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. 24

The Longitudinal Study of Australian Children

# Wave 8 Weighting and non-response

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

October 2020

# 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.

This paper is an extension of several previous weighting and non-response technical reports produced by the Australian Institute of Family Studies and the Australian Bureau of Statistics. The authors wish to thank Steve Xu for his work on developing the weighting for Wave 8.



*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. 24, Wave 8 Weighting and Non-Response

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

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

ISBN 978-1-76016-201-6 (Online)
ISBN 978-1-76016-202-3 (PDF)

# Contents

# List of figures



# List of tables

# Introduction

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

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

The two cohorts of children included in the study were:

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

The first wave of data collection took place in 2004, with subsequent main waves conducted every two years. Wave 8 of the Longitudinal Study of Australian Children was conducted in 2018 with B cohort children at age 14–15 years and K cohort children at age 18–19 years. The number of active participants continues to decrease from wave to wave, as a result of failure to maintain contact, participants opting out (including some instances where the study child has deceased), 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 mailout sample, of which 16,342 were contacted and 10,090 successfully recruited to participate in the study. Of these 10,090 children recruited in the Wave 1 sample, 6,164 children responded in Wave 8, and 5,232 children responded to all waves.

## The use of weighting in analysis

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

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

This can result in the composition of the active sample being skewed 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 8 of the Longitudinal Study of Australian Children, with a focus on the treatment of bias. We encourage data users to either make use of survey weights or incorporate into their models those variables we have identified in the weighting process as being related to response propensity. We also offer a timely reminder to users that LSAC is based on a clustered sample design using a primary sampling unit of postcode, and that this variable should be used when conducting statistical tests to avoid overstating significance.

## Summary of sample design properties

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

Table 1 provides a summary for reference.

**Table 1:** LSAC sample design properties

| Property | Description |
|---|---|
| **Scope** (the population about which inference is to be made) | Two cohorts of children (the B cohort who were 0–1 year and the K cohort who were 4–5 years old during 2004, the Wave 1 recruitment year). The scope excluded very remote areas of Australia. |
| **Coverage** (the population represented by the active participating sample) | For Wave 1 recruitment: The subset of Wave 1 scope who had contact records 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: One 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. |

## Responding units

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

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

Study Child was interviewed or not. Therefore, in Wave 8 records were considered responding if we were able to conduct an interview with the young person or any of the parental figures.

**Table 2:** Responding units by wave

| Cohort | Waves | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| B | P1 | P1 | P1 | P1 | P1 or SC | P1 or SC | P1 or SC | P1 or SC |
| K | P1 | P1 | P1 | P1 or SC | P1 or SC | P1 or SC | P1 or SC | SC or W7P1 or W7P2 or W7PLE |

# Summary of weighting in Waves 1–7

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

Weights for Waves 2–7 were calculated by adjusting previous wave weights for differential sample attrition in two stages (Cusack & Defina, 2014; Sipthorp & Daraganova, 2011; Sipthorp & Misson, 2007, 2009; Usback, 2018). At the first stage, a modelled response propensity factor was applied; at the second, the weights were adjusted to preserve stratum totals. Extreme weights were capped as a form of outlier treatment to avoid any particular child contributing much more than other children in the sample to a weighted estimate, because this can potentially lead to volatile statistics if any such child has unusual characteristics.

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

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

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

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

# Wave 8 weighting method

This section contains a brief description of the method used to create weights for Wave 8 data. The method is largely unchanged from Wave 7.

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 realign weight totals with known totals from the original sample. Both stages contribute to non-response bias reduction.

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

For calculation of cross-sectional weights, the final weights used in Wave 1 are adjusted for all additional non-response in the current wave – regardless of whether a unit responded to Waves 2–7.

## Initial weights

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

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

## Response propensity modelling

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

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

The method for selection of covariates to use in the response propensity model is largely unchanged from Wave 7. A stepwise model selection process is used that considers all possible covariates for the response propensity model (list of covariates provided in Appendix E).

This stepwise process calculates the score chi-square statistics of covariates not in the model and adds the largest covariate not yet in the model. If any covariates are no longer found to be significant ($p < 0.05$) then they are removed from the model. This model selection process resulted in a shortlist of variables to consider adding to the Wave 8 model.

The response propensity model for Wave 8 was then re-run on the shortlisted variables together with the variables used for derivation of Wave 7 weights. Variables that were no longer useful or significant ($p > 0.1$) were removed from the model. Variables used in Wave 7 derivation that were still useful predictors in the Wave 8 model were maintained where possible to achieve consistency over time.

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

- Parent 1 age
- Parent 2 age
- Mother's highest level of high school completed
- Mother's proficiency in spoken English
- Parent 1 self-completed questionnaire returned
- Parent 2 self-completed questionnaire returned
- Number of people in household (new)
- Parent 1: Rent home (new)
- Parent 1: Born in Australia (new)

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

- Parent 1 age
- Parent 2 age
- Parent 1 self-completed questionnaire returned
- Parent 2 self-completed questionnaire returned
- Parent 1 renting home indicator
- Parent 1: School completion (new)
- Parent 1: English as main language at home (new)
- SEIFA Index of Education and Occupation (new)
- Study Child: Indigenous (new)
- SEIFA Index of Economic Resources (new)

## Wave 7 variables used in the B cohort longitudinal weight model

- Parent 1: Study Child overall school achievement (new)
- Interviewed in November–February (Derived from gdatint)
- Parent 1: School completion (new)
- Parent 2 self-completed questionnaire returned
- SEIFA Index of Relative Socio-economic Advantage and Disadvantage (new)

## Wave 7 variables used in the K cohort longitudinal weight model

- Parent 1 how far will study child go in education
- Parent 1 rents home
- Parent 2 self-completed questionnaire returned
- SEIFA Index of Relative Socio-economic Advantage and Disadvantage (new)

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

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

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

# Stratum weight adjustment

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

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

This stratum weight adjustment is also known as post-stratification or calibration to benchmarks. There is a separate adjustment factor calculated for each stratum based on the sum of the response propensity adjusted weights compared to the benchmark of the count of children within that stratum, subject to individual sample weights not exceeding the lower weight cap of 0.33 or the upper weight cap of 3.5 (unchanged from Wave 7). This process of calculating the weight adjustment for each unit to satisfy the benchmark specified while simultaneously satisfying the weight caps specified is achieved iteratively through the ABS SAS implementation of the generalised regression estimator (GREGWT).

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

## Weight capping

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

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

The lower cap of 0.33 and the upper cap of 3.5 have remain unchanged from Wave 7.

