Australian Government

Department of Social Services

*Growing Up in Australia:*

The Longitudinal Study of Australian Children (LSAC)

LSAC Technical Paper No. 15


# Wave 6 Weighting and Non-Response

Amanda Norton and Kevin Monahan

May 2015

# Technical Paper

The Longitudinal Study of Australian Children: LSAC Technical paper No.15, Wave 6 Weighting and Non-Response

Authors: Amanda Norton and Kevin Monahan
Australian Bureau of Statistics



**For more information, write to:**

National Centre for Longitudinal Data
Policy Evidence Branch
Australian Government Department of Social Services
PO Box 7576
Canberra Business Centre ACT 2610

Email: NCLD@dss.gov.au

# Contents

# About the authors

Amanda Norton is an Assistant director in the Household Survey Methodology section of the Australian Bureau of Statistics.

Kevin Monahan is a Research Officer in the Household Survey Methodology section of the Australian Bureau of Statistics.

# Acknowledgements

*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 and the Australian Bureau of Statistics, with advice provided by a consortium of leading researchers.

The authors wish to thank **Emily Cunningham** and **Melissa Gare** for review and technical advice in the compilation of this weighting paper. The paper also draws on the work of Ryan Defina, Benedict Cusack, Ben Ingram and Daniel Fearnley, who have all contributed to developing these weighting methods.

# 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 ('baby') cohort, who were aged 0–1 years at the beginning of the study (born between March 2003 and February 2004); and
- the K ('kindergarten') 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 6 of the Longitudinal Study of Australian Children was conducted in 2014 with B-cohort children at age 10–11 years and K-cohort children at age 14–15 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 (for example, moving overseas). Some children are brought back into the sample after missing a wave if contact can be re-established (for example, 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, 7,301 children responded in Wave 6, and 6,717 children responded to all waves.

This weighting paper serves two purposes: describing the response properties and quality of the sample continuing into Wave 6; and describing the method and implementation of weight calculations to assist analysts make accurate population inferences from the LSAC sample. The method of producing weights is unchanged from Wave 5; however, the response propensity models used to adjust the weights have been re-fit.

## 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 single-year 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 6 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). We provide a summary here for reference.

**Table 1: LSAC sample design properties**

| Property | Description |
|---|---|
| **Target population** (whom the study is about) | Children growing up in Australia |
| **Scope** (the population about which inference is to be made) | Two single-year cohorts of children (B-cohort babies and K-cohort kindergarteners who were 0–1 years and 4–5 years old respectively during the Wave 1 recruitment year in 2004. Scope excluded very remote areas of Australia. |
| **Coverage** (the population represented by the active participating sample) | For Wave 1 recruitment: The subset of Wave 1 scope for whom contact records were available through Medicare, who could be contacted, and who agreed to participate in LSAC.<br>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<br>Stage 2 Unit: 1 Cluster of dwellings within postcode<br>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. |

# Summary of weighting in Waves 1–5

Weights for Wave 1 were calculated beginning with the inverse probability of selection for each child, and adjusting these weights to align to known population benchmarks (Soloff et al., 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 to 5 were calculated by adjusting previous wave weights for differential sample attrition in two stages (Cusack & Defina, 2014; Sipthorp & Misson, 2007, 2009; Sipthorp & Daraganova, 2011). 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 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 changes in the Wave 6 weighting

Wave 6 uses the same two-stage weighting method as Wave 5. The only difference is that the response propensity models have been created based on the Wave 6 responses, and are thus different to the models used for Wave 5.

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

- Cross-sectional weight model—uses all children from Wave 1 and Wave 1 data items to predict response propensity in Wave 6;
- Longitudinal weight model—uses 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.

The only change for the cross-sectional response propensity models is the addition of the variable indicating whether Parent 2 has returned the self-completed questionnaire (or a separate category if there is no Parent 2).

The longitudinal response propensity models are now restricted to using the data of only those children who had responded to all waves up to and including Wave 5, instead of using all children from Wave 5.

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

# Wave 6 weighting method

This section contains a brief description of the method used to create weights for Wave 6 data. The method is largely unchanged from Wave 5. For more detail refer to the LSAC Technical Paper no. 10 'Wave 5 weighting & non response' by Benedict Cusack and Ryan Defina (2014).

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 starting from the longitudinal weight from the previous wave of the study, and adjusting for any additional non-response to the current wave.

Cross-sectional weights begin with the final weight used in Wave 1 and adjust for all additional non-response to the current wave—regardless of whether a unit responded to Waves 2–5.

## Initial weights

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

- For Wave 6 longitudinal weights (which applies to those who have responded to all Waves 1, 2, 3, 4, 5 and 6), the initial weight for children in Wave 6 is the final longitudinal weight from Wave 5.
- For Wave 6 cross-sectional weights (which applies to all of those who responded in Wave 6), the initial weight for children in Wave 6 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 as its initial weight divided by 0.9.

### Selection of covariates for logistic regression non-response adjustment

The starting point for the response propensity models was to use the same model covariates from Wave 5 to achieve consistency over time; however, an investigation was undertaken to see if there were any significant changes that should be made to these models. This investigation involved using a model selection process to determine potential covariates to add to the model and then comparing combinations of these covariates with the Wave 5 variables to determine a model with strong consistency to Wave 5 but allowed significant improvements if any were identified. The two steps to this investigation are explained in more detail below.

Selecting covariates to use in the response propensity models involved using a stepwise model selection process using all possible covariates considered 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 ($p<0.05$) than they are removed from the model. These model selection processes resulted in a shortlist of variables to consider adding to the Wave 5 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 5 variables. Wave 5 variables that were clearly no longer useful or significant ($p>0.1$) were removed from the model. The other Wave 5 variables were maintained where possible to achieve consistency over time. New covariates were chosen by taking the combination with Wave 5 variables that resulted in the lowest Akaike Information Criterion.

### Wave 1 variables used in the cross-sectional weight models for both cohorts

- 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 (new).

### Wave 5 variables used in the B cohort longitudinal weight model

- Parent 1 age
- Mother's highest level of high school completed
- Parent 2 self-completed questionnaire returned
- Overall school achievement—teacher reported (new)
- Parent 1 housing tenure (new).

### Wave 5 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
- Reading ability—teacher reported (new)
- Parent 1 renting home indicator (new)
- Number of days each week someone in household helps study child with homework (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 re-align the sample composition within each stratum to the composition within each stratum as at Wave 1, and to re-align the sum of sample weights to be equal to the sum of the current sample size, such that the average sample weight is 1. The original selections were done by dividing each state into 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-response 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 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 have been 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.

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. As with previous waves, the weight caps have been applied to the sample weights, with a lower cap of 0.33, and an upper cap of 2.5.

# 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 6.

For the B cohort, out of 1,022 that did not respond to Wave 5, 89 responded to Wave 6. Out of the 1,349 that did not respond to at least one of Waves 2, 3, 4 or 5, 323 responded to Wave 6.

For the K cohort, out of 1,027 that did not respond to Wave 5, 77 responded to Wave 6. Out of the 1,301 that did not respond to at least one of Waves 2, 3, 4 or 5, 261 responded to Wave 6.

Table 2 below shows those who have responded after dropping out in a previous wave (sample reacquisition).

**Table 2: 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 |
|---|---|---|---|---|
| B | 133 | 135 | 129 | 89 |
| K | 135 | 119 | 94 | 77 |

## 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 3 and 4 below for updated counts. Note the total sample approached was 8,921 for the B cohort and 9,893 for the K cohort (including non-contacts).

**Table 3: Sample counts for the B cohort**

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

**Table 4: Sample counts for the K cohort**

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

- 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, i.e. fully responding to each wave since Wave 1.
- Cross-sectional attrition rate (%)—those not responding to that particular as a percentage of the Wave 1 sample.
- 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

Table 5 below shows the number of children with a sample weight at the lower cap of 0.33 and upper cap of 2.5 by cohort and by type of weight. These counts of units with weights at the caps have generally increased since Wave 5, especially for the cross-sectional weights. This is a result of decreasing response for some characteristics.

For the B cohort, the number of units at the upper cap has increased from 66 in Wave 5 to 116 for the cross-sectional weight, and decreased from 149 in Wave 5 to 142 for the longitudinal weight.

