina, 2014; Sipthorp & Daraganova, 2011; Sipthorp & Misson, 2007, 2009). 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 a sample weight is calculated that adds up to the number of children in the sample. 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 the response 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.

Summary of Wave 6 weighting

Wave 6 used the same two-stage weighting method as Wave 5. The response propensity models were created based on the Wave 6 responses.

Each cohort had both a longitudinal weight and a cross-sectional weight, resulting in four response propensity models, which were updated in Wave 6. The differences between the cross-sectional weight models and longitudinal weight models were as follows:

- cross-sectional weight model—used all children from Wave 1 and Wave 1 data items to predict response propensity in Wave 6;
- longitudinal weight model—used children who had responded to all waves up to and including Wave 5, and Wave 5 data items, to predict response propensity in Wave 6.

Response propensity models were also updated with the addition of the variable indicating whether Parent 2 had returned the self-completed questionnaire (or a separate category if there was no Parent 2).

The B-cohort longitudinal weight model had two variables added and two variables removed. The two variables added were overall school achievement of the study child (teacher reported) and Parent 1's housing tenure. The variables removed were SEIFA Economic Resources score (no relationship to Wave 6 non-response) and mother's proficiency in spoken English (not collected in Wave 5).

The K-cohort longitudinal weight model had three variables added and two variables removed. The three variables added were language and literacy skills of the study child (teacher reported), whether Parent 1 rents their home and how many days each week someone in the household helps the study child with homework. The variables removed were SEIFA Economic Resources score (no relationship to Wave 6 non-response) and mother's proficiency in spoken English (not collected in Wave 5).

Update to the propensity model

Part of the weighting process for the LSAC survey involves adjusting for non-response by particular characteristics that may have different attrition than average. This is achieved by developing a propensity model based on responses from the previous waves using logistic regression applied to relevant covariates.

For the longitudinal weights for Wave 5, the propensity model should account for non-response in Wave 5 among those units that have responded to all previous waves from Wave 1 to 4. However, this model was previously developed using respondents to Wave 4 regardless of responses to previous waves. This was incorrect as response propensity adjustments applied in Waves 2–4 had already accounted for those units that had not responded to one of these waves, so they should not have been included in longitudinal modelling again. Once the correct response flags were applied, the previously identified model was no longer optimal as some of the covariates were no longer significant (see details below).

The wave responses were corrected and a stepwise process was applied to create an updated logistic model using candidate variables. The models were assessed using key measures such as the Wald statistic of each covariate (a measure of their significance in the model), the AIC value (a trade-off between model fit and over fitting) and the C statistic (the AUC or a measure of the discriminatory power of the model). The newly created model is outlined below and the difference between the previous and current population weights are examined.

Tables 2 and 3 show the previous covariates used.

Table 2: B-cohort previous model (achieving a C-stat of 0.716)

Variable	Description
DF03DP1	Parent 1 age
DCNFSER	SEIFA Economic Resources 2011 score (*no longer significant)
DFD08M1	Mother level school completion
DP2SCD	Parent 2 self-completed data present
DFD11M2	Mother's proficiency in spoken English (*no longer significant)

Table 3: K-cohort previous model (achieving a C-stat of 0.685)

Variable	Description
FF03FP1	Parent 1 age (*shows some evidence of significance)
FCNFSER	SEIFA Economic Resources 2011 score (*no longer significant)
FFD08M1	Mother level school completion
FFD11M2	Mother's proficiency in spoken English
FP2SCD	Parent 2 self-completed data present

Numerous covariates were examined using the correct inclusion criteria and Tables 4 and 5 show the results in the updated model.

Table 4: B-cohort updated model (significant and relevant covariates) (achieving a C-stat of 0.732)

Variable	Description
DF03DP1	Parent 1 age
DFD08A3A	Parent 1 highest qualification
DP2SCD	Parent 2 self-completed data present

Table 5: K-cohort updated model (significant and relevant covariates) (achieving a C-stat of 0.708)

Variable	Description
FF03FP1	Parent 1 age
FP2SCD	Parent 2 self-completed data present
FFD08A1	Parent 1 school completion
FLC08T1B	T/C reading progress
FLC08A3A	Parent 1 overall school achievement
FFD11M2	Mother's proficiency in spoken English

Reweighting of Waves 5 and 6 using the updated propensity model

With the correction of the propensity model completed, the data for Waves 5 and 6 were reweighted making use of the updated model. Tables 6–9 below show the differences in the population weights when using the old and updated models for Wave 5 and Wave 6. Population weights are calculated by multiplying the sample weights by a constant factor, based on the sampling fraction, so that the sum of the weights add up to the population total at Wave 1 (for comparability and consistency). This factor depends on cohort and model type. The absolute difference comparisons shown below may seem large but the population weights themselves can range from 20 to 170 as opposed to the sample weights, which are constrained between 0.33 and 3.5.

Table 6: Wave 5 weight differences B cohort

Absolute difference comparison	Frequency	Percentage	Cumulative frequency	Cumulative percentage
Diff. < 5	3,456	91.96	3,456	91.96
5 < Diff. < 10	176	4.68	3,632	96.64
Diff. > 10	126	3.36	3,758	100.00

Table 7: Wave 5 weight differences K cohort

Absolute difference comparison	Frequency	Percentage	Cumulative frequency	Cumulative percentage
Diff. < 5	3,301	89.65	3,301	89.65
5 < Diff. < 10	257	6.98	3,558	96.63
Diff. > 10	124	3.37	3,682	100.00

Table 8: Wave 6 weight differences B cohort

Absolute difference comparison	Frequency	Percentage	Cumulative frequency	Cumulative percentage
Diff. < 5	3,117	90.58	3,117	90.58
5 < Diff. < 10	210	6.10	3,327	96.69
Diff. > 10	114	3.31	3,441	100.00

Table 9: Wave 6 weight differences K cohort

Absolute difference comparison	Frequency	Percentage	Cumulative frequency	Cumulative percentage
Diff. < 5	2,950	90.05	2,950	90.05
5 < Diff. < 10	209	6.38	3,159	96.43
Diff. > 10	117	3.57	3,276	100.00

Wave 7 weighting method

This section contains a brief description of the method used to create weights for Wave 7 data. The method is largely unchanged from Wave 5 with some slight corrections made, as discussed above. 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.

Cross-sectional weights begin with the final weight used in Wave 1 and adjust for all additional non-responses in the current wave—regardless of whether a unit responded in Waves 2–6.

Initial weights

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

- For Wave 7 longitudinal weights (which applies to those who have responded to all Waves 1, 2, 3, 4, 5, 6 and 7), the initial weight for children in Wave 7 is the final corrected longitudinal weight from Wave 6.
- For Wave 7 cross-sectional weights (which applies to all of those who responded in Wave 7), the initial weight for children in Wave 7 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 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 at 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 6. A stepwise model selection process is used that considers all possible covariates for the response propensity model (list of variables considered 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. These model selection processes resulted in a shortlist of variables to consider adding to the Wave 6 models.

The variables that showed the strongest effects (the highest score chi-square statistic) in the model selection process were then added in various combinations with Wave 6 variables. Wave 6 variables that were clearly no longer significant ($p > 0.1$) were removed from the model. The other variables used in Wave 6 that were still useful predictors for Wave 7 were maintained where possible to achieve consistency over time. New covariates were chosen by taking the combination with Wave 6 variables that resulted in the lowest Akaike Information Criterion (AIC).