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

# Further characteristics of response across waves

## Reacquisition of sample from previous waves

In this context, the reacquisition of sample refers to gaining a full response from a participant who was not considered responding in a previous wave. For the B cohort, out of 1,726 that did not respond to Wave 7, 134 responded to Wave 8. Out of the 2,079 that did not respond to at least one of Waves 2, 3, 4, 5, 6 or 7, 405 responded to Wave 8.

For the K cohort, out of 1,894 that did not respond to Wave 7, 301 responded to Wave 8. Out of the 2,191 that did not respond to at least one of Waves 2, 3, 4, 5, 6 or 7, 527 responded to Wave 8.

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

**Table 3:** Sample reacquisition for Waves 3, 4, 5, 6

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

For the K cohort, there were 301 units that responded to Wave 8 that did not respond to Wave 7. Of these 301 units, there were 230 units where the study child was interviewed. Refer to Table 2 for a definition of what is a considered a responding unit.

## Total responding sample for each wave

The fully responding sample at various stages in the sample drives the calibration and hence weighting process. Observe Tables 4 and 5 for updated counts.

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

| Wave | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Cross-sectional response | 5,107 | 4,606 | 4,386 | 4,242 | 4,085 | 3,764 | 3,381 | 3127 |
| Longitudinal response | – | 4,606 | 4,253 | 3,997 | 3,758 | 3,441 | 3,028 | 2722 |
| Cross-sectional attrition rate (%) | – | 9.8 | 14.1 | 16.9 | 20.0 | 26.3 | 33.8 | 38.8 |
| Longitudinal attrition rate (%) | – | 9.8 | 7.7 | 6.0 | 6.0 | 8.4 | 12.0 | 10.1 |

**Table 5:** Sample counts for the K cohort

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

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

## Number of children with weight at cap

Tables 6 and 7 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.

For the B cohort, the number of units at the upper cap has increased from 42 in Wave 7 to 43 for the cross-sectional weight, and increased from 18 in Wave 7 to 20 for the longitudinal weight.

**Table 6:** Counts of capped sample weights for Wave 8 – B cohort

| | Cross-sectional | | Longitudinal | |
|---|---|---|---|---|
| State | Lower cap (0.33) | Upper cap (3.5) | Lower cap (0.33) | Upper cap (3.5) |
| NSW | 0 | 19 | 2 | 6 |
| Vic. | 0 | 10 | 0 | 7 |
| Qld | 14 | 5 | 9 | 4 |
| SA | 6 | 5 | 0 | 1 |
| WA | 10 | 2 | 1 | 0 |
| Tas. | 10 | 1 | 16 | 2 |
| NT | 12 | 0 | 16 | 0 |
| ACT | 2 | 1 | 4 | 0 |
| **Australia** | **54** | **43** | **48** | **20** |

For the K cohort, the number of units at the upper cap increased from 22 in Wave 7 to 25 for the cross-sectional weight, and increased from 9 in Wave 7 to 11 for the longitudinal weight.

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

| | Cross-sectional | | Longitudinal | |
|---|---|---|---|---|
| State | Lower cap (0.33) | Upper cap (3.5) | Lower cap (0.33) | Upper cap (3.5) |
| NSW | 0 | 10 | 0 | 3 |
| Vic. | 0 | 7 | 0 | 4 |
| Qld | 0 | 5 | 0 | 4 |
| SA | 0 | 1 | 0 | 0 |
| WA | 0 | 2 | 4 | 0 |
| Tas. | 22 | 0 | 8 | 0 |
| NT | 21 | 0 | 7 | 0 |
| ACT | 2 | 0 | 3 | 0 |
| **Australia** | **45** | **25** | **22** | **11** |

# Conclusion

Sample attrition has continued again in this wave; with the responding sample around 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 probability of response. The weights developed help to correct for different response patterns, allowing users to better analyse the data and draw more accurate conclusions about the population, being the two cohorts of children (the B cohort who were 0–1 year and the K cohort who were 4–5 years old during 2004, the Wave 1 recruitment year).

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

# Bibliography

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

# Appendix A: Glossary of terms and abbreviations

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

| Term | Definition |
|---|---|
| ABS | Australian Bureau of Statistics |
| Attrition | Process of sample size shrinking over time due to any mechanism |
| Cohort | Sample with a particular characteristics, e.g. B cohort ages 0–1 years in first wave |
| Coverage | Population represented by the remaining active participants |
| Cross-sectional | Pertaining to a statistic at one time point, typically broken down by characteristics at that time point |
| Design effect | Penalty factor to variance due to sample tending to be similar within selected postcode clusters |
| Estimation | Process of calculating a descriptive statistic from sample using weight, acknowledging the presence of sampling error |
| F2F | Face-to-face |
| Longitudinal | Pertaining to a statistic involving many time points, typically with a focus on evolution of participants over time |
| LSAC | Longitudinal Study of Australian Children |
| Missing data | Data absent either from non-response or partial response |
| Non response | Failure to acquire survey response due to non-contact or refusal (opt-out) |
| P1 | Parent 1, 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 post-strata for weighting |
| Recruited sample | Subset of selected sample who agreed to participate in Wave 1 |
| Response propensity | Chance that a particular individual or group will respond to a given wave |
| Respondent | or Participant or Active Participant: Any child (family) active in the study |
| Selected sample | Selection of children (families) approached at time of Wave 1 recruitment |
| Stratification | Process of dividing population into strata for selection |
| Stratum (Strata) | Cell(s) of population from which set number of children selected in sample |
| Study variable | Any variable collected in the study that data users wish to analyse |
| Weight | Value for a respondent to correct, up or down, for representativeness based on characteristics of responding sample |

# Appendix B: Description of Wave 8 weights

**Table B1:** Description of Wave 8 weights

| SAS name | Cohort | Type | Waves cases responded to |
|---|---|---|---|
| hweight | B | Population | 1 & 8 |
| hweights | B | Sample | 1 & 8 |
| bcdefghwt | B | Population | 1, 2, 3, 4, 5, 6, 7 & 8 |
| bcdefghwts | B | Sample | 1, 2, 3, 4, 5, 6, 7 & 8 |
| jweight | K | Population | 1 & 8 |
| jweights | K | Sample | 1 & 8 |
| defghijwts | K | Population | 1, 2, 3, 4, 5, 6,7 & 8 |
| defghijwt | K | Sample | 1, 2, 3, 4, 5, 6, 7 & 8 |