**Table 5: Counts of capped weights for Wave 6—B cohort**

| | Cross-sectional | | Longitudinal | |
|---|---|---|---|---|
| State | Lower cap (0.33) | Upper cap (2.5) | Lower cap (0.33) | Upper cap (2.5) |
| NSW | 0 | 54 | 0 | 63 |
| VIC | 0 | 25 | 0 | 31 |
| QLD | 18 | 16 | 18 | 20 |
| SA | 2 | 8 | 1 | 9 |
| WA | 2 | 8 | 3 | 10 |
| TAS | 3 | 3 | 6 | 5 |
| NT | 13 | 0 | 17 | 0 |
| ACT | 0 | 2 | 0 | 4 |
| AUS | 38 | 116 | 45 | 142 |

For the K cohort, the number of units at the upper cap has increased from 40 in Wave 5 to 74 for the cross-sectional weight, and increased from 92 in Wave 5 to 121 for the longitudinal weight.

**Table 6: Counts of capped weights for Wave 6—K cohort**

| | Cross-sectional | | Longitudinal | |
|---|---|---|---|---|
| State | Lower cap (0.33) | Upper cap (2.5) | Lower cap (0.33) | Upper cap (2.5) |
| NSW | 0 | 32 | 0 | 51 |
| VIC | 0 | 20 | 0 | 32 |
| QLD | 0 | 14 | 0 | 25 |
| SA | 0 | 4 | 0 | 6 |
| WA | 0 | 3 | 0 | 4 |
| TAS | 9 | 1 | 20 | 2 |
| NT | 31 | 0 | 20 | 0 |
| ACT | 3 | 0 | 4 | 1 |
| AUS | 43 | 74 | 44 | 121 |

# Conclusion

Sample attrition has increased in this wave; however, the responding sample 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 more weights at the weight caps, due to the increased non-response in this wave. 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 (ie. that particular characteristics cause non-response). The models reflect the observed response patterns and the weights developed ensure that the change in sample composition can be adjusted for in any analysis.

# Bibliography

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

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

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

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

Engle, R. (1983). *Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics*. Handbook of Econometrics II. Elsevier. Pp. 796–801

Fairclough, D.L. (2010). *Design and analysis of quality of life studies in clinical trials*. 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, DB. (1987). *Statistical analysis with missing data*. Vol. 539. New York: Wiley.

Little, R.J.A. (1986). Survey Nonresponse Adjustments for Estimates of Means. *International Statistical Review, vol. 54*, pp. 139–157.

Pfeffermann, D. (1993). The Role of Sampling Weights when Modelling Survey Data. *International Statistical Review, vol. 61*, pp 317–337.

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

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

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

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

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

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

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

# 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.

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

# Appendix B: Description of Wave 6 weights

**Table B1: Description of Wave 6 weights**

| SAS name | Cohort | Type | Waves cases responded to |
|---|---|---|---|
| fweight | B | Population | 1 & 6 |
| fweights | B | Sample | 1 & 6 |
| bcdefwt | B | Population | 1, 2, 3, 4, 5 & 6 |
| bcdefwts | B | Sample | 1, 2, 3, 4, 5 & 6 |
| hweight | K | Population | 1 & 6 |
| hweights | K | Sample | 1 & 6 |
| defghwts | K | Population | 1, 2, 3, 4, 5 & 6 |
| defghwt | K | Sample | 1, 2, 3, 4, 5 & 6 |

# 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 | 13.9 | 0.0002 |
| AF03M3 | Parent 2 age | 1 | 21.9 | <0.0001 |
| AFD08M1 | Mother's highest year of high school completed | 4 | 112.2 | <0.0001 |
| AFD11M2 | Mother's proficiency in spoken English | 4 | 87.3 | <0.0001 |
| AP1SCD | Parent 1 self-completed questionnaire returned | 1 | 19.8 | <0.0001 |
| AP2SCD | Parent 2 self-completed questionnaire returned | 2 | 26.5 | <0.0001 |

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 |
|---|---|---|---|---|
| EF03EP1 | Parent 1 age | 1 | 3.7 | 0.0559 |
| EFD08M1 | Mother's highest year of high school completed | 3 | 11.5 | 0.0094 |
| EP2SCD | Parent 2 self-completed questionnaire returned | 2 | 42.4 | <0.0001 |
| ELC08T3B | Overall school achievement of study child (teacher reported) | 6 | 24.0 | 0.0005 |
| EHO04A5 | Parent 1 housing tenure | 3 | 21.0 | 0.0001 |

**Table C3: K cohort—cross-sectional weights**

| Variable name | Description | DF | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|
| CF03M2 | Parent 1 age | 1 | 20.4 | <0.0001 |
| CF03M3 | Parent 2 age | 1 | 10.7 | 0.0011 |
| CFD08M1 | Mother's highest year of high school completed | 4 | 90.1 | <0.0001 |
| CFD11M2 | Mother's proficiency in spoken English | 4 | 63.5 | <0.0001 |
| CP1SCD | Parent 1 self-completed questionnaire returned | 1 | 10.4 | 0.0013 |
| CP2SCD | Parent 2 self-completed questionnaire returned | 2 | 55.4 | <0.0001 |

**Table C4: K cohort—longitudinal weights**

| Variable name | Description | DF | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|
| GF03GP1 | Parent 1 age | 1 | 8.0 | 0.0048 |
| GFD08M1 | Mother's highest year of high school completed | 4 | 13.0 | 0.0111 |
| GP2SCD | Parent 2 self-completed questionnaire returned | 2 | 70.5 | <0.0001 |
| GLC08T1B | Reading ability (teacher reported) | 6 | 20.7 | 0.0021 |
| GHO04A3B | Parent 1 rents home | 2 | 15.8 | 0.0004 |
| GHE11A3E | Number of days household helps study child with homework each week | 5 | 22.7 | 0.0004 |

# Appendix D: Odds ratio estimates for variables in Wave 6 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 | 1.520 | 1.276 | 1.811 |
| af03m2 | Parent 1 age | 1.176 | 1.080 | 1.281 |
| afd08m1 1 vs 5 | Mother completed Year 12 or equivalent | 2.470 | 1.530 | 3.986 |
| afd08m1 2 vs 5 | Mother completed Year 11 or equivalent | 1.325 | 0.799 | 2.197 |
| afd08m1 3 vs 5 | Mother completed Year 10 or equivalent | 1.188 | 0.726 | 1.942 |
| afd08m1 4 vs 5 | Mother completed Year 9 or equivalent | 0.817 | 0.463 | 1.441 |
| afd11m2 0 vs 4 | Not Applicable to Mother's proficiency in spoken English | 4.066 | 1.657 | 9.979 |
| afd11m2 1 vs 4 | Mother speaks English Very well | 1.938 | 0.775 | 4.841 |
| afd11m2 2 vs 4 | Mother speaks English Well | 1.333 | 0.516 | 3.445 |
| afd11m2 3 vs 4 | Mother speaks English Not well | 2.295 | 0.858 | 6.141 |
| ap1scd 0 vs 1 | Parent 1 did not return self-completed questionnaire | 0.591 | 0.469 | 0.745 |
| ap2scd -9 vs 1 | No Parent 2 in household | 1.169 | 0.690 | 1.981 |
| ap2scd 0 vs 1 | Parent 2 did not return self-completed questionnaire | 0.568 | 0.455 | 0.710 |

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

| Effect | Description | Point estimate | 95% Wald confidence limits | |
|---|---|---|---|---|
| ef03ep1 | Parent 1 age | 1.112 | 0.997 | 1.240 |
| efd08m1 1 vs 4 | Mother completed Year 12 or equivalent | 2.073 | 1.212 | 3.547 |
| efd08m1 2 vs 4 | Mother completed Year 11 or equivalent | 1.774 | 0.950 | 3.315 |
| efd08m1 3 vs 4 | Mother completed Year 10 or equivalent | 1.401 | 0.792 | 2.478 |
| ep2scd -9 vs 1 | No Parent 2 in household | 0.594 | 0.415 | 0.851 |
| ep2scd 0 vs 1 | Parent 2 did not return self-completed questionnaire | 0.407 | 0.311 | 0.534 |
| elc08t3b -9 vs 5 | Not Applicable overall school achievement | 0.522 | 0.202 | 1.346 |
| elc08t3b 0 vs 5 | Missing overall school achievement | 0.400 | 0.113 | 1.416 |
| elc08t3b 1 vs 5 | Overall school achievement Far below average | 0.553 | 0.188 | 1.621 |
| elc08t3b 2 vs 5 | Overall school achievement Below average | 0.460 | 0.178 | 1.188 |
| elc08t3b 3 vs 5 | Overall school achievement Average | 0.835 | 0.330 | 2.116 |
| elc08t3b 4 vs 5 | Overall school achievement Above Average | 0.974 | 0.378 | 2.509 |
| eho04a5 1 vs 6 | Parent 1 housing tenure: Being paid off by P1 and/or partner | 3.107 | 1.797 | 5.372 |
| eho04a5 2 vs 6 | Parent 1 housing tenure: Owned outright by P1 and/or partner | 3.904 | 2.032 | 7.502 |
| eho04a5 3 vs 6 | Parent 1 housing tenure: Rented or boarded by P1 and/or partner | 2.283 | 1.308 | 3.987 |