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

Wave 1 variables used in the K-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 (new)

Wave 6 variables used in the B-cohort longitudinal weight model

- Matrix reasoning score missing (new)
- Parent 1 age
- Matrix reasoning (new)
- Mother: English as main language at home (new)
- Parent 2 self-completed questionnaire returned
- Parent 1 renting home indicator
- Interviewed in Nov., Dec., Jan. or Feb. (derived from fdatint) (new)
- Participation in checkpoint health interview (new)

Wave 6 variables used in the K-cohort longitudinal weight model

- Parent 1 age
- Mother's highest level of high school completed
- Parent 2 self-completed questionnaire returned
- Parent 1 renting home indicator
- How far study child will go in education (new)
- Parent 1 SEIFA decile of relative socio-economic advantage (new)

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

Odds ratio estimates for the levels of 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 re-align the sample composition within each stratum at each wave to the composition within each stratum as at Wave 1, and to re-align 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 ("met"/"exmet"), 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 0.33 or the upper weight cap of 3.5 (changed for Wave 7 from the previous waves' value of 2.5). 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 during 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 number of units assigned weights at the usual caps (lower 0.33 and upper 2.5) has been increasing each study wave. This is an expected result due to increasing attrition rates over time. However, this effect has raised concerns as to whether the weighting caps were still appropriate at current levels as there is the potential for bias to be introduced in the estimate if a large number of weights are constrained. As the responding sample becomes smaller with each successive wave of the study it is likely that even more units will be given weights at the caps as certain groups become less represented.

As a result, a new upper cap of 3.5 has been introduced and is intended to stay in place for several further waves before requiring review. The upper cap of 3.5 was chosen as it doesn't constrain too many units and will continue to be appropriate in future waves. The lower cap of 0.33 remains unchanged from Wave 6. More detail on the number of units now appearing at the caps can be seen in Tables 13 and 14 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 fully responding in a previous wave. Consider the following acquisition figures for Wave 7.

For the B cohort, out of 1,343 that did not respond to Wave 6, 124 responded to Wave 7. Out of the 1,666 that did not respond to at least one of Waves 2, 3, 4, 5 or 6, 353 responded to Wave 7.

For the K cohort, out of 1,446 that did not respond to Wave 6, 120 responded to Wave 7. Out of the 1,707 that did not respond to at least one of Waves 2, 3, 4, 5 or 6, 297 responded to Wave 7.

Table 10 shows those who have responded after previously being a "non-responder" in a previous wave (sample reacquisition).

Table 10: Sample reacquisition for waves 3, 4 and 5

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
B	133	135	129	89	124
K	135	119	94	77	120

Total responding sample for each wave

The fully responding sample at various stages in the sample drives the calibration and hence weighting process. Observe Tables 11 and 12 below for updated counts.

Table 11: Sample counts for the B cohort

Wave	1	2	3	4	5	6	7
Cross-sectional response	5,107	4,606	4,386	4,242	4,085	3,764	3,381
Longitudinal response	–	4,606	4,253	3,997	3,758	3,441	3,028
Cross-sectional attrition rate (%)	–	9.8	14.1	16.9	20.0	26.3	33.8
Longitudinal attrition rate (%)	–	9.8	7.7	6.0	6.0	8.4	12.0

Table 12: Sample counts for the K cohort

Wave	1	2	3	4	5	6	7
Cross-sectional response	4,983	4,464	4,331	4,169	3,956	3,537	3,089
Longitudinal response	–	4,464	4,196	3,940	3,682	3,276	2,792
Cross-sectional attrition rate (%)	–	10.4	13.1	16.3	20.6	29.0	38.0
Longitudinal attrition rate (%)	–	10.4	6.0	6.1	6.6	11.0	14.8

- 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, and all waves beforehand, as a percentage of the previous wave's longitudinal response.

Number of children with weight at cap

Tables 13 and 14 below show the number of children with a sample weight at the lower cap of 0.33 and upper cap of 3.5 by cohort and by type of weight. The counts of units with weights at the lower cap have generally increased since Wave 6, especially for the cross-sectional weights. The counts of units with weights at the upper cap, however, have decreased significantly due to the increase of the upper cap from 2.5 in Wave 6 to 3.5 in Wave 7.

For the B cohort, the number of units at the upper cap has decreased from 116 in Wave 6 to 42 for the cross-sectional weight, and decreased from 142 in Wave 6 to 18 for the longitudinal weight.

Table 13: Counts of capped sample weights for Wave 7—B cohort

State	Cross-sectional		Longitudinal	
	Lower cap (0.33)	Upper cap (3.5)	Lower cap (0.33)	Upper cap (3.5)
NSW	0	17	0	8
VIC.	0	8	0	3
QLD	13	5	9	3
SA	1	5	3	0
WA	22	5	13	2
TAS.	9	1	11	2
NT	11	0	13	0
ACT	0	1	0	0
AUS.	56	42	49	18

For the K cohort, the number of units at the upper cap has decreased from 74 in Wave 6 to 22 for the cross-sectional weight, and decreased from 121 in Wave 6 to 9 for the longitudinal weight.

Table 14: Counts of capped sample weights for wave 7—K cohort

State	Cross-sectional		Longitudinal	
	Lower cap (0.33)	Upper cap (3.5)	Lower cap (0.33)	Upper cap (3.5)
NSW	0	8	0	2
VIC.	0	6	0	2
QLD	5	4	11	5
SA	0	2	0	0
WA	0	2	0	0
TAS.	18	0	18	0
NT	31	0	2	0
ACT	1	0	0	0
AUS.	55	22	31	9

Conclusion

Sample attrition has continued again in this wave; however, the responding sample still remains above 3,000 for both cohorts. 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 response. The weights developed help to correct for different response patterns, allowing users to analyse the data and draw conclusions about the population.

There are far less weights at the upper weight cap for this wave due to the increase in the upper cap from 2.5 to 3.5. The weight capping ensures that no unit contributes too much or too little to any analysis done using this data.

The response propensity models have changed for this wave. This represents a change in the observed response; however, care should be taken when using this observed behaviour to infer causal relationships (i.e., that particular characteristics cause non-response). The models reflect the observed response patterns and the weights developed provide a tool that may be useful for adjusting for changes in sample composition in analysis.

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.

Norton, A., & Monahan, K. (2015). *LSAC Technical Paper No. 15: Wave 6 weighting and non-response*. Melbourne: Australian Institute of Family Studies.

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

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.

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.