# 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 | 29.4 | <0.0001 |
| af03m3 | Parent 2 age | 1 | 6.8 | 0.0090 |
| anpeople | Number of people in household | 1 | 29.2 | <0.0001 |
| aho04a3b | Parent 1 rents home | 1 | 16.6 | <0.0001 |
| afd08m1 | Mother's highest level of schooling completed | 4 | 112.3 | <0.0001 |
| afd11m2 | Mother's proficiency in spoken English | 5 | 22.7 | 0.0004 |
| ap1scd | Parent 1 self-completed questionnaire returned | 1 | 14.9 | 0.0001 |
| ap2scd | Parent 2 self-completed questionnaire returned | 2 | 20.8 | <0.0001 |
| zf09m2 | Parent 1 born in Australia | 1 | 7.4 | 0.0065 |

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 |
|---|---|---|---|---|
| glc08a3a | Parent 1: Study Child overall school achievement | 6 | 12.3 | 0.0566 |
| gfd08a1 | Parent 1's highest level of schooling completed | 5 | 35.3 | <0.0001 |
| gcnfsad2 | SEIFA Index of Relative Socio-Economic Advantage and Disadvantage | 1 | 2.8 | 0.0959 |
| gp2scd | Parent 2 self-complete questionnaire returned | 2 | 53.6 | <0.0001 |
| eoy | Interviewed in November–February (Derived from gdatint) | 1 | 13.6 | 0.0002 |

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

| Variable name | Description | DF | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|
| cf03m2 | Parent 1 age | 1 | 15.7 | <0.0001 |
| cf03m3 | Parent 2 age | 1 | 3.6 | 0.0576 |
| cp1scd | Parent 1 self-completed questionnaire returned | 1 | 9.1 | 0.0026 |
| cp2scd | Parent 2 self-completed questionnaire returned | 2 | 33.5 | <0.0001 |
| cho04a3b | Parent 1 rents home | 1 | 17.1 | <0.0001 |
| cfd08a1 | Parent 1's highest level of schooling completed | 5 | 89.1 | <0.0001 |
| cf11m2 | Parent 1: English as main language at home | 1 | 47.5 | <0.0001 |
| ccnfseo | SEIFA Index of Education and Occupation | 1 | 19.6 | <0.0001 |
| zf12m1 | Study Child: Indigenous | 3 | 20.9 | 0.0001 |
| ccnfser | SEIFA Index of Economic Resources | 1 | 7.1 | 0.0076 |

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

| Variable name | Description | DF | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|
| icnfsad2 | SEIFA Index of Relative Socio-Economic Advantage and Disadvantage | 1 | 8.65 | 0.0033 |
| ip2scd | Parent 2 self-completed questionnaire returned | 2 | 19.9 | <0.0001 |
| iho04a3b | Parent 1 rents home | 2 | 17.2 | 0.0002 |
| ihe13a | Parent 1: How far study child will go in education | 6 | 22.9 | 0.0008 |

# Appendix D: Odds ratio estimates for variables in Wave 8 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 interval | |
|---|---|---|---|---|
| af03m2 | Parent 1 age | 1.041 | 1.026 | 1.057 |
| af03m3 | Parent 2 age | 1.019 | 1.005 | 1.033 |
| anpeople | Number of people in household | 0.866 | 0.822 | 0.913 |
| afd08m1 1 | **Mother: Highest level of schooling completed Year 12 or equivalent – reference category** | | | |
| afd08m1 2 vs 1 | Mother completed Year 11 or equivalent | 0.501 | 0.415 | 0.606 |
| afd08m1 3 vs 1 | Mother completed Year 10 or equivalent | 0.523 | 0.444 | 0.616 |
| afd08m1 4 vs 1 | Mother completed Year 9 or equivalent | 0.426 | 0.298 | 0.609 |
| afd08m1 6 vs 1 | Mother completed Year 8 or below | 0.322 | 0.192 | 0.54 |
| aho04a3b 1 vs 2 | Parent 1 Rents home vs **Parent 1 does not rent home** | 0.748 | 0.651 | 0.86 |
| afd11m2 -9 | **Mother's proficiency in spoken English – Not asked – reference category** | | | |
| afd11m2 -4 vs -9 | Mother's proficiency in spoken English – Refusal | 0.348 | 0.081 | 1.485 |
| afd11m2 1 vs -9 | Mother speaks English Very well | 0.652 | 0.522 | 0.815 |
| afd11m2 2 vs -9 | Mother speaks English Well | 0.532 | 0.369 | 0.766 |
| afd11m2 3 vs -9 | Mother speaks English Not well | 0.813 | 0.504 | 1.311 |
| afd11m2 4 vs -9 | Mother speaks English Not at all | 0.635 | 0.251 | 1.608 |
| ap1scd 0 vs 1 | Parent 1 did not return self-completed questionnaire vs **Parent 1 did return self-completed questionnaire** | 0.635 | 0.505 | 0.8 |
| ap2scd 1 | **Parent 2 self-complete questionnaire returned - reference category** | | | |
| ap2scd -9 vs 1 | No Parent 2 in household | 0.765 | 0.47 | 1.244 |
| ap2scd 0 vs 1 | Parent 2 did not return self-completed questionnaire | 0.611 | 0.494 | 0.756 |
| zf09m2 0 vs 1101 | Parent 1: Born in Australia vs **Parent 1: Born elsewhere** | 0.781 | 0.654 | 0.933 |

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

| Effect | Description | Point estimate | 95% Wald confidence interval | |
|---|---|---|---|---|
| gcnfsad2 | SEIFA Index of Relative Socio-Economic Advantage and Disadvantage | 1.001 | 1.000 | 1.003 |
| glc08a3a 3 | **Parent 1: Study Child overall school achievement average – reference category** | | | |
| glc08a3a -9 vs 3 | Parent 1: Study Child overall school achievement not asked | 0.895 | 0.362 | 2.21 |
| glc08a3a -2 vs 3 | Parent 1: Study Child overall school achievement don't know | 0.415 | 0.104 | 1.649 |
| glc08a3a 1 vs 3 | Parent 1: Study Child overall school achievement excellent | 1.576 | 1.073 | 2.316 |
| glc08a3a 2 vs 6 | Parent 1: Study Child overall school achievement above average | 1.194 | 0.881 | 1.617 |
| glc08a3a 4 vs 3 | Parent 1: Study Child overall school achievement below average | 0.804 | 0.507 | 1.275 |
| glc08a3a 5 vs 3 | Parent 1: Study Child overall school achievement well below average | 4.849 | 0.645 | 36.47 |
| gp2scd 1 | **Parent 2 self-complete questionnaire returned – reference category** | | | |
| gp2scd -9 vs 1 | No parent 2 in household | 0.52 | 0.367 | 0.737 |
| gp2scd 0 vs 1 | Parent 2 did not return self-completed questionnaire | 0.346 | 0.260 | 0.461 |
| gfd08a1 1 | **Parent 1: School completion Year 12 or below - reference category** | | | |
| gfd08a1 -2 vs 1 | Parent 1: School completion - Don't know | 0.065 | 0.023 | 0.183 |
| gfd08a1 2 vs 1 | Parent 1: School completion Year 11 or equivalent | 0.642 | 0.428 | 0.961 |
| gfd08a1 3 vs 1 | Parent 1: School completion Year 10 or equivalent | 0.618 | 0.443 | 0.861 |
| gfd08a1 4 vs 1 | Parent 1: School completion Year 9 of equivalent | 0.878 | 0.357 | 2.16 |
| gfd08a1 6 vs 1 | Parent 1: Never attended school | 0.505 | 0.162 | 1.576 |
| eoy 1 vs 0 | Interviewed in November–February (derived from gdatint) vs **Not interviewed In November–February** | 0.556 | 0.407 | 0.759 |