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

| Effect | Description | Point estimate | 95% Wald confidence limits | |
|---|---|---|---|---|
| cf03m2 | Parent 1 age | 1.199 | 1.108 | 1.297 |
| cf03m3 | Parent 2 age | 1.405 | 1.146 | 1.722 |
| cfd08m1 1 vs 5 | Mother completed Year 12 or equivalent | 2.926 | 1.949 | 4.394 |
| cfd08m1 2 vs 5 | Mother completed Year 11 or equivalent | 1.810 | 1.177 | 2.785 |
| cfd08m1 3 vs 5 | Mother completed Year 10 or equivalent | 1.693 | 1.115 | 2.572 |
| cfd08m1 4 vs 5 | Mother completed Year 9 or equivalent | 1.103 | 0.681 | 1.786 |
| cfd11m2 0 vs 4 | Not Applicable to Mother's proficiency in spoken English | 1.437 | 0.651 | 3.173 |
| cfd11m2 1 vs 4 | Mother speaks English Very well | 0.874 | 0.388 | 1.967 |
| cfd11m2 2 vs 4 | Mother speaks English Well | 0.498 | 0.217 | 1.145 |
| cfd11m2 3 vs 4 | Mother speaks English Not well | 0.900 | 0.377 | 2.148 |
| cp1scd 0 vs 1 | Parent 1 did not return self-completed questionnaire | 0.697 | 0.560 | 0.868 |
| cp2scd -9 vs 1 | No Parent 2 in household | 0.949 | 0.549 | 1.640 |
| cp2scd 0 vs 1 | Parent 2 did not return self-completed questionnaire | 0.447 | 0.362 | 0.554 |

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

| Effect | Description | Point estimate | 95% Wald confidence limits | |
|---|---|---|---|---|
| gf03gp1 | Parent 1 age | 1.157 | 1.046 | 1.281 |
| gfd08m1 1 vs 5 | Mother completed Year 12 or equivalent | 2.930 | 1.604 | 5.352 |
| gfd08m1 2 vs 5 | Mother completed Year 11 or equivalent | 2.393 | 1.257 | 4.555 |
| gfd08m1 3 vs 5 | Mother completed Year 10 or equivalent | 2.653 | 1.423 | 4.947 |
| gfd08m1 4 vs 5 | Mother completed Year 9 or equivalent | 2.322 | 1.041 | 5.181 |
| gp2scd -9 vs 1 | No Parent 2 in household | 0.412 | 0.305 | 0.556 |
| gp2scd 0 vs 1 | Parent 2 did not return self-completed questionnaire | 0.358 | 0.278 | 0.460 |
| glc08t1b -9 vs 5 | Not Applicable language and literacy skills | 0.408 | 0.192 | 0.870 |
| glc08t1b 0 vs 5 | Missing language and literacy skills | 0.425 | 0.170 | 1.059 |
| glc08t1b 1 vs 5 | Language and literacy skills Far below average | 0.315 | 0.123 | 0.802 |
| glc08t1b 2 vs 5 | Language and literacy skills Below average | 0.489 | 0.222 | 1.074 |
| glc08t1b 3 vs 5 | Language and literacy skills Average | 0.615 | 0.290 | 1.301 |
| glc08t1b 4 vs 5 | Language and literacy skills Above average | 0.725 | 0.341 | 1.541 |
| gho04a3b -9 vs 2 | Not Applicable to Parent 1 renting home | 0.242 | 0.119 | 0.494 |
| gho04a3b 1 vs 2 | Parent 1 renting home | 0.847 | 0.649 | 1.104 |
| ghe11a3e -9 vs 5 | Not applicable to days per week helping with homework | 1.594 | 0.929 | 2.736 |
| ghe11a3e 1 vs 5 | Someone in household helps with homework 5 or more days per week | 1.283 | 0.795 | 2.071 |
| ghe11a3e 2 vs 5 | Someone in household helps with homework 3 or 4 days per week | 2.091 | 1.330 | 3.288 |
| ghe11a3e 3 vs 5 | Someone in household helps with homework 1 or 2 days per week | 2.149 | 1.412 | 3.269 |
| ghe11a3e 4 vs 5 | Someone in household helps with homework less than once per week | 2.264 | 1.442 | 3.556 |

# 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. of people in household |
| ansib | 0/1—No. of 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 |

| Table E2: Wave 5 data items considered for B cohort—longitudinal weight | |
|---|---|
| **Variable name** | **Variable label** |
| eahactd | 8/9—P1—Home activities index (v4) |
| eahacte | 8/9—P1—Home activities index (v5) |
| ecnfsad2 | 8/9—SEIFA—Advantage/Disadvantage—2011—SA2—Score |
| ecnfsad2d | 8/9—SEIFA—Advantage/Disadvantage—2011—SA2—Deciles—National |
| ecnfser2 | 8/9—Home—SEIFA Economic Resources—2011—SA2—Score |
| ecnfser2d | 8/9—Home—SEIFA Economic Resources—2011—SA2—Deciles—National |
| ef01em | M@8/9—Present for wave |
| ef01ep2 | P2@8/9—Present for wave |
| ef03ep1 | P1@8/9—Age |
| ef03ep2 | P2@8/9—Age |
| ef11em | M@8/9—Language other than English spoken at home |
| ef11ep1 | P1@8/9—Language other than English spoken at home |
| ef11m1 | 8/9—SC—Main language spoken at home |
| efd08a1 | 8/9—P1—F2F A1.1/A1.2/A1.3+W1-4—School completion |
| efd08a2a | 8/9—P1—F2F A1.2/A1.3+W1-4—Completed other qualification |
| efd08a3a | 8/9—P1—F2F A1.2/A1.3+W1-4—Highest qualification |
| efd08m1 | 8/9—M—F2F A1.1/A1.2/A1.3+W1-4—School completion |
| efemp | 8/9—F—Employment status |
| eho04a1 | 8/9—P1—F2F J3—Housing tenure |
| eho04a3b | 8/9—P1—F2F P1.6.2—Rent home |
| eho04a5 | 8/9—P1—F2F J3—Housing tenure |
| elc08t3b | 8/9—T/C—Teach 22.3—Overall school achievement |
| ematreas | 8/9—Matrix reasoning |
| ememp | 8/9—M—Employment status |
| enpeople | 8/9—No. of people in household |
| ensib | 8/9—No. of siblings of SC in household |
| ep2 | 8/9—SC has 2 parents in the home |
| ep2scd | 8/9—Parent 2 self-completed data present |
| zf02ep1 | P1@8/9—Sex |
| zf09ep1 | P1@8/9—Country of birth |
| zf12ep1 | P1@8/9—Indigenous status |
| ehe11a3e | 8/9—P1—F2F C6.2—How often help child with homework |
| ehb24a | 8/9—Teach 16—Activity during organised activities |
| ehe09a | 8/9—F2F M8.1—Extra curricular—any |

| 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. of people in household |
| cnsib | 4/5—No. of 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 |

**Table E4: Wave 5 data items considered for K cohort—longitudinal weight**

| Variable name | Variable label |
|---|---|
| gcnfsad2 | 12/13—SEIFA—Advantage/Disadvantage—2011—SA2—Score |
| gcnfsad2d | 12/13—SEIFA—Advantage/Disadvantage—2011—SA2—Deciles—National |
| gcnfser2 | 12/13—Home—SEIFA Economic Resources—2011—SA2—Score |
| gcnfser2d | 12/13—Home—SEIFA Economic Resources—2011—SA2—Deciles—National |
| gf01gm | M@12/13—Present for wave |
| gf01gp2 | P2@12/13—Present for wave |
| gf03gp1 | P1@12/13—Age |
| gf03gp2 | P2@12/13—Age |
| gf11gm | M@12/13—Language other than English spoken at home |
| gf11gp1 | P1@12/13—Language other than English spoken at home |
| gf11m1 | 12/13—SC—Main language spoken at home |
| gfd08a1 | 12/13—P1—F2F A1.1/A1.2/A1.3+W1-4—School completion |
| gfd08m1 | 12/13—M—F2F A1.1/A1.2/A1.3+W1-4—School completion |
| gfemp | 12/13—F—Employment status |
| ghe02a6d | 12/13—P1—F2F M1.5—Everyday activities with SC |
| gho04a3b | 12/13—P1—F2F P1.6.2—Rent home |
| gho04a5 | 12/13—P1—F2F P1—Housing tenure |
| glc08t1b | 12/13—T/C—Teach 17—Reading progress |
| gmemp | 12/13—M—Employment status |
| gnpeople | 12/13—No. of people in household |
| gnsib | 12/13—No. of siblings of SC in household |
| gp2 | 12/13—SC has 2 parents in the home |
| gp2scd | 12/13—Parent 2 self-completed data present |
| zf02gp1 | P1@12/13—Sex |
| zf09gp1 | P1@12/13—Country of birth |
| zf12gp1 | P1@12/13—Indigenous status |
| ghe13a | 12/13—F2F C7.0—How far SC will go in education |
| glc08a1a | 12/13—P1—F2F C7.1—Reading progress |
| glc08a2a | 12/13—P1—F2F C7.2—Maths progress |
| glc08a3a | 12/13—P1—F2F C7.3—Overall school achievement |
| ghe11a3e | 12/13—P1—F2F C6.2—How often help child with homework |
| ghe09a1 | 12/13—F2F M8.1—Extra curricular—any |