Appendix A: Glossary of terms and abbreviations

Many technical terms are used in this paper, 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
Akaike Information Criterion (AIC)	A measure of the relative quality of statistical models for the same set of data, used to inform model selection
Attrition	Process of sample size shrinking over time due to any mechanism
Cohort	Sample with a particular characteristic, e.g. B cohort aged 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, the parent with whom the LSAC face-to-face interview is conducted, generally the child's mother
P2	Parent 2, the child's second parent
Partial response	Acquisition of data for some study modules but not others
Post-stratification	Process of dividing population into groups for the purpose of weighting to benchmark totals
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	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 a set number of children were 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 7 weights

Table B1: Description of Wave 7 weights

SAS name	Cohort	Type	Waves cases responded to
gweight	B	Population	1 & 7
gweights	B	Sample	1 & 7
bcdefgwt	B	Population	1, 2, 3, 4, 5, 6 & 7
bcdefgwts	B	Sample	1, 2, 3, 4, 5, 6 & 7
iweight	K	Population	1 & 7
iweights	K	Sample	1 & 7
defghiwts	K	Population	1, 2, 3, 4, 5, 6 & 7
defghiwt	K	Sample	1, 2, 3, 4, 5, 6 & 7

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	31.6	< 0.0001
AF03M3	Parent 2 age	1	5.7	0.0166
AFD08M1	Mother's highest year of high school completed	4	115.1	< 0.0001
AFD11M2	Mother's proficiency in spoken English	4	76.0	< 0.0001
AP1SCD	Parent 1 self-completed questionnaire returned	1	23.9	< 0.0001
AP2SCD	Parent 2 self-completed questionnaire returned	2	23.9	< 0.0001

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
MISS_MATREAS	Flag for non-participation / non-completion of matrix reasoning task	1	1.7	0.1874
FF03FP1	Parent 1 age	1	5.4	0.0198
FMATREAS	Matrix reasoning	1	7.1	0.0076
FF11FM	Mother, language other than English spoken at home	1	6.2	0.0127
FP2SCD	Parent 2 self-complete data	2	29.8	< 0.0001
FHO04A5	Parent 1 housing tenure	4	14.0	0.0073
EOY	Interviewed in November–February	1	9.0	0.0028
CHCP_RESP	Response status in CHCP	2	125.7	< 0.0001

Table C3: K cohort—cross-sectional weights

Variable name	Description	DF	Wald Chi-Square	Pr > ChiSq
CF03M2	Parent 1 age	1	21.0	< 0.0001
CF03M3	Parent 2 age	1	5.9	0.0148
CFD08M1	Mother's highest year of high school completed	4	90.2	< 0.0001
CFD11M2	Mother's proficiency in spoken English	4	37.2	< 0.0001
CP1SCD	Parent 1 self-completed questionnaire returned	1	6.9	0.0087
CP2SCD	Parent 2 self-completed questionnaire returned	2	57.0	< 0.0001
CHO04A3B	Parent 1 rents home	1	24.5	< 0.0001

Table C4: K cohort—longitudinal weights

Variable name	Description	DF	Wald Chi-Square	Pr > ChiSq
HF03HP1	Parent 1 age	1	8.4	0.0038
HFD08M1	Mother's highest year of high school completed	4	10.8	0.0288
HCNFSAD2D	SEIFA Index of Relative Socio-economic Advantage and Disadvantage	9	20.4	0.0157
GP2SCD	Parent 2 self-completed questionnaire returned	2	69.9	< 0.0001
HHO04A3B	Parent 1 rents home	2	17.8	0.0001
HHE131A	Parent 1—How far study child will go in education	6	24.9	0.0004

Appendix D: Odds ratio estimates for variables in Wave 7 response propensity models

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

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

Effect	Description	Point estimate	95% Wald confidence limits	
af03m3	Parent 2 age	1.017	1.003	1.032
af03m2	Parent 1 age	1.043	1.028	1.058
afd08m1 1 vs 5	Mother completed Year 12 or equivalent	3.282	1.989	5.417
afd08m1 2 vs 5	Mother completed Year 11 or equivalent	1.718	1.016	2.906
afd08m1 3 vs 5	Mother completed Year 10 or equivalent	1.698	1.016	2.837
afd08m1 4 vs 5	Mother completed Year 9 or equivalent	1.268	0.702	2.29
afd11m2 0 vs 4	Not applicable to mother's proficiency in spoken English	1.881	0.779	4.538
afd11m2 1 vs 4	Mother speaks English very well	0.968	0.394	2.377
afd11m2 2 vs 4	Mother speaks English well	0.702	0.276	1.784
afd11m2 3 vs 4	Mother speaks English not well	0.854	0.324	2.25
ap1scd 0 vs 1	Parent 1 did not return self-completed questionnaire	0.566	0.451	0.711
ap2scd -9 vs 1	No Parent 2 in household	0.705	0.43	1.153
ap2scd 0 vs 1	Parent 2 did not return self-completed questionnaire	0.591	0.478	0.732

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

Effect	Description	Point estimate	95% Wald confidence limits	
MISS_MATREAS	Flag for non-participation / non-completion of matrix reasoning task	1.555	0.807	2.996
ff03fp1	Parent 1 age	1.025	1.004	1.047
fmatreas	Matrix reasoning	1.051	1.013	1.091
ff11fm_col 1 vs 2	Mother, language other than English spoken at home	1.471	1.086	1.992
fp2scd -9 vs 1	No Parent 2 in household	0.583	0.419	0.81
fp2scd 0 vs 1	Parent 2 did not return self-completed questionnaire	0.505	0.391	0.65
fho04a5_col -9 vs 6	Not applicable housing tenure	0.278	0.094	0.822
fho04a5_col 1 vs 6	House being paid off	0.965	0.433	2.148
fho04a5_col 2 vs 6	House owned outright	1.299	0.553	3.055
fho04a5_col 3 vs 6	House rented	1.023	0.452	2.317
EOY 0 vs 1	Interviewed in November–February	1.628	1.183	2.239
CHCP 0 vs 1	Participated in Child Health Check Point (between Waves 6 and 7)	0.5	0.304	0.822
CHCP_RESP 1 vs 3	Response status In Child Health Check Point (Status indicates reason for non-response)	3.651	1.73	7.707

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

Effect	Description	Point estimate	95% Wald confidence limits	
cf03m3	Parent 2 age	1.017	1.003	1.03
cf03m2	Parent 1 age	1.033	1.019	1.047
cfd08m1 1 vs 5	Mother completed Year 12 or equivalent	2.208	1.466	3.326
cfd08m1 2 vs 5	Mother completed Year 11 or equivalent	1.444	0.938	2.226
cfd08m1 3 vs 5	Mother completed Year 10 or equivalent	1.213	0.796	1.847
cfd08m1 4 vs 5	Mother completed Year 9 or equivalent	0.897	0.551	1.461
cfd11m2 0 vs 4	Not applicable to mother's proficiency in spoken English	1.342	0.607	2.968
cfd11m2 1 vs 4	Mother speaks English very well	0.937	0.416	2.11
cfd11m2 2 vs 4	Mother speaks English well	0.618	0.268	1.422
cfd11m2 3 vs 4	Mother speaks English not well	0.846	0.354	2.023
cp1scd 0 vs 1	Parent 1 did not return self-completed questionnaire	0.745	0.599	0.928
cp2scd -9 vs 1	No Parent 2 in household	0.887	0.529	1.486
cp2scd 0 vs 1	Parent 2 did not return self-completed questionnaire	0.45	0.366	0.554
cho04a3b 1 vs 2	Parent 1 rents home	0.695	0.601	0.802

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

Effect	Description	Point estimate	95% Wald confidence limits	
hf03hp1	Parent 1 age	1.026	1.008	1.044
hfd08m1_col 1 vs 5	Mother completed Year 12 or equivalent	1.385	0.618	3.106
hfd08m1_col 2 vs 5	Mother completed Year 11 or equivalent	1.185	0.512	2.746
hfd08m1_col 3 vs 5	Mother completed Year 10 or equivalent	0.928	0.409	2.105
hfd08m1_col 4 vs 5	Mother completed Year 9 or equivalent	0.83	0.316	2.182
hcnfsad2d 1 vs 10	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National 1st Decile	1.0	0.612	1.634
hcnfsad2d 2 vs 10	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National 2nd Decile	0.997	0.631	1.574
hcnfsad2d 3 vs 10	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National 3rd Decile	0.852	0.542	1.34
hcnfsad2d 4 vs 10	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National 4th Decile	1.002	0.644	1.561
hcnfsad2d 5 vs 10	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National 5th Decile	1.37	0.857	2.188
hcnfsad2d 6 vs 10	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National 6th Decile	1.018	0.658	1.574
hcnfsad2d 7 vs 10	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National 7th Decile	1.727	1.059	2.816
hcnfsad2d 8 vs 10	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National 8th Decile	0.688	0.451	1.051