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

| Effect | Description | Point estimate | 95% Wald confidence interval | |
|---|---|---|---|---|
| cf03m3 | Parent 2 age | 1.028 | 1.014 | 1.042 |
| cf03m2 | Parent 1 age | 1.013 | 1.000 | 1.026 |
| cf11m2 0 vs 1201 | Parent 1: English as main language at home vs **Parent 1: Main language at home non-English** | 0.549 | 0.463 | 0.651 |
| cp1scd 0 vs 1 | Parent 1 did not return self-completed questionnaire vs **Parent 1 did return self-completed questionnaire** | 0.710 | 0.569 | 0.887 |
| cp2scd 1 | **Parent 2 did return self-completed questionnaire – reference category** | | | |
| cp2scd -9 vs 1 | No Parent 2 in household | 0.893 | 0.534 | 1.493 |
| cp2scd 0 vs 1 | Parent 2 did not return self-completed questionnaire | 0.539 | 0.437 | 0.664 |
| cho04a3b 1 vs 2 | Parent 1 rents home vs **Parent 1 does not rent home** | 0.734 | 0.635 | 0.85 |
| zf12m1 1 | **Study child Indigenous Status: No – reference category** | | | |
| zf12m1 2 vs 1 | Study child Indigenous Status: Yes, Aboriginal | 0.437 | 0.305 | 0.626 |
| zf12m1 3 vs 1 | Study child Indigenous Status: Yes, Torres Strait Islander | 0.693 | 0.166 | 2.89 |
| zf12m1 4 vs 1 | Study child both Indigenous and Torres Strait Islander | 0.677 | 0.214 | 2.145 |
| cfd08a1 1 | **Parent 1 never attended school – reference category** | | | |
| cfd08a1 2 vs 1 | Parent 1: School completion Year 11 or equivalent | 0.694 | 0.578 | 0.833 |
| cfd08a1 3 vs 1 | Parent 1: School completion Year 10 or equivalent | 0.594 | 0.508 | 0.693 |
| cfd08a1 4 vs 1 | Parent 1: School completion Year 9 or equivalent | 0.336 | 0.246 | 0.458 |
| cfd08a1 5 vs 1 | Parent 1: School completion Year 8 or below | 0.49 | 0.329 | 0.731 |
| cfd08a1 6 vs 1 | **Parent 1: Never attended school** | 6.448 | 0.752 | 55.306 |
| ccnfser | SEIFA Index of Economic Resources | 0.998 | 0.997 | 1.000 |
| ccnfseo | SEIFA Index of Education and Occupation | 1.003 | 1.002 | 1.005 |

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

| Effect | Description | Point estimate | 95% Wald confidence interval | |
|---|---|---|---|---|
| icnfsad2 | SEIFA Index of Relative Socio-Economic Advantage and Disadvantage | 1.003 | 1.001 | 1.004 |
| iho04a3b 2 | **Parent 1 does not rent home – reference category** | | | |
| iho04a3b -9 vs 2 | Parent 1 rents home - Not asked | 0.206 | 0.094 | 0.451 |
| iho04a3b 1 vs 2 | Parent 1 rents home | 0.709 | 0.508 | 0.99 |
| ip2scd 1 | **Parent 2 self-complete questionnaire returned – reference category** | | | |
| ip2scd -9 vs 1 | Parent 2: Self-complete questionnaire not asked | 0.856 | 0.600 | 1.222 |
| ip2scd 0 vs 1 | Parent 2: Self-complete questionnaire not returned | 0.518 | 0.386 | 0.694 |
| ihe13a 4 | **Parent 1: Study child will go to university and complete a degree – reference category** | | | |
| ihe13a -9 vs 4 | Parent 1: How far study child will go in education not asked | 1.007 | 0.607 | 1.672 |
| ihe13a -3 vs 4 | Parent 1: How far study child will go in education – Refused or not answered | 0.558 | 0.311 | 1.004 |
| ihe13a 1 vs 4 | Parent 1: Study child will leave school before finishing secondary school | 0.693 | 0.28 | 1.716 |
| ihe13a 2 vs 4 | Parent 1: Study child will complete secondary school | 0.603 | 0.417 | 0.872 |
| ihe13a 3 vs 4 | Parent 1: Study child will complete a trade or vocational training course | 0.644 | 0.453 | 0.914 |
| ihe13a 5 vs 4 | Parent 1: Study child will obtain post-graduate qualifications at a university e.g. Master degree or Doctoral degree | 2.118 | 1.151 | 3.897 |

# 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 |
| zf02m1 | SC—F2F A3—Sex |

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

| Variable name | Variable label |
|---|---|
| gcnfsad2 | 12/13 - SEIFA - Index of Relative Socio-Economic Advantage and Disadvantage - 2016 - SA2 - Score |
| gcnfsad2d | 12/13 - SEIFA - Index of Relative Socio-Economic Advantage and Disadvantage - 2016 - SA2 - Score - Deciles - National |
| gcnfser2 | 12/13 - Home - SEIFA Economic Resources - 2016 - SA2 - Score |
| gcnfser2d | 12/13 - Home - SEIFA Economic Resources - 2016 - SA2 - Deciles - National |
| gf01gm | M@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 - Language other than English spoken at home |
| gfd08a1 | 12/13 - P1 - P1 CAI A1.1/A1.3+W1-6 - School completion |
| gfd08a2a | 12/13 - P1 - P1 CAI A1.2/A1.3+W1-6 - Finished other post-secondary qualification |
| gfd08a3a | 12/13 - P1 - P1 CAI A1.3+W1-6 - Highest qualification |
| gfd08m1 | 12/13 - M - P1CAI A1.1/A1.3+W1-6 - School completion |
| gfemp | 12/13 - F - Employment status |
| gho04a1 | 12/13 - P1 CAI I1.6-1.8 - Home ownership |
| gho04a3b | 2/13 - P1 CAI I1.6.2 - Rent home |
| gho04a5 | 12/13 - P1 CAI I1.6-1.8 - Housing tenure |
| glc08a3a | 12/13 - P1 - P1 CAI C7.5 - Overall school achievement |
| gmemp | 12/13 - M - employment status |
| gnpeople | 12/13 - No. people in household |
| gnsib | 12/13 - No. siblings of SC in household |
| gp2 | 12/13 - SC has 2 parents In the home |
| gp2scd | 12/13 - Parent 2 Self-complete data present |
| zf02gp1 | P1@12/13 - Sex |
| zf09gp1 | P1@12/13 - Country of birth |
| zf12gp1 | P1@12/13 - Indigenous status |
| ghe11a3e | 12/13 - P1 - P1 CAI C6.2 - How often help child with homework |
| ghe09a1 | 12/13 - P1 CAI F8.4 - Extra-curricular - Any |
| gdatint | 12/13 - Date of interview |
| gf01gp2 | P2@12/13 - Present for wave |