# 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

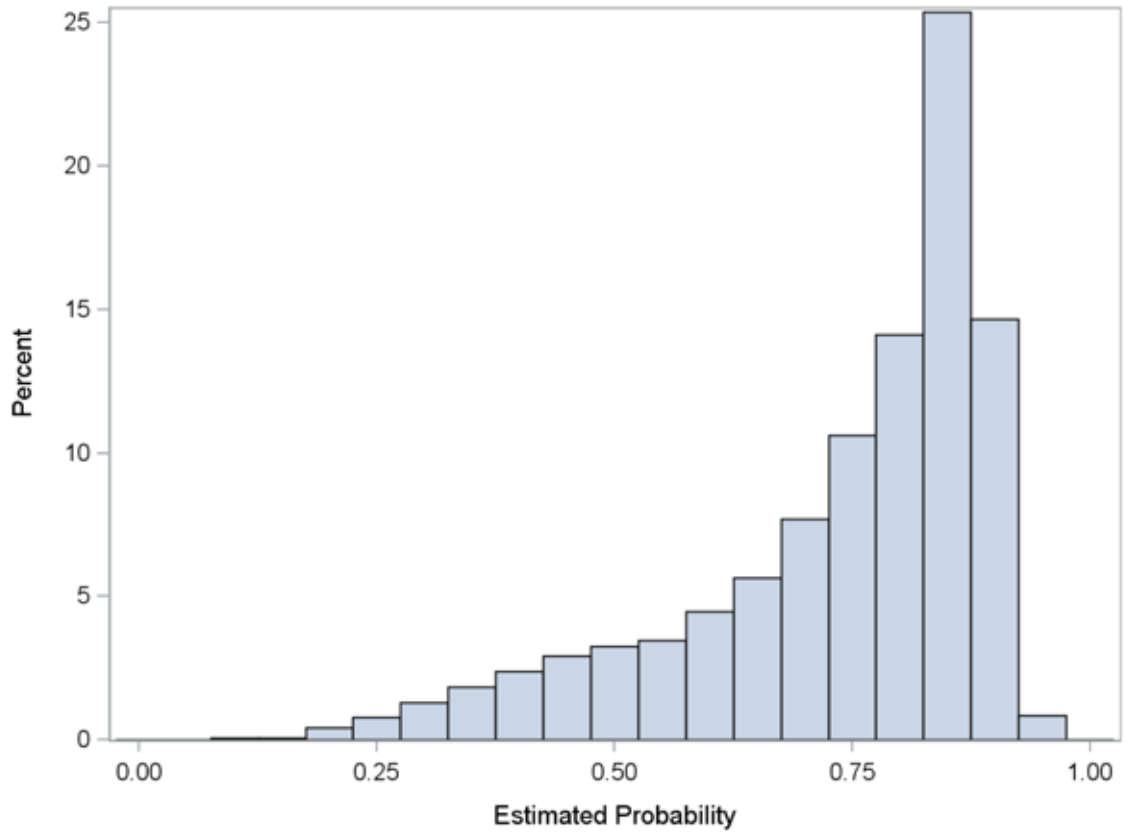


**Figure F1: Distribution of estimated response propensities—B cohort cross-sectional weight**

**Table F1: Analysis variable: estimated probability—B cohort cross-sectional weight**

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | N |
|---|---|---|---|---|---|---|---|
| 0.7370263 | 0.1622659 | 0.0836033 | 0.9510988 | 0.8560761 | 0.8674956 | 3763.99 | 5107 |

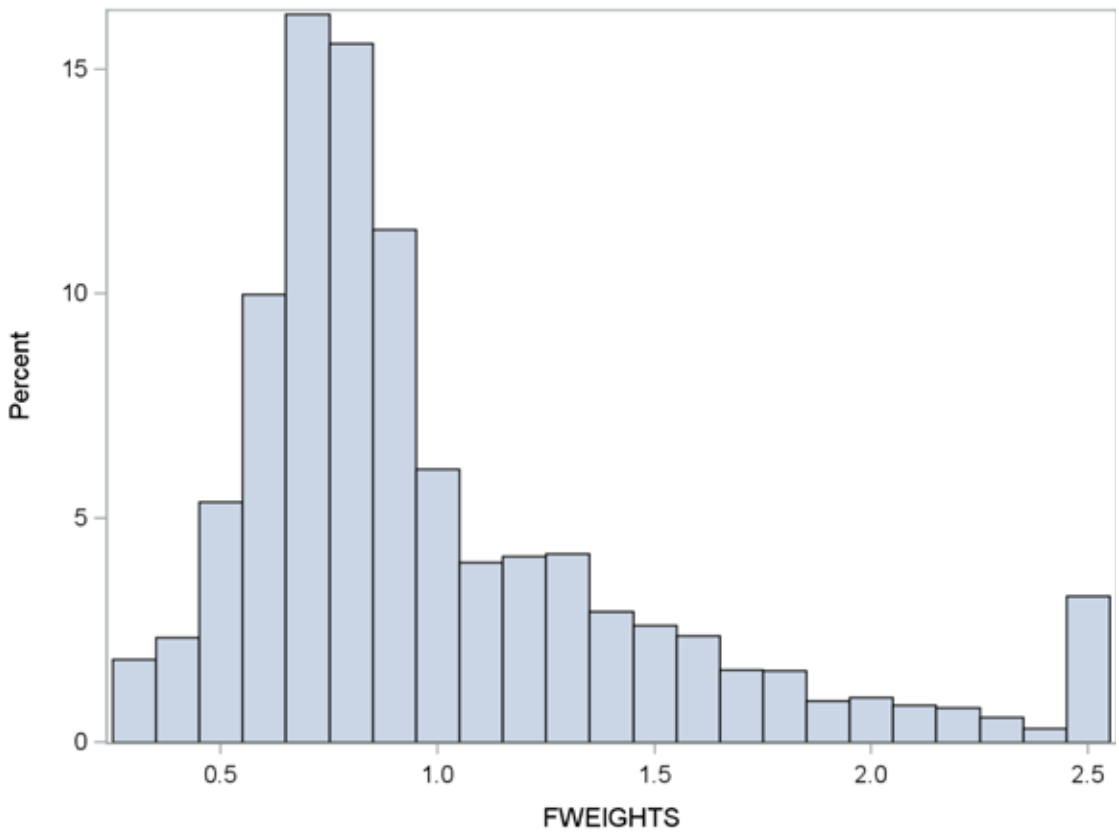


**Figure F2: Distribution of final sample weight for Wave 6—B cohort cross-sectional weight**

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

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | N |
|---|---|---|---|---|---|---|---|
| 1.0000000 | 0.4892489 | 0.3300000 | 2.5000000 | 2.5000000 | 2.1700000 | 3764.00 | 3764 |