Table continued over page

Effect	Description	Point estimate	95% Wald confidence limits	
hcnfsad2d 9 vs 10	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National 9th Decile	1.282	0.823	1.998
hp2scd -9 vs 1	No Parent 2 in household	0.539	0.404	0.718
hp2scd 0 vs 1	Parent 2 did not return self-completed questionnaire	0.362	0.284	0.461
hho04a3b -9 vs 2	Not applicable to Parent 1 renting home	0.217	0.102	0.46
hho04a3b 1 vs 2	Parent 1 renting home	1.157	0.866	1.546
hhe13a -9 vs 5	Not applicable to study child will go in education	0.339	0.15	0.764
hhe13a -2 vs 5	Don't know to study child will go in education	1.746	0.582	5.234
hhe13a 1 vs 5	Study child will leave school before finishing secondary school	0.282	0.15	0.531
hhe13a 2 vs 5	Study child will complete secondary school	0.602	0.375	0.967
hhe13a 3 vs 5	Study child will complete a trade or vocational training course	0.593	0.377	0.931
hhe13a 4 vs 5	Study child will go to university and complete a degree	0.737	0.492	1.104

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—Rents home
aho04a5	0/1—P1—F2F L5—Housing tenure
aho09a1a1	0/1—P1—F2F L11—Safe neighbourhood
anpeople	0/1—No. of people in household
ansib	0/1—No. of siblings of study child in household
ap1scd	0/1—Parent 1 self-completed data present
ap2	0/1—Study child has two 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 6 data items considered for B cohort—longitudinal weight

Variable name	Variable label
fcnfsad2	10/11—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Score
fcnfsad2d	10/11—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National
fcnfser2	10/11—Home—SEIFA Economic Resources—2011—SA2—Score
fcnfser2d	10/11—Home—SEIFA Economic Resources—2011—SA2—Deciles—National
ff01fm	M@10/11—Present for wave
ff03fp1	10/11—P1@10/11—Age
ff03fp2	10/11—P2@10/11—Age
ff11fm	10/11—M@10/11—Main language spoken at home
ff11fp1	10/11—P1@10/11—Main language spoken at home
ff11m1	10/11—SC—Main language spoken at home
ffd08a1	10/11—P1—F2F W1-3+A1.1/A1.2/A1.3—School completion
ffd08a2a	10/11—P1—F2F W1-3+A1.2/A1.3—Completed other qualification
ffd08a3a	10/11—P1—F2F W1-3+A1.2/A1.3—Highest qualification
ffd08m1	10/11—M—F2F W1-3+A1.1/A1.2/A1.3—School completion
ffemp	10/11—F—Employment status
fho04a1	10/11—P1—F2F P3—Home ownership
fho04a3b	10/11—P1—F2F P3.2—Rents home
fho04a5	10/11—P1—F2F P3—Housing tenure
flc08t3b	10/11—T/C—Teach 22.3—Overall school achievement
fmatreas	10/11—Matrix reasoning imputed
fmemp	10/11—M—Employment status
fnpeople	10/11—No. of people in household
fnsib	10/11—No. of siblings of study child in household
fp2	10/11—Study child has two parents in the home
fp2scd	10/11—Parent 2 self-complete data present
zf02fp1	P1@10/11—Sex
zf09fp1	P1@10/11—Country of birth
zf12fp1	P1@10/11—Indigenous status
fhe11a3e	10/11—P1—F2F C17.2—How often help child with homework
fhb24a	10/11—Teach 16—Activity during organised activities
fhe09a	10/11—F2F M8.1—Extracurricular—Any
CHCP	derived flag based on participation in CHCP
CHCP_RESP	derived variable based on participation or reason for non-response to CHCP
EOY	derived flag for being surveyed in Nov., Dec. or Jan.
month of interview	derived from datint
ff01fp2	P2@10/11—Present for wave
fid40h	10/11—F2F T1.1—Parent consent for Matrix Reasoning
fid44a1	10/11—Matrix Reasoning completed
fid44a2	10/11—F2F T1.6—Reason Matrix Reasoning not completed
fid44b	10/11—F2F T1.7—Study child stayed focused on Matrix Reasoning
fid44c	10/11—F2F T1.8—Parent present during Matrix Reasoning
fid44d	10/11—F2F T1.9—Sibling present during Matrix Reasoning

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—Rents home
cho04a5	4/5—P1—F2F L5—Housing tenure
cho09a1a1	4/5—P1—F2F L11—Safe neighbourhood
cnpeople	4/5—No. of people in household
cnsib	4/5—No. of siblings of study child in household
cp1scd	4/5—Parent 1 self-completed data present
cp2	4/5—Study child has two 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 6 data items considered for K cohort—longitudinal weight

Variable name	Variable label
hcnfsad2	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Score
hcnfsad2d	14/15—SEIFA—Index of Relative Socio-economic Advantage and Disadvantage—2011—SA2—Deciles—National
hcnfser2	14/15—Home—SEIFA Economic Resources—2011—SA2—Score
hcnfser2d	14/15—Home—SEIFA Economic Resources—2011—SA2—Deciles—National
hf01hm	M@14/15—Present for wave
hf01hp2	P2@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—Main language spoken at home
hfd08a1	14/15—P1—F2F A1.1/A1.3+W1—5—School completion
hfd08m1	14/15—M—F2F A1.1/A1.3+W1—5—School completion
hfemp	14/15—F—Employment status
hf22ahp1	Parent 1@14/15—Study child helps with everyday activities—Parent 1
hho04a3b	14/15—P1—F2F P1.6.2—Rents home
hho04a5	14/15—P1—F2F P1.6-8—Housing tenure
hlc08t1b	14/15—T/C—Teach 18—Reading progress
hmemp	14/15—M—Employment status
hnpeople	14/15—No. of people in household
hnsib	14/15—No. of siblings of study child in household
hp2	14/15—Study child has two 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
hhe13a	14/15—P1—F2F C7.1—How far study child will go in education
hlc08a3a	14/15—P1—F2F C7.4—Overall school achievement
ghe11a3e	12/13—P1—F2F C6.2—How often help child with homework
stratum	Stratum
zf02m1	SC—F2F A3—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	SD	Minimum	Maximum	Mode	Range	Sum	n
0.6620323	0.1703521	0.0845640	0.9234743	0.7972682	0.8389103	3381.00	5,107

Figure F2:   Distribution of final sample weight for wave 7—B cohort cross-sectional weight

Table F2: Analysis variable: GWEIGHTS—B cohort cross-sectional weight

Mean	SD	Minimum	Maximum	Mode	Range	Sum	n
1.0000000	0.5817048	0.3300000	3.5000000	0.3300000	3.1700000	3381.00	3,381

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	SD	Minimum	Maximum	Mode	Range	Sum	n
0.8799768	0.1118793	0.1228189	0.9841539	0.9654372	0.8613350	3028.00	3,441