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

| Variable name | Variable label |
|---|---|
| caangb | 4/5—P1—Angry parenting (v3) |
| cahact | 4/5—P1—Home activities index |
| ccnfsad | 4/5—Home—SEIFA Advantage/Disadvantage |
| ccnfseo | 4/5—Home—SEIFA Education & Occupation |
| ccnfser | 4/5—Home—SEIFA Economic Resources |
| cf01cm | 4/5—M@4/5—Present for wave |
| cf01m3 | 4/5—P2@W1—Present for wave |
| cf03m2 | 4/5—P1@W1—F2F A4—Age |
| cf03m3 | 4/5—P2@W1—F2F A4—Age |
| cf11cm | 4/5—M@4/5—F2F A12—Main language spoken at home |
| cf11m1 | 4/5—SC—F2F A12—Main language spoken at home |
| cf11m2 | 4/5—P1@W1—F2F A12—Main language spoken at home |
| cfd08a1 | 4/5—P1—F2F H3—School completion |
| cfd08m1 | 4/5—M—F2F H3—School completion |
| cfd11m2 | 4/5—M—F2F H10—Proficiency in spoken English |
| cho04a3b | 4/5—P1—F2F L4—Rent home |
| cho04a5 | 4/5—P1—F2F L5—Housing tenure |
| cho09a1a1 | 4/5—P1—F2F L11—Safe neighbourhood |

| | |
|---|---|
| cnpeople | 4/5—No. 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 |
| Stratum | Stratum |
| zf02m1 | SC - F2F A3 - Sex |

**Table E4:** Wave 7 data items considered for K cohort – longitudinal weight

| Variable name | Variable label |
|---|---|
| icnfsad2 | 16/17 - SEIFA - Index of Relative Socio-Economic Advantage and Disadvantage - 2016 - SA2 - Score |
| icnfsad2d | 16/17 - SEIFA - Index of Relative Socio-Economic Advantage and Disadvantage - 2016 - SA2 - Score - Deciles - National |
| icnfser2 | 16/17 - Home - SEIFA Economic Resources - 2016 - SA2 - Score |
| icnfser2d | 16/17 - Home - SEIFA Economic Resources - 2016 - SA2 - Deciles - National |
| if01im | M@16/17 - Present for wave |
| if03ip1 | P1@16/17 - Age |
| if03ip2 | P2@16/17 - Age |
| lf11im | M@16/17 - Language other than English spoken at home |
| if11ip1 | P1@16/17 - Language other than English spoken at home |
| if11m1 | 16/17 - SC - Language other than English spoken at home |
| ifd08a1 | 16/17 - P1 - P1 CASI B1.1/B1.3+W1-6 - School completion |
| ihe13a | 16/17 - P1 CASI C6 - How far SC will go in education |
| stratum | 16/27 - Stratum |
| ifd08m1 | 16/17 - M - P1 CASI B1.1/B1.3+W1-6 - School completion |
| ifemp | 16/17 - F - Employment status |
| if01ip2 | P2@16/17 - Present for wave |
| iho04a3b | 16/17 - P1 CAI I1.6.2 - Rent home |
| iho04a5 | 16/17 - P1 CAI I1.6-8 - Housing tenure |
| imemp | 16/17 - M - employment status |
| inpeople | 16/17 - No. people in household |
| insib | 16/17 - No. siblings of SC in household |
| ip2 | 16/17 - SC has 2 parents In the home |
| ip2scd | 16/17 - Parent 2 Self-complete data present |
| zf02ip1 | P1@16/17 - Sex |
| zf09ip1 | P1@16/17 - Country of birth |
| zf12ip1 | P1@16/17 - Indigenous status |
| ihe38a1 | 16/17 - P1 - P1 CAI B8.7.1+W6 - Mother's interest in education |
| ihe38a2 | 16/17 - P1 - P1 CAI B8.7.2+W6 - Father's interest in education |
| zf02m1 | 16/17 - SC - Sex |

# Appendix F: Distributional checks of non-response modelling

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

## B cohort – cross-sectional weight

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

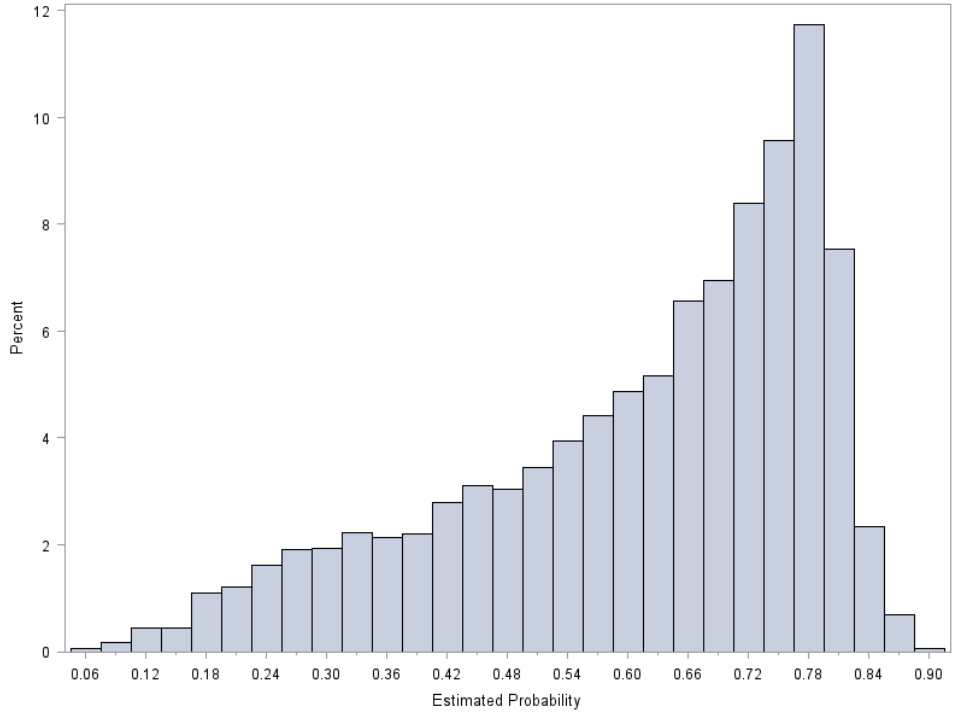


**Table F1:** Analysis variable: estimated probability – B cohort cross-sectional weight

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | *N* |
|---|---|---|---|---|---|---|---|
| 0.6129700 | 0.17765 | 0.053671 | 0.895916 | 0.765239 | 0.84229 | 3,127 | 5,107 |