## B cohort—longitudinal weight

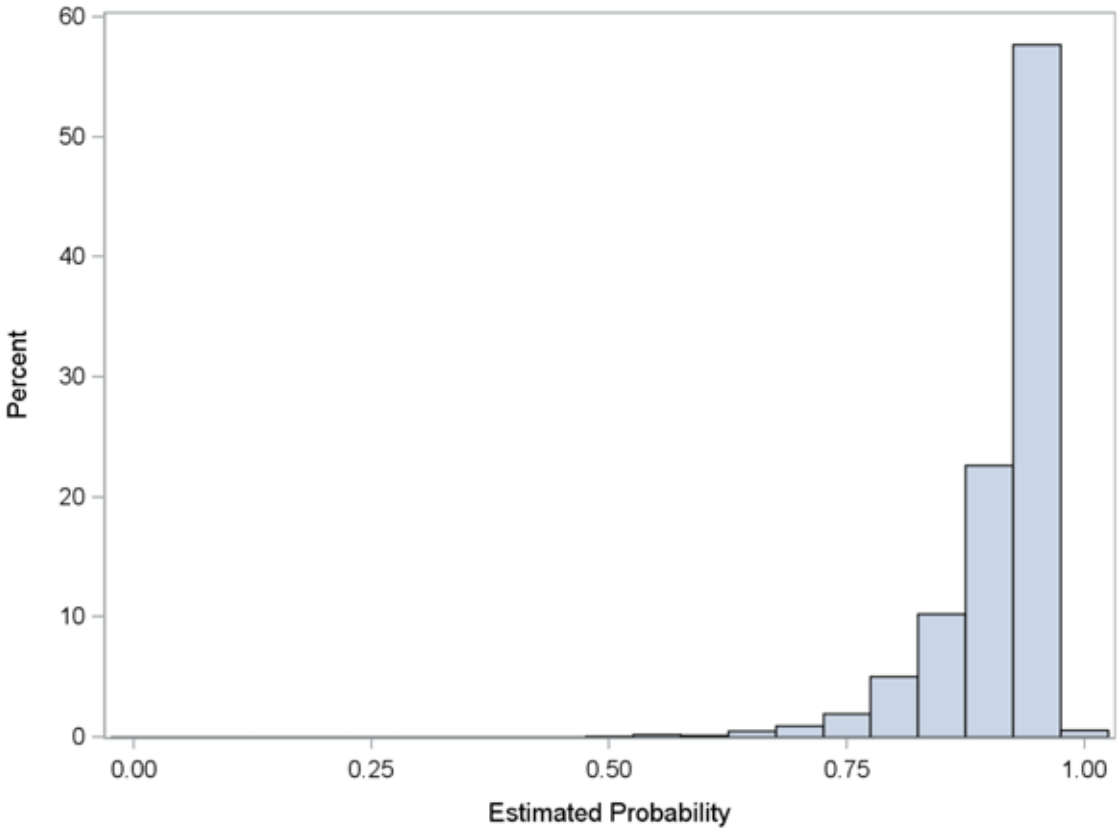


**Figure F3: Distribution of estimated response propensities—B cohort longitudinal weight**

**Table F3: Analysis variable: estimated probability—B cohort longitudinal weight**

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | N |
|---|---|---|---|---|---|---|---|
| 0.9156466 | 0.0626117 | 0.5055649 | 0.9801787 | 0.9591276 | 0.4746139 | 3441.00 | 3758 |

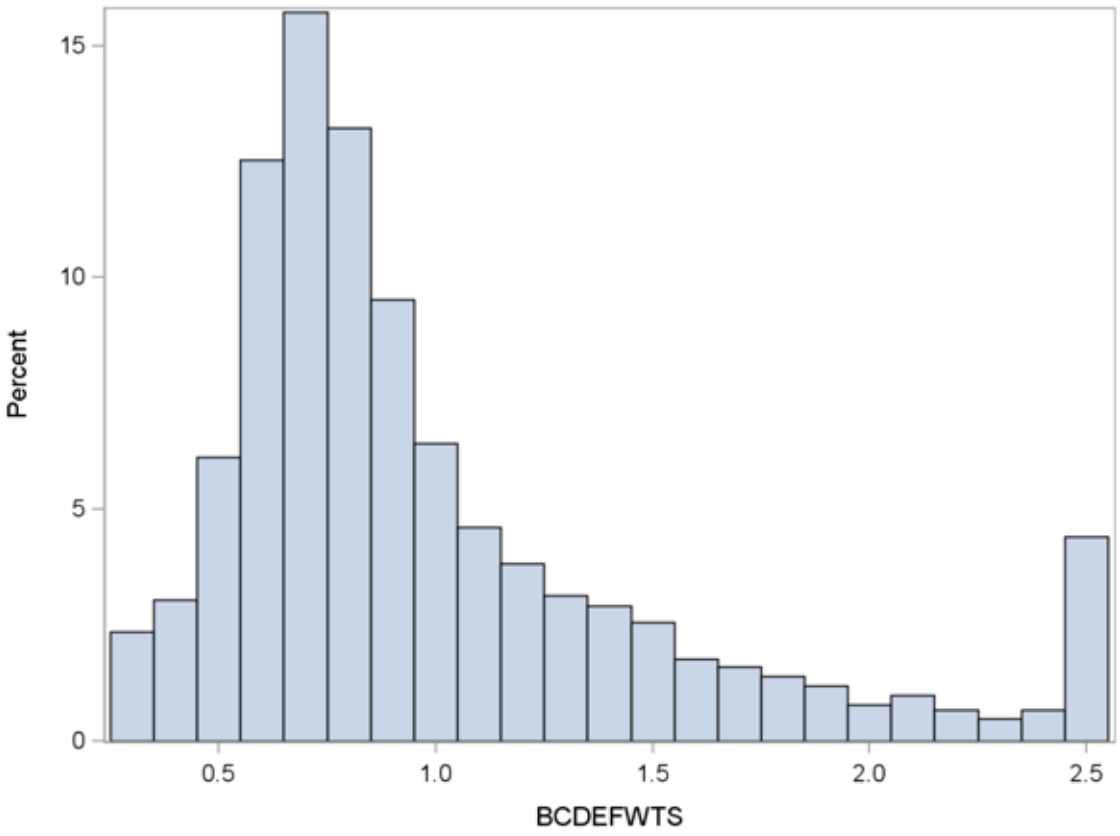


**Figure F4:. Distribution of final sample weight for Wave 6—B cohort longitudinal weight**

**Table F4: Analysis variable: BCDEFWTS—B cohort longitudinal weight**

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | N |
|---|---|---|---|---|---|---|---|
| 1.0000000 | 0.5255565 | 0.3300000 | 2.5000000 | 2.5000000 | 2.1700000 | 3441.00 | 3441 |

## K cohort—cross-sectional weight

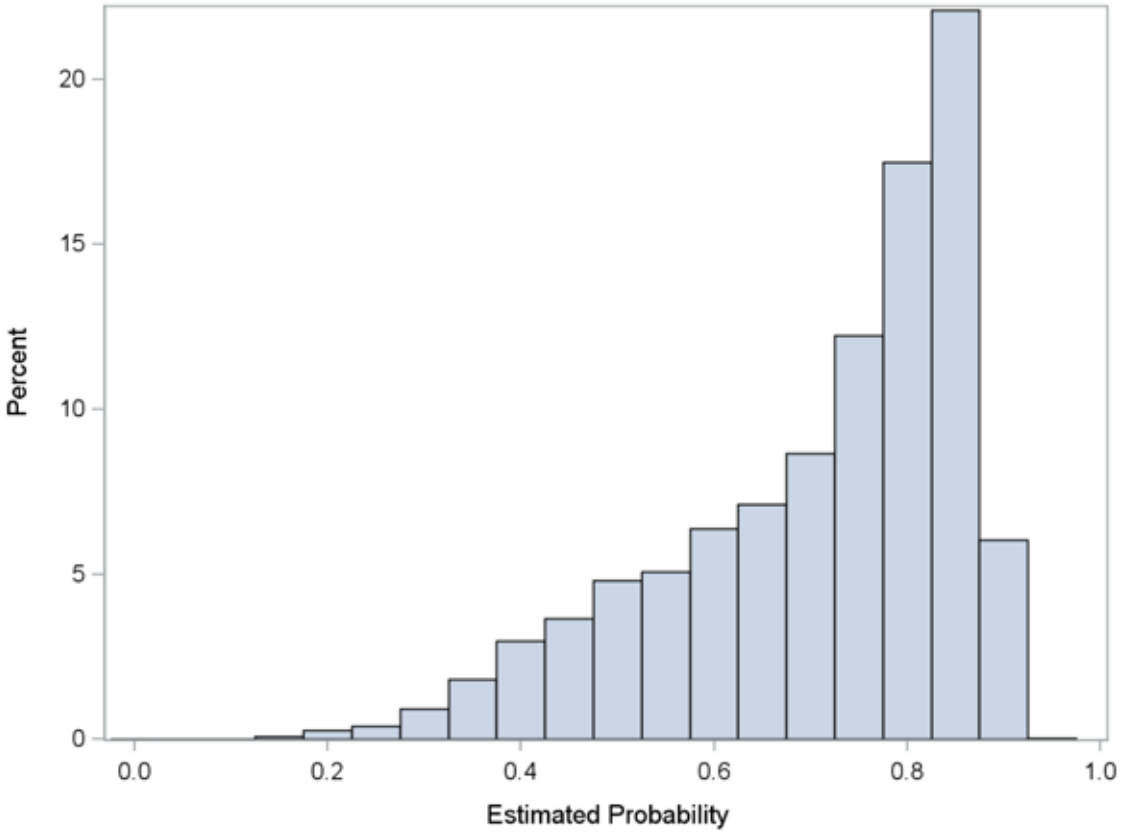


**Figure F5: Distribution of estimated response propensities—K cohort cross-sectional weight**

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

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | N |
|---|---|---|---|---|---|---|---|
| 0.7098128 | 0.1532877 | 0.1476436 | 0.9361922 | 0.8475647 | 0.7885486 | 3537.00 | 4983 |