Figure F4: Distribution of final sample weight for Wave 7—B cohort longitudinal weight

Table F4: Analysis variable: BCDEFGWTS—B cohort longitudinal weight

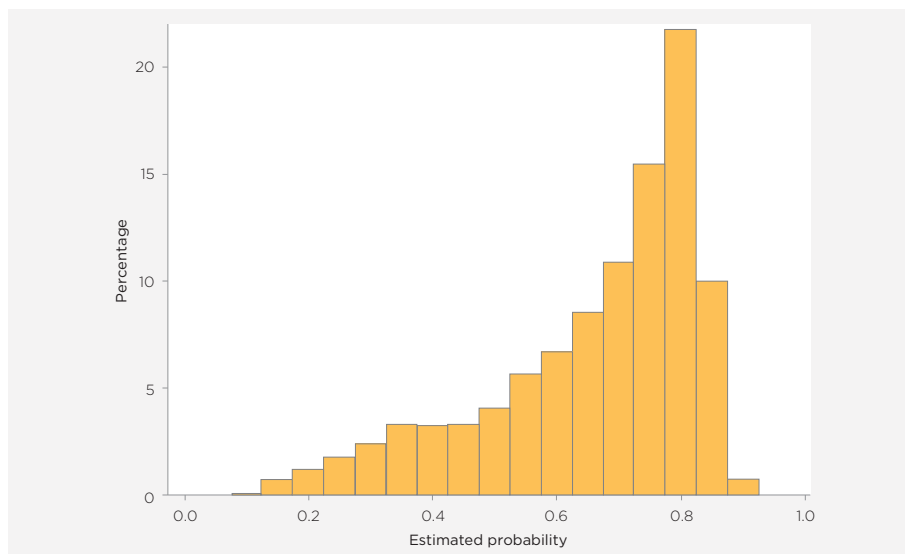| Mean	SD	Minimum	Maximum	Mode	Range	Sum	n
1.0000000	0.5730721	0.3300000	3.5000000	0.3300000	3.1700000	3028.00	3,028

K cohort—cross-sectional weight
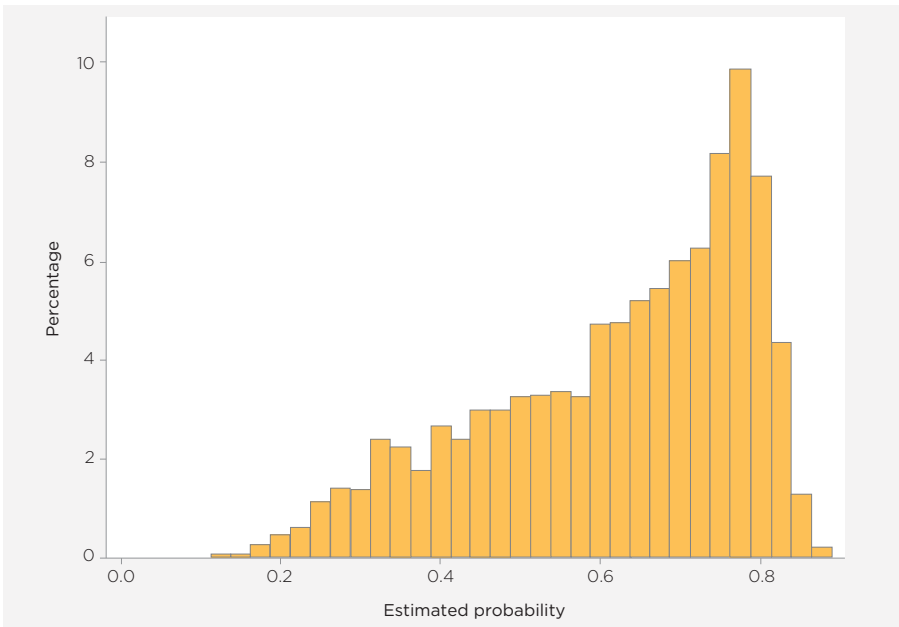
Figure F5:   Distribution of estimated response propensities—K cohort cross-sectional weight

Table F5: Analysis variable: estimated probability—K cohort cross-sectional weight

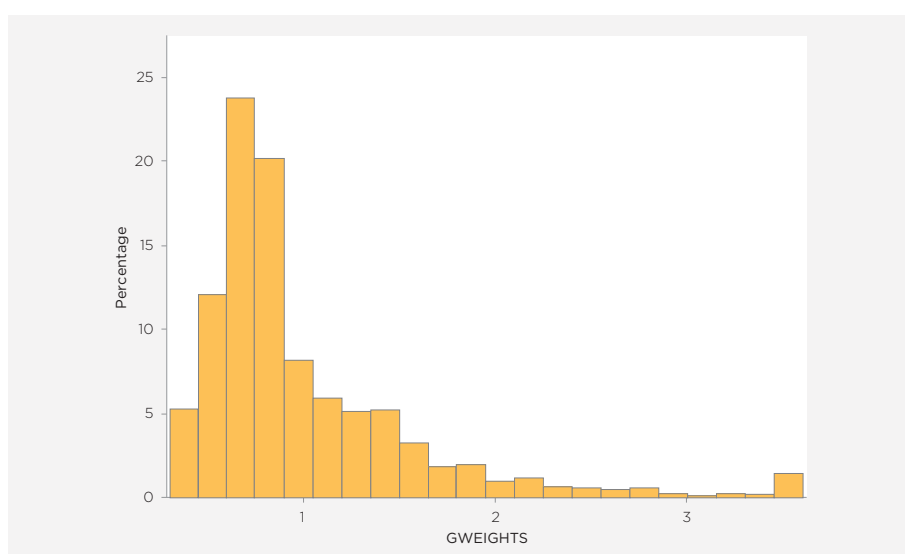| Mean	SD	Minimum	Maximum	Mode	Range	Sum	n
0.6199076	0.1632044	0.1172021	0.8771693	0.7625012	0.7599672	3089.00	4,983
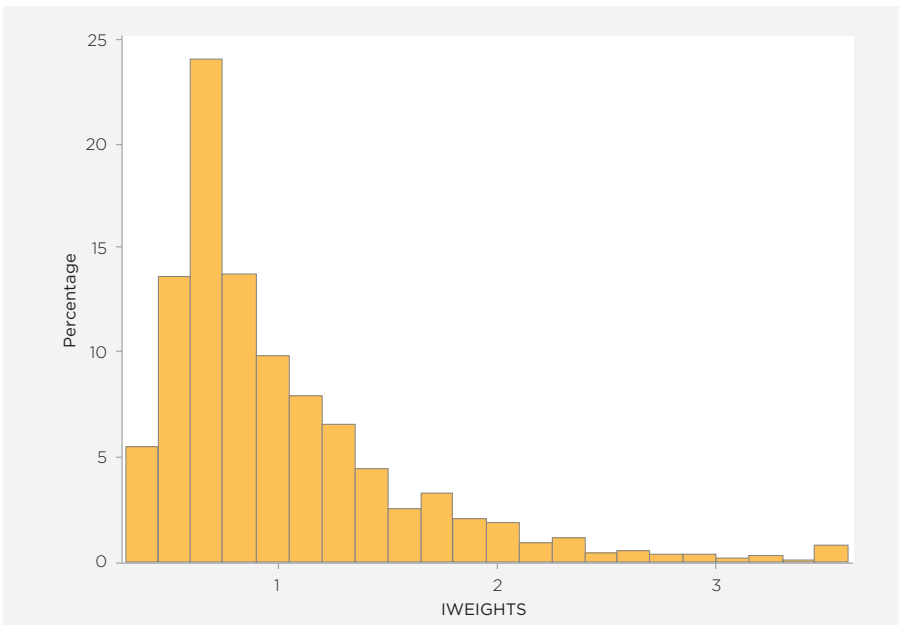
Figure F6:   Distribution of final sample weight for Wave 7—K cohort cross-sectional weight