**Figure F2:** Distribution of final sample weight for Wave 8 – B cohort cross-sectional weight

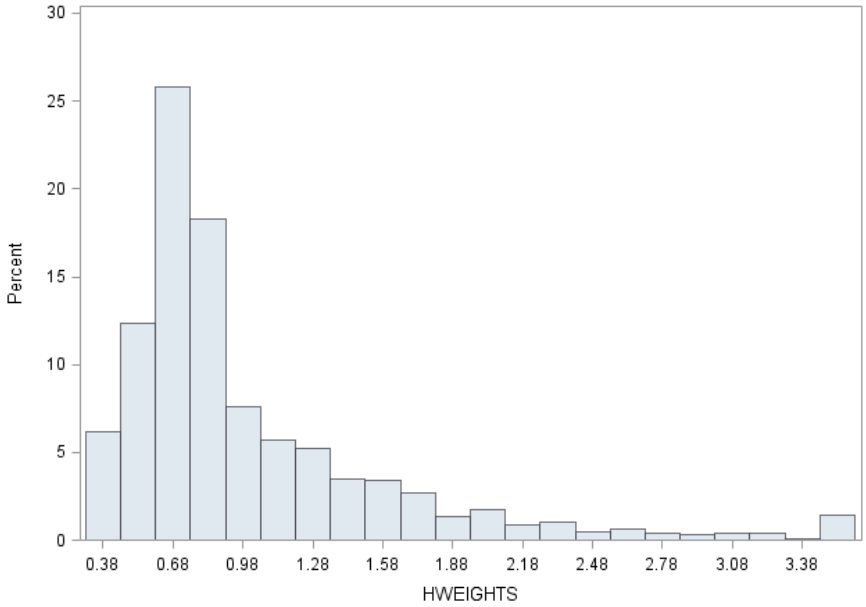


**Table F2:** Analysis variable: GWEIGHTS—B cohort cross-sectional weight

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | *N* |
|---|---|---|---|---|---|---|---|
| 1.0000000 | 0.60351 | 0.3300000 | 3.5000000 | 0.3300000 | 3.1700000 | 3,127 | 3,127 |

# B cohort – longitudinal weight

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

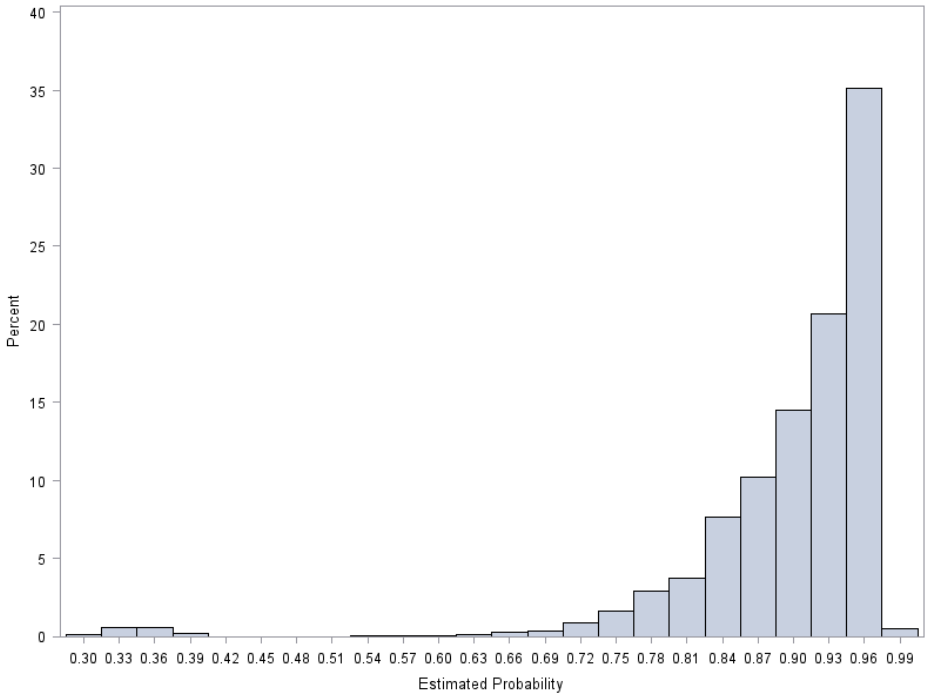


**Table F3:** Analysis variable: estimated probability – B cohort longitudinal weight

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | *N* |
|---|---|---|---|---|---|---|---|
| 0.8989430 | 0.09005 | 0.296077 | 0.990482 | 0.959677 | 0.6944 | 2,722 | 3,028 |

**Figure F4:** Distribution of final sample weight for Wave 8 – B cohort longitudinal weight

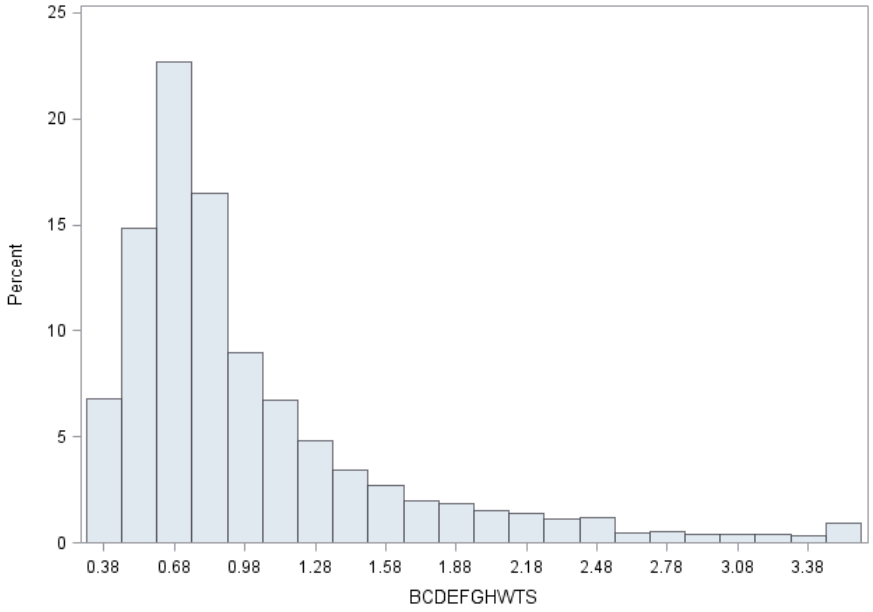


**Table F4:** Analysis variable: BCDEFGWTS – B cohort longitudinal weight

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | *N* |
|---|---|---|---|---|---|---|---|
| 1.0000000 | 0.60904 | 0.3300000 | 3.5000000 | 0.3300000 | 3.1700000 | 2,722 | 2,722 |