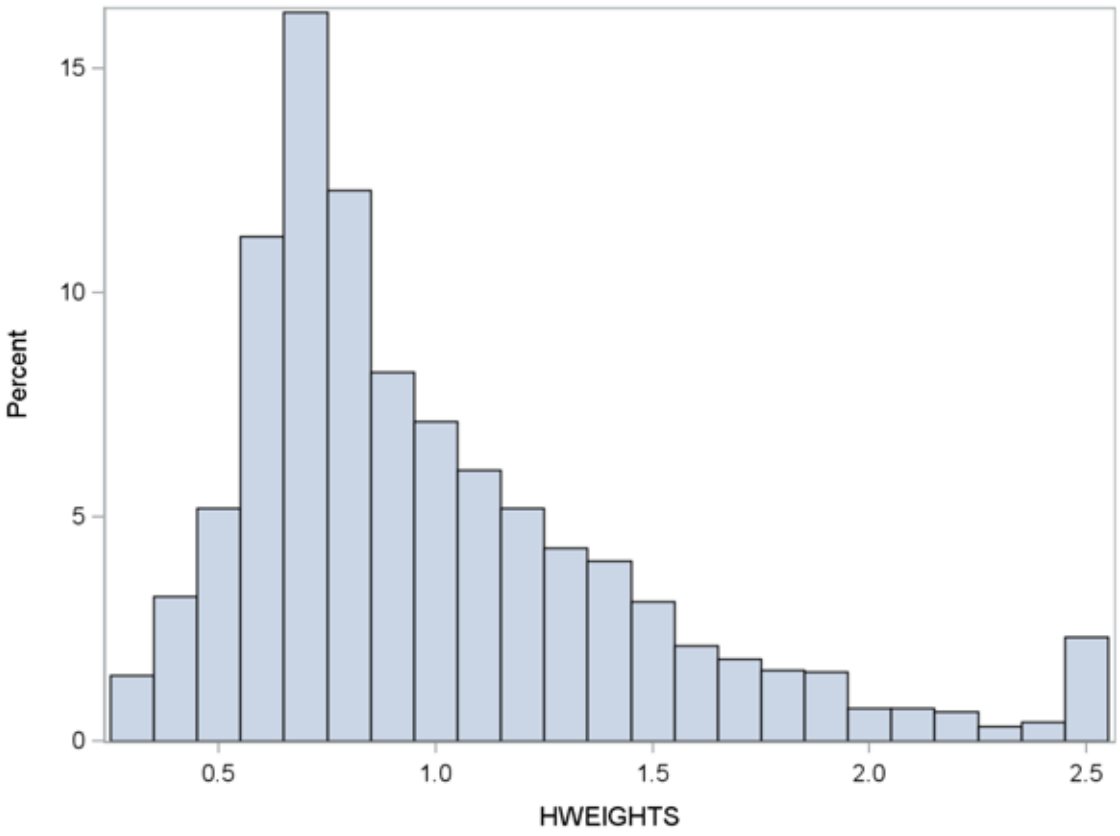


**Figure F6: Distribution of final sample weight for Wave 6—K cohort cross-sectional weight**

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

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | N |
|---|---|---|---|---|---|---|---|
| 1.0000000 | 0.4702167 | 0.3300000 | 2.5000000 | 2.5000000 | 2.1700000 | 3537.00 | 3537 |

## K cohort—longitudinal weight

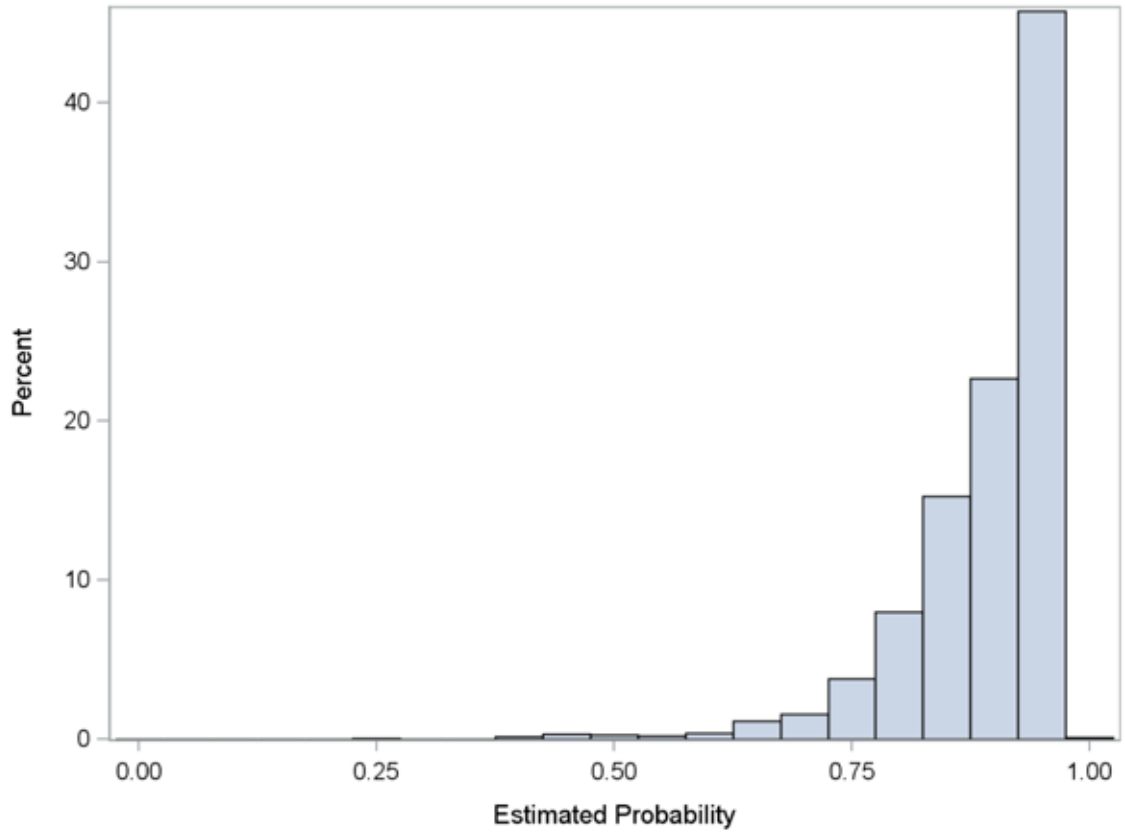


**Figure F7: Distribution of estimated response propensities—K cohort longitudinal weight**

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

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | N |
|---|---|---|---|---|---|---|---|
| 0.8897338 | 0.0861107 | 0.1583724 | 0.9785689 | 0.9488212 | 0.8201965 | 3276.00 | 3682 |