Table F6: Analysis variable: IWEIGHTS—K cohort cross-sectional weight

Mean	SD	Minimum	Maximum	Mode	Range	Sum	n
1.0000000	0.5506967	0.3300000	3.5000000	0.3300000	3.1700000	3089.00	3,089

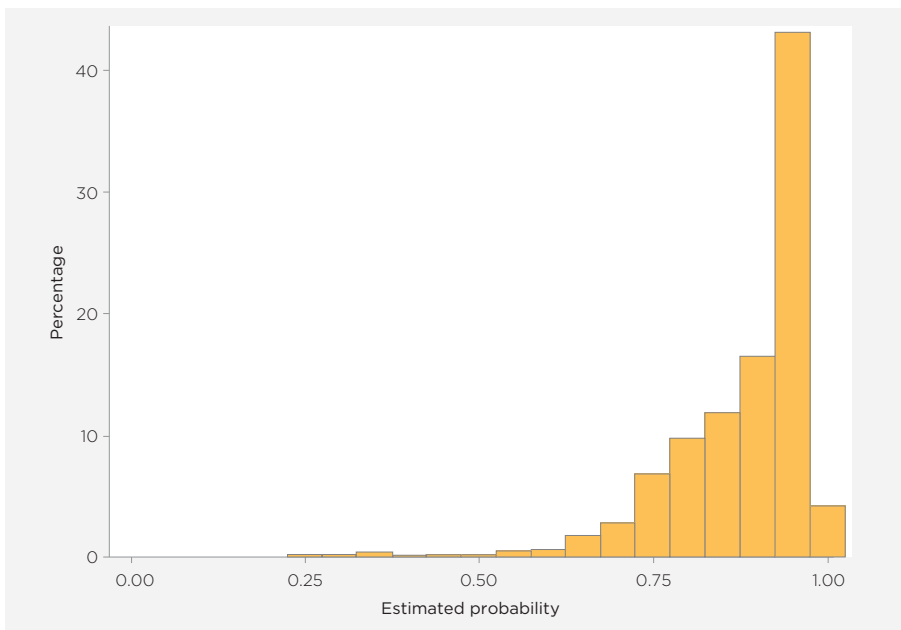## K cohort—longitudinal weight
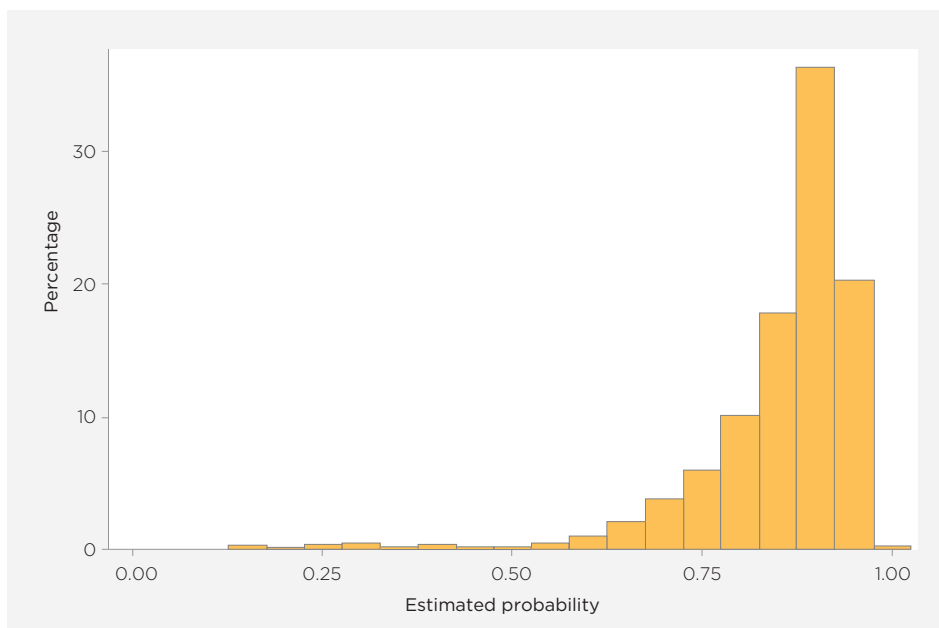
Figure F7: Distribution of estimated response propensities—K cohort longitudinal weight

Table F7: Analysis variable: estimated probability—K cohort longitudinal weight

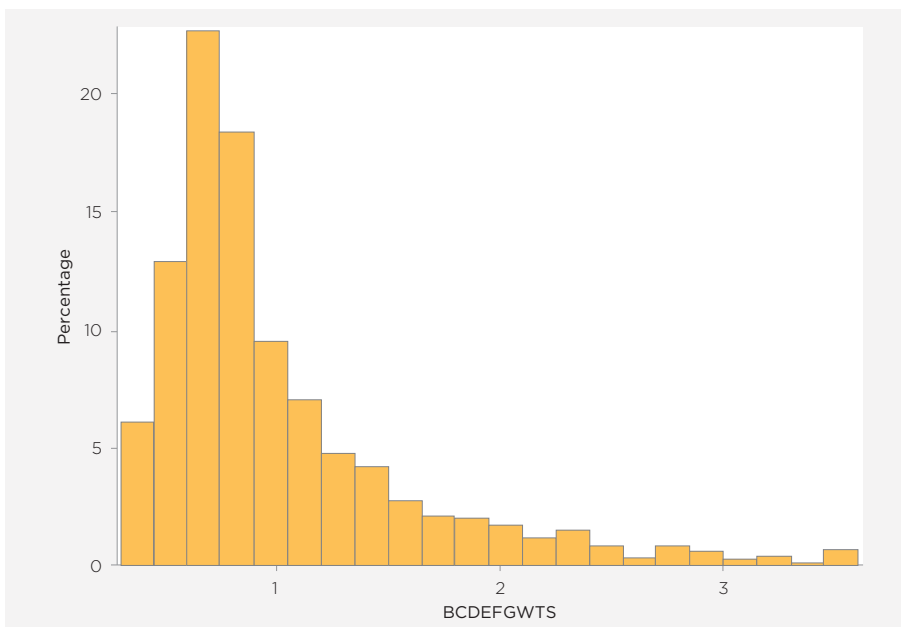| Mean	SD	Minimum	Maximum	Mode	Range	Sum	n
0.8522533	0.1130539	0.1273007	0.9832894	0.9122197	0.8559887	2791.98	3,276
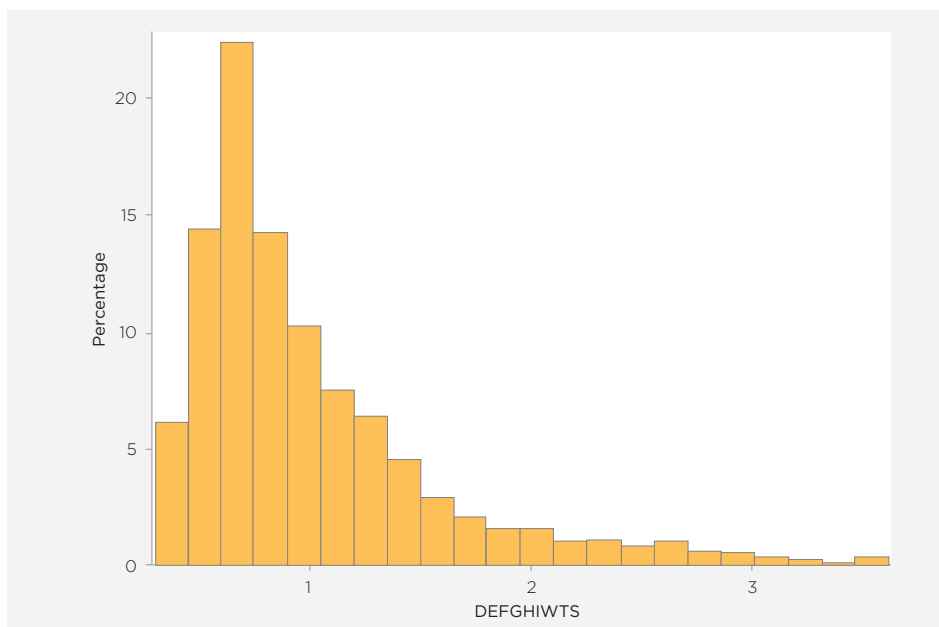
Figure F8: Distribution of final sample weight for Wave 7—K cohort longitudinal weight