# K cohort – cross-sectional weight

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

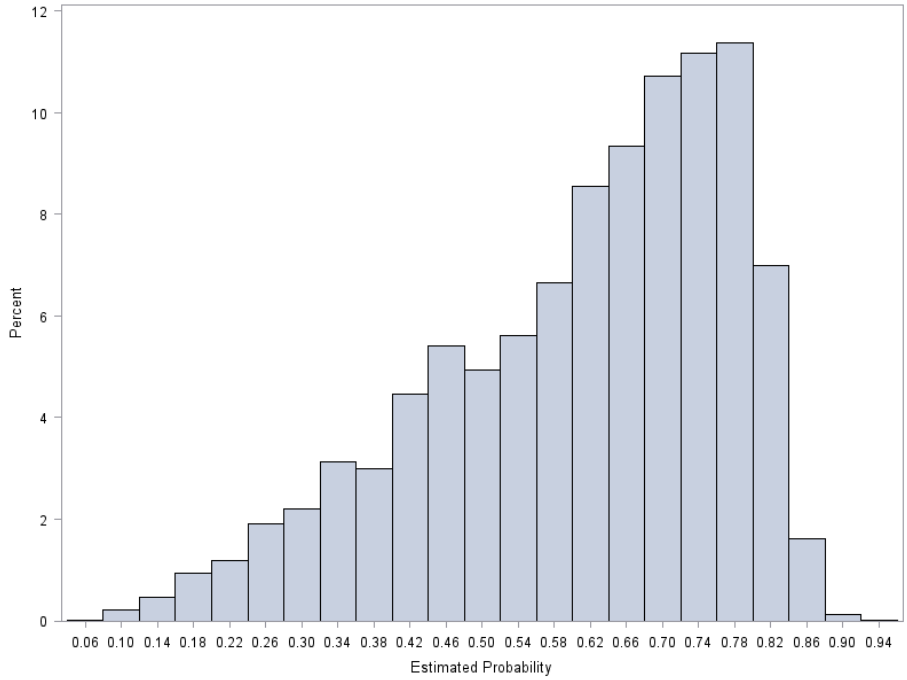


**Table F5:** Analysis variable: estimated probability – K cohort cross-sectional weight

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | *N* |
|---|---|---|---|---|---|---|---|
| 0.6094720 | 0.1664000 | 0.0545860 | 0.9427050 | 0.6662930 | 0.8881200 | 3,037 | 4,983 |

**Figure F6:** Distribution of final sample weight for Wave 8 – K cohort cross-sectional weight

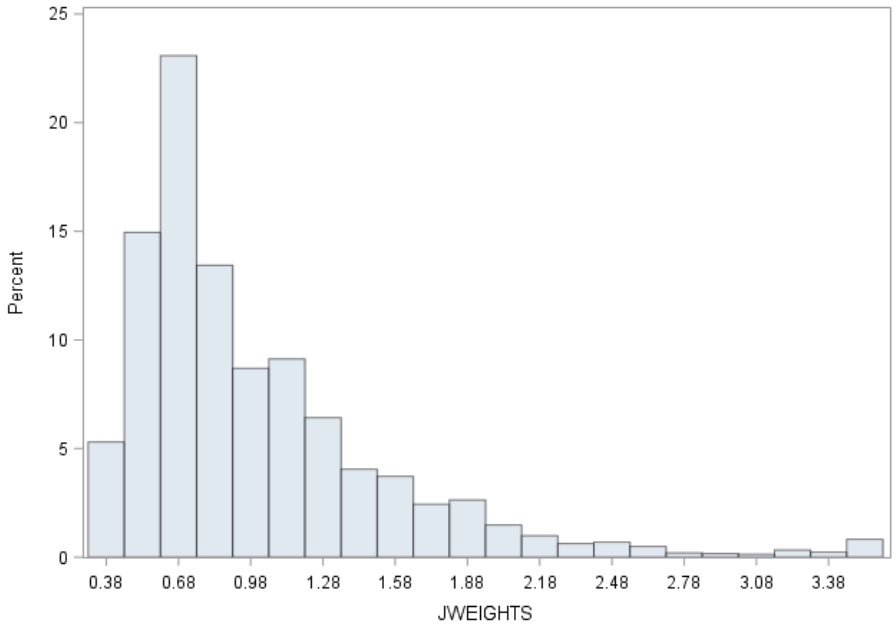


**Table F6:** Analysis variable: IWEIGHTS – K cohort cross-sectional weight

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | *N* |
|---|---|---|---|---|---|---|---|
| 1.0000000 | 0.5535300 | 0.3300000 | 3.5000000 | 0.3300000 | 3.1700000 | 3,037 | 3,037 |