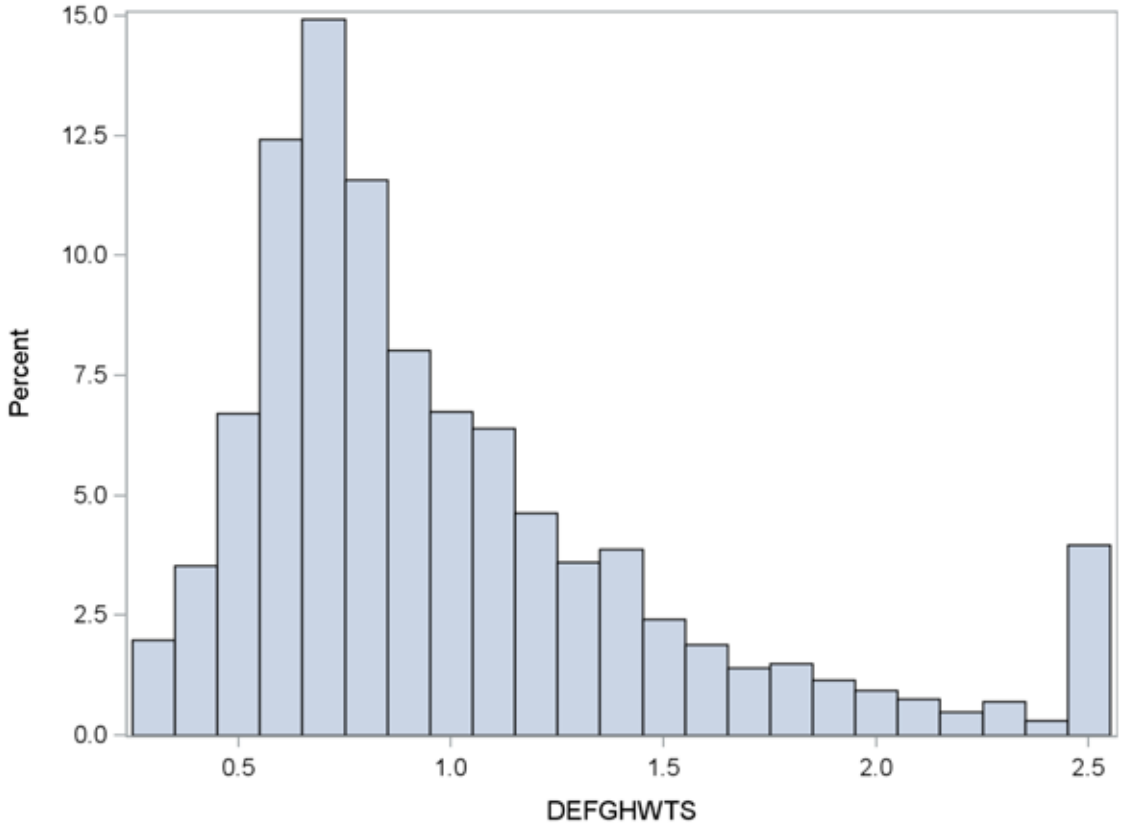


**Figure F8: Distribution of final sample weight for Wave 6—K cohort longitudinal weight**

**Table F8: Analysis variable: DEFGHWTS—K cohort longitudinal weight**

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | N |
|---|---|---|---|---|---|---|---|
| 1.0000000 | 0.5098827 | 0.3300000 | 2.5000000 | 2.5000000 | 2.1700000 | 3276.00 | 3276 |

# Appendix G: Non-response to instruments

**Table G1: Non-response to instruments**

| | Eligible | Responding | %Wave 1 | Response rate % |
|---|---|---|---|---|
| | | **B cohort** | | |
| **Wave 6 (issued sample = 4483)** | | | | |
| Interview | 3764 | 3764 | 73.7 | 100.0 |
| P1CASI | 3759 | 3668 | 71.8 | 97.6 |
| P2SC | 3198 | 2312 | na | 72.3 |
| PLECATI | 559 | 398 | na | 71.2 |
| TEACH | 3762 | 3100 | na | 82.4 |
| ACASIB | 3648 | 3597 | 70.4 | 98.6 |
| TUD | 3649 | 3460 | 67.8 | 94.8 |
| MR | 3648 | 3585 | 70.2 | 98.3 |
| **Wave 5 (issued sample = 4658)** | | | | |
| Interview | 4,085 | 4,085 | 80.0 | 100.0 |
| P1CASI | 4,077 | 4,010 | 78.5 | 98.4 |
| P2SC | 3,512 | 2,444 | na | 69.6 |
| PLECATI | 537 | 404 | na | 75.2 |
| TEACH | 4,021 | 3,490 | na | 86.8 |
| | | **K cohort** | | |
| **Wave 6 (issued sample = 4395)** | | | | |
| 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 |
| **Wave 5 (issued sample = 4551)** | | | | |
| Interview | 3,956 | 3,956 | 77.5 | 100.0 |
| P1CASI | 3,952 | 3,857 | 77.4 | 97.6 |
| P2SC | 3,277 | 2,333 | na | 71.2 |
| PLECATI | 614 | 464 | na | 75.6 |
| TEACH | 3,857 | 3,225 | na | 83.6 |
| ACASI | 3,873 | 3,844 | 77.1 | 99.3 |
| CSRK | 3,872 | 3,850 | 77.3 | 99.4 |
| TUD | 3,871 | 3,649 | 73.2 | 94.3 |

| 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 6, 3% of Parent 1's did not complete the P1 CASI.

## Parent 2 self-completed forms

The response rate for Wave 6 Parent 2's was around 74% compared with 70% in Wave 5.

## Parent Living Elsewhere (PLE) instrument

Of the eligible PLE's that interviewers attempted to contact, 73% responded.

## Teacher self-completed form

The teacher forms continue to achieve good response rates (over 81%). When compared to 85% in Wave 5. In Wave 6 teacher forms for the B cohort were again sent to the study child's main classroom teacher. However due to the majority of the K cohort children attending high school in Waves 5 and 6, the teacher forms for the K study children were sent to their English teacher. Importantly, this change in protocol did not negatively affect teacher response rates.

## Child interview

The response rate for the Time Use Diary (TUD) for the K cohort remains high at 91% compared with 94% in Wave 5. In Wave 6 the B cohort completed the TUD for the first time and the response rate was 95%. The combined response rate was 93%.

## Instrument response rate by characteristics of families

Based on Wave 1 characteristics, the response rates to the instruments in Wave 6 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 5 to 29% 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 6% 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 to 10% lower than the full sample.
- When combined parental income was at least $1000pw, Parent 2 was 9% more likely and the PLE was 7% more likely to take part in an interview than when combined parental income was below $1000pw.
- Similarly, where Parent 1 was employed Parent 2 was 5% more likely to take part in an interview compared to where Parent 1 was not employed.
- South Australia had the highest response rate to the Parent 2 form (82%); the lowest was in New South Wales (69%).
- The highest response rate to the teacher questionnaire was in Tasmania (89%); teachers in the Northern Territory had the lowest response rate (83%).
- Study children from Queensland and the ACT had the highest response rate to the TUD (96%), while those from Tasmania had the lowest (92%).

## K cohort

The following differences in response were observed:

- Aboriginal and Torres Strait Islander children were under-represented across all parent and teacher forms, with a response rate 5 to 32% lower than the non-Aboriginal and Torres Strait Islander sample.
- Indigenous children also had a lower response rate to the TUD (83%) when compared to the non-Indigenous sample (91%).
- 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 6% lower than the full sample. Where Parent 1 spoke a language other than English at home, Parent 2 response rates were 11% lower than families where Parent 1 spoke only English.
- When combined parental income was at least $1000pw, Parent 2 and the PLE were 8 to 10% more likely to take part in an interview than when the combined parental income was below $1000pw.
- Similarly, where Parent 1 was employed Parent 2 was 6% more likely to take part in an interview compared to where Parent 1 was not employed.
- Western Australia had the highest response rate to the P2 form (81%); Victoria had the lowest (73%).
- The highest response rate to the teacher questionnaire was from Tasmania (88%); the lowest was from the Northern Territory (68%).
- Study children from South Australia had the highest response rate to the TUD (93%), while those from Tasmania had the lowest (82%).

# 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 | ACASIB | TUD |
|---|---|---|---|---|---|---|---|
| **Full sample** | **84.0** | **97.6** | **72.3** | **71.2** | **84.4** | **98.6** | **94.8** |
| | **(4,483)** | **(3,759)** | **(3,198)** | **(559)** | **(3,672)** | **(3,648)** | **(3,649)** |
| Study child Indigenous | 60.9 | 95.3 | 55.8 | 43.5 | 80.0 | 97.0 | 92.9 |
| | (174) | (106) | (77) | (23) | (100) | (99) | (99) |
| Study child non-Indigenous | 84.9 | 97.6 | 72.7 | 72.4 | 84.5 | 98.6 | 94.9 |
| | (4,309) | (3,653) | (3,121) | (536) | (3,572) | (3,549) | (3,550) |
| Parent 1 LOTE spoken | 78.5 | 93.8 | 61.7 | 67.7 | 82.9 | 98.2 | 92.4 |
| | (573) | (448) | (405) | (31) | (427) | (433) | (433) |
| Parent 1 English only | 84.8 | 98.1 | 73.8 | 71.4 | 84.6 | 98.7 | 95.1 |
| | (3,910) | (3,311) | (2,793) | (528) | (3,245) | (3,215) | (3,216) |
| Parent 1 Employed | 87.6 | 97.8 | 74.6 | 71.9 | 84.5 | 98.7 | 95.4 |
| | (2,303) | (2,016) | (1,750) | (299) | (1,985) | (1,966) | (1,966) |
| Parent 1 Not Employed | 80.1 | 97.4 | 69.4 | 70.4 | 84.2 | 98.5 | 94.1 |
| | (2,172) | (1,736) | (1,441) | (260) | (1,680) | (1,675) | (1,676) |
| Parental income <$1000 | 78.4 | 97.4 | 67.1 | 67.3 | 83.1 | 98.4 | 93.6 |
| | (1,859) | (1,456) | (1,164) | (263) | (1,419) | (1,407) | (1,408) |
| Parental income >=$1000 | 88.6 | 97.9 | 76.5 | 74.7 | 85.2 | 98.8 | 95.8 |
| | (2,390) | (2,115) | (1,875) | (273) | (2,072) | (2,063) | (2,063) |
| NSW | 82.2 | 97.3 | 69.4 | 72.3 | 84.0 | 98.8 | 95.5 |
| | (1,406) | (1,153) | (992) | (159) | (1,124) | (1,119) | (1,119) |
| VIC | 82.8 | 98.2 | 71.8 | 67.8 | 84.5 | 98.8 | 94.3 |
| | (1,082) | (895) | (785) | (115) | (880) | (881) | (881) |
| QLD | 83.6 | 97.3 | 72.0 | 72.7 | 83.9 | 98.2 | 95.5 |
| | (920) | (769) | (642) | (128) | (747) | (740) | (740) |
| SA | 85.7 | 98.6 | 81.5 | 80.8 | 85.3 | 99.3 | 93.5 |
| | (335) | (287) | (232) | (52) | (285) | (277) | (278) |
| WA | 87.0 | 97.5 | 73.5 | 67.7 | 84.9 | 97.7 | 94.6 |
| | (460) | (399) | (340) | (62) | (384) | (386) | (386) |
| TAS | 94.3 | 98.0 | 76.2 | 66.7 | 88.8 | 98.9 | 91.6 |
| | (106) | (100) | (84) | (15) | (98) | (95) | (95) |