Table F8: Analysis variable: DEFGHIWTS—K cohort longitudinal weight

Mean	SD	Minimum	Maximum	Mode	Range	Sum	n
1.0000000	0.5651332	0.3300000	3.5000000	0.3300000	3.1700000	2792.00	2,792

Appendix G: Non-response to instruments

Table G1: Non-response to instruments

	Eligible	Responding	%Wave 1	Response rate %
				B cohort
Wave 7 (issued sample = 4,319)				
Interview	3,381	3,381	66.2	100.0
P1CASI	3,374	3,287	64.4	97.4
P2SC	2,794	1,999	na	71.5
PLECATI	507	325	na	64.1
TEACH	3,333	2,567	na	77.0
ACASI	3,238	3,212	62.9	99.2
CSRB	3,238	3,224	63.1	99.6
TUD	3,237	2,684	52.6	82.9
Wave 6 (issued sample = 4,483)				
Interview	3,764	3,764	73.7	100.0
P1CASI	3,759	3,668	71.8	97.6
P2SC	3,198	2,312	na	72.3
PLECATI	559	398	na	71.2
TEACH	3,762	3,100	na	82.4
ACASIB	3,648	3,597	70.4	98.6
TUD	3,649	3,460	67.8	94.8
MR	3,648	3,585	70.2	98.3
				K cohort
Wave 7 (issued sample = 4,175)				
Interview	3,089	3,089	62.0	100.0
P1CASI	3,048	3,003	60.3	98.5
P2SC	2,467	1,775	na	71.9
PLECATI	488	270	na	55.3
ACASI*	2,959	2,937	58.9	99.3
EXEC*	3,035	2,604	52.3	85.8
Wave 6 (issued sample = 4,395)				
Interview	3,537	3,537	71.0	100.0
P1CASI	3,526	3,376	67.8	95.7
P2SC	2,904	2,212	na	76.2
PLECATI	554	420	na	75.8
TEACH	3,413	2,692	na	78.9
ACASI*	3,386	3,323	66.5	98.1
CSRK	3,388	3,317	66.6	97.9
TUD*	3,387	3,071	61.6	90.7
EXEC*	3,386	3,333	66.9	98.4
GJA*	3,386	3,281	65.8	96.9

Instrument	Description
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

Parent 1 CASI

Of the families interviewed in Wave 7, only 2% of Parent 1's did not complete the P1 CASI.

Parent 2 self-completed forms

The response rate for Wave 7 Parent 2s was around 70% compared with 74% in Wave 6.

Parent Living Elsewhere (PLE) instrument

Of the eligible PLEs that interviewers attempted to contact in Wave 7 around 60% responded.

Teacher self-completed form

The teacher forms continue to achieve good response rates (over 77%) compared to 82% in Wave 6. In Wave 7 teacher forms for the B cohort were sent to the study child's English teacher as the majority of the children are at high school. Teacher forms for the K cohort were not sent for Wave 7.

Child interview

The response rate for the Time Use Diary (TUD) for the B cohort remains high at 83% but has dropped off a reasonable amount compared with 95% in Wave 6. The K cohort did not complete the TUD in Wave 7.

Instrument response rate by characteristics of families

Based on Wave 1 characteristics, the response rates to the instruments in Wave 7 were only marginally different from the full responding sample for most of the subpopulations. Larger differences in response rates are described below.

B cohort

The following differences in response were observed:

- Aboriginal and Torres Strait Islander children were under-represented across the parent interviews (F2F, PLECATI, P2SC) and the teacher questionnaire with response rates 10–25% lower than the non-Aboriginal and Torres Strait Islander sample.

- Where Parent 1 spoke a language other than English at home families had an interview response rate 9% lower than the full sample. Where Parent 1 spoke a language other than English at home, Parent 2 and the PLE had response rates 4–8% lower than the full sample.

- When combined parental income was at least $1,000pw, Parent 2 was 12% more likely and the PLE was 11% more likely to take part in an interview than when combined parental income was below $1,000pw.

- Similarly, when Parent 1 was employed, Parent 2 was 6% more likely to take part in an interview compared to when Parent 1 was not employed.

- ACT had the highest response rate to the Parent 2 form (79%); the lowest was in New South Wales (69%).

- The highest response rate to the teacher questionnaire was in Tasmania (86%); teachers in South Australia had the lowest response rate (73%).

- Study children from Tasmania had the highest response rate to the TUD (91%), while those from Queensland had the lowest (77%).

K cohort

The following differences in response were observed:

- Aboriginal and Torres Strait Islander children were under-represented across the majority of parent forms in the range of 4–26% when compared with the non-Aboriginal and Torres Strait Islander sample. However, the PLE was an exception with the Aboriginal and Torres Strait Islander children matching the non-Aboriginal and Torres Strait Islander sample at 55%.

- There were lower response rates for study families where Parent 1 spoke a language other than English at home; these families had an interview response rate 4% lower than the full sample. Where Parent 1 spoke a language other than English at home, Parent 2 response rates were 6% lower than families where Parent 1 spoke only English.

- When combined parental income was at least $1,000pw, Parent 2 and the PLE were 3–10% more likely to take part in an interview than when the combined parental income was below $1,000pw.

- Western Australia had the highest response rate to the P2 form (76%); Norther Territory had the lowest (56%).