## K cohort – longitudinal weight

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

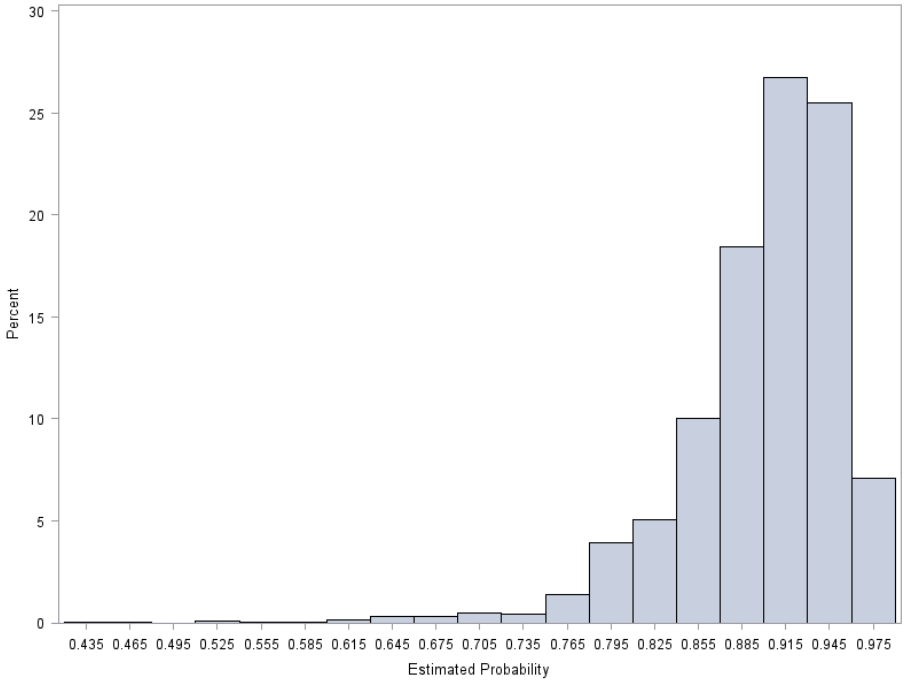


**Table F7:** Analysis variable: estimated probability – K cohort longitudinal weight

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | *N* |
|---|---|---|---|---|---|---|---|
| 0.8989970 | 0.0566500 | 0.4309880 | 0.9780040 | 0.9370920 | 0.5470200 | 2,510 | 2,792 |

**Figure F8:** Distribution of final sample weight for Wave 8 – K cohort longitudinal weight

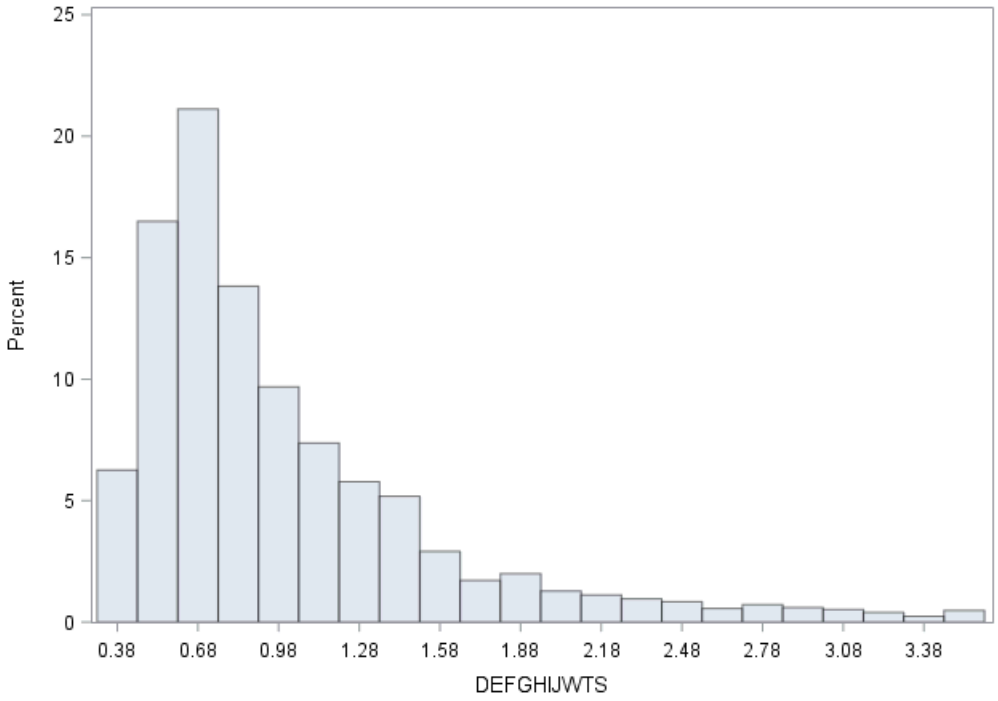


**Table F8:** Analysis variable: DEFGHIWTS – K cohort longitudinal weight

| Mean | Std Dev | Minimum | Maximum | Mode | Range | Sum | *N* |
|---|---|---|---|---|---|---|---|
| 1.0000000 | 0.5860100 | 0.3300000 | 3.5000000 | 0.3300000 | 3.1700000 | 2,510 | 2,510 |

# 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 8 (issued sample = 4,030)** | | | | |
| Interview | 3,127 | 3,127 | 61.2 | 100.0 |
| P1CASI | 3,123 | 3,086 | 60.4 | 98.8 |
| P2SC | 2,575 | 1,854 | na | 72.0 |
| PLECATI | 521 | 319 | na | 61.2 |
| TEACH | 3,060 | 2,317 | na | 75.7 |
| ACASI | 3,036 | 3,011 | 59.0 | 99.2 |
| TUD | 3,036 | 2,827 | 55.4 | 83.6 |
| | 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 8 (issued sample = 3,943)** | | | | |
| YP or P Interview | 3,037 | 3,037 | 60.9 | 100.0 |
| YP Interview | 3,037 | 2,708 | 54.3 | 89.2 |
| CASI | 2,708 | 2,656 | 53.3 | 98.1 |
| CAWI | 2,708 | 1,908 | 38.3 | 70.5 |
| CAWSI | 800 | 596 | na | 74.5 |
| PCATI - W7P1 | 3,207 | 2,635 | na | 82.2 |
| PCATI - W7P2 | 2,057 | 1,681 | na | 81.7 |
| PCATI - W7PLE | 473 | 317 | na | 67.0 |

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

# B instrument response

**CASI**

Of the B families interviewed in Wave 8, only 1% of Parent 1s did not complete the P1 CASI. Of the B study children who participated in Wave 8 again only 1% did not complete their ACASI.

**Parent 2 self-completed forms**

The response rate for Wave 8 B Parent 2s remains similar to Wave 7 at 72%.

**Parent Living Elsewhere (PLE) instrument**

Of the eligible B PLEs that interviewers attempted to contact in Wave 8 around 60% responded.

**Teacher self-completed form**

The teacher forms continue to achieve good response rates (76%) compared to 77% in Wave 7.

**Child interview**

The response rate for the Time Use Diary (TUD) for the B cohort remains high at 84%.

# K instrument response

**Interview**

In Wave 8 young people were approached prior to their parents for the first time. Study families were considered to be participating in Wave 8 if either the young person or one of the Wave 7 parental figures were interviewed.

**CAWI**

In Wave 8 the CAWI was introduced and the take up rate was around 70% of young people who participated In Wave 8.

**Parent interview**

In Wave 8 parents were interviewed via a telephone interview. The proportion of parents living elsewhere participating increased between Wave 7 and Wave 8 from 55% to 67%. Further, the total number of parents living elsewhere who were interviewed increased.