**Table H1: B cohort non-response to forms**

| Response rate % (N) | F2F | P1CASI | P2SC | PLE CATI | TEACH | ACASIB | TUD |
|---|---|---|---|---|---|---|---|
| **Full sample** | **84.0** | **97.6** | **72.3** | **71.2** | **84.4** | **98.6** | **94.8** |
| | **(4,483)** | **(3,759)** | **(3,198)** | **(559)** | **(3,672)** | **(3,648)** | **(3,649)** |
| ACT | 88.2 | 97.8 | 81.1 | 46.2 | 84.1 | 97.8 | 95.5 |
| | (102) | (90) | (74) | (13) | (88) | (89) | (89) |
| NT | 91.7 | 92.4 | 71.4 | 80.0 | 83.3 | 100.0 | 91.8 |
| | (72) | (66) | (49) | (15) | (66) | (61) | (61) |
| Capital city | 84.6 | 97.5 | 73.7 | 71.5 | 85.3 | 98.7 | 95.1 |
| | (2,862) | (2,418) | (2,072) | (340) | (2,359) | (2,353) | (2,353) |
| Rest of state | 82.9 | 97.7 | 69.9 | 70.5 | 82.7 | 98.4 | 94.2 |
| | (1,610) | (1,331) | (1,118) | (217) | (1,304) | (1,285) | (1,286) |
| Study child male | 83.8 | 98.0 | 71.1 | 68.6 | 83.4 | 98.2 | 94.0 |
| | (2,302) | (1,929) | (1,631) | (293) | (1,884) | (1,871) | (1,872) |
| Study child female | 84.1 | 97.1 | 73.6 | 74.1 | 85.5 | 99.0 | 95.7 |
| | (2,181) | (1,830) | (1,567) | (266) | (1,788) | (1,777) | (1,777) |

# 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** | **80.5** | **95.7** | **76.2** | **75.8** | **78.9** | **97.9** | **98.1** | **90.7** |
| | **(4395)** | **(3526)** | **(2904)** | **(554)** | **(3413)** | **(3388)** | **(3386)** | **(3387)** |
| Study child Indigenous | 58.9 | 88.0 | 44.4 | 47.1 | 74.4 | 93.3 | 97.3 | 82.7 |
| | (141) | (83) | (54) | (17) | (78) | (75) | (75) | (75) |
| Study child non-Indigenous | 81.2 | 96.0 | 76.8 | 76.7 | 79.0 | 98.0 | 98.2 | 90.9 |
| | (4252) | (3441) | (2849) | (536) | (3333) | (3311) | (3309) | (3310) |
| Parent 1 LOTE spoken | 74.4 | 89.4 | 66.6 | 76.6 | 76.3 | 97.9 | 98.4 | 89.6 |
| | (610) | (452) | (395) | (47) | (431) | (431) | (430) | (431) |
| Parent 1 English only | 81.5 | 96.7 | 77.7 | 75.7 | 79.2 | 97.9 | 98.1 | 90.8 |
| | (3785) | (3074) | (2509) | (507) | (2982) | (2957) | (2956) | (2956) |
| Parent 1 Employed | 83.7 | 97.0 | 78.5 | 76.4 | 78.9 | 98.1 | 98.6 | 91.6 |
| | (2589) | (2162) | (1810) | (356) | (2110) | (2075) | (2074) | (2075) |
| Parent 1 Not Employed | 76.0 | 93.8 | 72.4 | 75.0 | 78.9 | 97.6 | 97.5 | 89.2 |
| | (1800) | (1361) | (1092) | (196) | (1300) | (1310) | (1309) | (1309) |
| Parental income <$1000 | 64.2 | 76.2 | 69.4 | 71.6 | 76.7 | 97.3 | 97.3 | 87.0 |
| | (1859) | (1456) | (847) | (271) | (1130) | (1126) | (1126) | (1126) |
| Parental income >=$1000 | 90.1 | 99.1 | 79.7 | 79.5 | 80.2 | 98.2 | 98.6 | 92.9 |
| | (2390) | (2115) | (1909) | (249) | (2101) | (2082) | (2081) | (2081) |
| NSW | 80.5 | 95.2 | 77.2 | 72.0 | 76.3 | 98.3 | 99.1 | 92.4 |
| | (1371) | (1100) | (902) | (168) | (1067) | (1064) | (1064) | (1064) |
| VIC | 78.4 | 95.7 | 73.3 | 77.8 | 80.0 | 97.8 | 97.7 | 89.5 |
| | (1083) | (846) | (703) | (135) | (822) | (809) | (809) | (809) |
| QLD | 80.4 | 96.4 | 75.1 | 79.3 | 77.7 | 98.0 | 97.1 | 90.2 |
| | (862) | (690) | (571) | (121) | (664) | (663) | (663) | (663) |
| SA | 80.8 | 97.6 | 78.8 | 81.8 | 81.3 | 99.2 | 98.0 | 92.7 |
| | (317) | (255) | (212) | (44) | (251) | (248) | (248) | (248) |
| WA | 81.7 | 96.8 | 80.7 | 67.4 | 82.0 | 97.2 | 97.7 | 89.8 |
| | (453) | (370) | (311) | (46) | (356) | (354) | (354) | (354) |

**Table I1: K cohort non-response to forms**

| Response rate % (N) | F2F | P1 CASI | P2SC | PLE CATI | TEACH | CSRK | ACASI | TUD |
|---|---|---|---|---|---|---|---|---|
| **Full sample** | **80.5** | **95.7** | **76.2** | **75.8** | **78.9** | **97.9** | **98.1** | **90.7** |
| | **(4395)** | **(3526)** | **(2904)** | **(554)** | **(3413)** | **(3388)** | **(3386)** | **(3387)** |
| TAS | 89.0 | 95.6 | 73.5 | 85.7 | 87.9 | 96.4 | 100.0 | 81.8 |
| | (127) | (113) | (83) | (21) | (107) | (110) | (110) | (110) |
| ACT | 88.6 | 94.6 | 76.9 | 85.7 | 85.1 | 95.4 | 98.9 | 92.0 |
| | (105) | (92) | (78) | (7) | (87) | (87) | (87) | (87) |
| NT | 77.9 | 86.7 | 75.0 | 58.3 | 67.8 | 96.2 | 98.0 | 92.3 |
| | (77) | (60) | (44) | (12) | (59) | (53) | (51) | (52) |
| Capital city | 80.2 | 95.8 | 76.5 | 74.8 | 78.4 | 97.8 | 98.0 | 92.1 |
| | (2764) | (2212) | (1849) | (310) | (2143) | (2136) | (2136) | (2136) |
| Rest of state | 81.0 | 95.7 | 75.6 | 77.0 | 79.8 | 98.0 | 98.4 | 88.2 |
| | (1624) | (1309) | (1052) | (243) | (1265) | (1247) | (1245) | (1246) |
| Study child male | 80.2 | 95.4 | 76.5 | 76.7 | 77.5 | 97.6 | 98.0 | 89.6 |
| | (2245) | (1794) | (1488) | (292) | (1744) | (1729) | (1728) | (1728) |
| Study child female | 80.8 | 96.1 | 75.8 | 74.8 | 80.3 | 98.2 | 98.3 | 91.8 |
| | (2150) | (1732) | (1416) | (262) | (1668) | (1659) | (1658) | (1659) |