Appendix H: B cohort non-response to forms for subpopulations

Table H1: B cohort non-response to forms

Response rate % (*n*)	F2F	P1CASI	P2SC	PLE CATI	TEACH	CSRB	ACASIB	TUD
Full sample	**78.3**	**97.4**	**71.5**	**64.1**	**77.0**	**99.6**	**99.2**	**82.3**
	(4,319)	**(3,374)**	**(2,794)**	**(507)**	**(3,333)**	**(3,238)**	**(3,238)**	**(3,238)**
Study child Indigenous	54.7	95.4	53.3	40.0	67.8	100.0	96.3	70.4
	(159)	(87)	(60)	(15)	(87)	(81)	(81)	(81)
Study child non-Indigenous	79.2	97.5	71.9	64.8	77.3	99.6	99.3	82.6
	(4,160)	(3,287)	(2,734)	(492)	(3,246)	(3,157)	(3,157)	(3,157)
Parent 1 LOTE spoken	69.4	96.8	63.7	60.0	73.1	99.7	98.7	85.3
	(546)	(375)	(342)	(30)	(386)	(373)	(373)	(373)
Parent 1 English only	79.6	97.5	72.6	64.4	77.1	99.5	99.3	81.9
	(3,773)	(2,999)	(2,452)	(477)	(2,965)	(2,865)	(2,865)	(2,865)
Parent 1 employed	82.0	98.2	74.1	66.8	78.0	99.5	99.2	82.3
	(2,244)	(1,836)	(1,561)	(268)	(1,816)	(1,766)	(1,766)	(1,766)
Parent 1 not employed	74.3	96.5	68.3	76.9	75.9	99.7	99.2	82.3
	(2,067)	(1,531)	(1,226)	(238)	(1,510)	(1,465)	(1,465)	(1,465)
Parental income < $1,000	73.4	97.4	65.4	58.1	76.1	99.7	99.3	81.2
	(1,770)	(1,294)	(1,004)	(234)	(1,274)	(1,252)	(1,252)	(1,252)
Parental income >= $1,000	80.6	97.7	78.0	69.0	77.3	99.6	99.2	82.8
	(2,390)	(1,925)	(1,616)	(255)	(1,905)	(1,840)	(1,840)	(1,840)
NSW	76.6	97.2	68.6	66.9	74.3	99.9	99.4	84.4
	(1,350)	(1,031)	(873)	(142)	(1,016)	(988)	(988)	(988)
VIC.	76.4	97.6	72.5	58.1	79.4	99.6	99.1	80.8
	(1,036)	(789)	(672)	(86)	(780)	(760)	(760)	(760)
QLD	80.4	97.3	71.6	63.2	77.9	99.4	99.1	76.8
	(890)	(716)	(573)	(117)	(705)	(684)	(684)	(684)
SA	77.7	99.2	74.6	61.8	72.9	99.6	100.0	81.6
	(323)	(251)	(205)	(55)	(251)	(245)	(245)	(245)
WA	80.9	96.9	73.3	66.7	78.5	98.8	99.4	87.3
	(444)	(358)	(288)	(63)	(354)	(339)	(339)	(339)
TAS.	84.9	94.4	72.6	73.3	85.6	100.0	95.4	90.8
	(106)	(90)	(73)	(15)	(90)	(87)	(87)	(87)
ACT	80.2	100.0	78.8	53.3	78.5	98.7	100.0	85.7
	(101)	(80)	(66)	(15)	(79)	(77)	(77)	(77)
NT	85.5	96.6	77.3	78.6	75.9	100.0	98.3	86.2
	(69)	(59)	(44)	(14)	(58)	(58)	(58)	(58)
Capital city	78.7	97.4	73.1	61.4	76.5	99.6	99.4	83.7
	(2,763)	(2,169)	(1,807)	(311)	(2,143)	(2,077)	(2,077)	(2,077)
Rest of state	77.6	97.4	68.8	68.2	77.9	99.7	98.8	79.8
	(1,545)	(1,197)	(980)	(195)	(1,183)	(1,153)	(1,153)	(1,153)
Study child male	78.3	97.7	71.7	62.8	77.2	99.5	98.9	81.5
	(2,217)	(1,733)	(1,435)	(269)	(1,707)	(1,667)	(1,667)	(1,667)
Study child female	78.3	97.1	71.4	65.5	76.8	99.6	99.6	83.1
	(2,102)	(1,641)	(1,359)	(238)	(1,626)	(1,571)	(1,571)	(1,571)

Appendix I: K cohort non-response to forms for subpopulations

Table I1: K cohort non-response to forms

Response rate % (*n*)	F2F	P1 CASI	P2SC	PLE CATI	TEACH	CSRK	ACASI	TUD
Full sample	**74.0**	**98.5**	**71.9**	**55.3**	**na**	**na**	**99.2**	**na**
	(4,175)	**(3,048)**	**(2,467)**	**(488)**	**na**	**na**	**(2,959)**	**na**
Study child Indigenous	54.3	94.1	46.7	55.6	na	na	92.5	na
	(129)	(68)	(45)	(18)	na	na	(67)	na
Study child non-Indigenous	74.6	98.6	72.4	55.4	na	na	99.4	na
	(4,044)	(2,978)	(2,421)	(469)	na	na	(2,890)	na
Parent 1 LOTE spoken	70.0	95.8	67.0	57.1	na	na	100.0	na
	(563)	(384)	(324)	(35)	na	na	(376)	na
Parent 1 English only	74.6	98.9	72.7	55.2	na	na	99.1	na
	(3,612)	(2,664)	(2,143)	(453)	na	na	(2,583)	na
Parent 1 employed	76.4	99.1	73.6	56.3	na	na	99.6	na
	(2,498)	(1,889)	(1,551)	(320)	na	na	(1,830)	na
Parent 1 not employed	70.4	97.6	69.1	53.6	na	na	98.7	na
	(1,673)	(1,156)	(914)	(166)	na	na	(1,126)	na
Parental income < $1,000	66.8	97.3	64.9	54.2	na	na	99.2	na
	(1,504)	(983)	(680)	(216)	na	na	(949)	na
Parental income >= $1,000	79.2	99.3	75.7	57.7	na	na	99.2	na
	(2,431)	(1,909)	(1,665)	(246)	na	na	(1,860)	na
NSW	74.2	98.2	72.2	38.1	na	na	99.7	na
	(1,296)	(943)	(780)	(168)	na	na	(927)	na
VIC.	69.5	98.9	72.3	45.9	na	na	99.4	na
	(1,022)	(703)	(566)	(135)	na	na	(666)	na
QLD	76.4	97.7	69.5	49.6	na	na	99.2	na
	(822)	(621)	(491)	(121)	na	na	(605)	na
SA	72.1	99.1	72.3	52.3	na	na	99.5	na
	(298)	(214)	(173)	(44)	na	na	(209)	na
WA	77.4	99.4	76.3	69.6	na	na	98.5	na
	(434)	(331)	(278)	(46)	na	na	(323)	na
TAS.	85.5	98.1	75.3	76.2	na	na	96.1	na
	(124)	(105)	(73)	(21)	na	na	(102)	na
ACT	76.7	100.0	71.6	85.7	na	na	100.0	na
	(103)	(79)	(67)	(7)	na	na	(78)	na
NT	69.7	100.0	56.4	58.3	na	na	100.0	na
	(76)	(52)	(39)	(12)	na	na	(49)	na
Capital city	74.7	98.4	72.9	60.9	na	na	99.4	na
	(2,609)	(1,923)	(1,578)	(276)	na	na	(1,864)	na
Rest of state	72.9	98.7	70.3	48.1	na	na	99.0	na
	(1,559)	(1,122)	(886)	(212)	na	na	(1,091)	na
Study child male	73.7	98.7	72.2	53.1	na	na	99.2	na
	(2,141)	(1,551)	(1,258)	(258)	na	na	(1,505)	na
Study child female	74.3	98.3	71.7	57.8	na	na	99.3	na
	(2,034)	(1,497)	(1209)	(230)	na	na	(1,454)	na

Note: na = Not collected in Wave 7 from K cohort.