The Longitudinal Study of Australian Children:

LSAC Technical paper No. 10

Wave 5 weighting & non response

Benedict Cusack and Ryan Defina

November 2013

Australian Government

Australian Institute of Family Studies

Table of Contents

About the authors

Benedict Cusack is an Assistant Director in the Statistical Services Branch of the Australian Bureau of Statistics. He has an academic background in theoretical and experimental physics, and moved into the fields of statistics and survey methodology when he joined the Australian Bureau of Statistics in 2005. He has extensive experience in sample designs and estimation for cross-sectional surveys, particularly relating to official statistical collections such as the Quarterly Business Indicators, Monthly Retail Trade, the Australian Health Survey and the Australian Aboriginal and Torres Strait Islander Health Survey. He has experience with child education and child protection statistics through brief contracts with the Hamilton-Wentworth District School Board (Canada) and the NSW Department of Community Services. He is currently the lecturer for the internal ABS course Model Assisted Methods. He has directed and coordinated weighting analysis and implementation for Wave 5 of Longitudinal Survey of Australian Children.

Ryan Defina is a Senior Research Officer in the *Statistical Services Branch of the Australian Bureau of Statistics*. He holds degrees in actuarial and theoretical statistics, and is currently engaged in postgraduate studies specialising in generalised linear models and replicate methods as applied to the design and analysis of survey data. He has experience in providing methodological support for a number of ABS surveys including the Australian Health Survey, Monthly Retail Trade, Quarterly Business Indicators & Capital Expenditure surveys. Within the Statistical Society of Australia, Ryan holds the offices of New South Wales Branch Secretary and foundation member of the Continued Professional Development Committee. He is the chief research investigator for weighting analysis in Wave 5 of the Longitudinal Survey of Australian Children.

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 **Stephen Horn, Mark Sipthorp, Galina Daraganova, Suzanne Brown, Olathe Butler** and **Melissa Gare** for review and technical advice in the compilation of this weighting paper.

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 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 representativity 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

ERP	Estimated Resident Population
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

Introduction

The Longitudinal Survey of Australian Children's sample is intended to be representative of Australian children (citizens, permanent residents and applicants for permanent residency) in each of the two selected age cohorts, infants (ideally children aged under 12 months) and children aged 4 years, allowing the assessment of developmental outcomes from infancy until middle childhood. (Soloff, Lawrence & Johnstone, 2005).

The LSAC study has an accelerated cross-sequential design, with the two cohorts of children selected according to the following specifications:

- 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. In 2005, 2007 and 2009 and 2011 parents were also sent a between-waves mail survey. (AIFS, 2012.)

Wave 5 of the Longitudinal Study of Australian Children was conducted in 2012 with B-cohort children at age 8-9 and K-cohort children at age 12-13. 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 sample after missing a wave if contact can be re-established (for example if they return from overseas). The active sample is now less than 50% of the selected sample; however the sample size still remains firmly above the 2,500 children per cohort (around 1% of target population) originally envisaged in the LSAC design.

This weighting paper serves two purposes: describing the response properties and quality of the sample continuing into Wave 5; and describing the method and implementation of weight calculations to assist analysts make accurate population inferences from the LSAC sample.

The Use of Weighting in Analysis

Surveys often use probability samples to allow inferences about the population to be drawn within statistical variation. 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.

Representativeness of the sample can be affected by missing data from 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 can't be contacted or don't 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 probability design 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 5 of the Longitudinal Study of Australia 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 representativity. We also offer a timely reminder to users that LSAC is based on a clustered sample design, and hence that use of the cluster variable is required for valid inference when conducting statistical tests to avoid overstating significance.

Summary of Sample Design

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 survey 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 kindies) who were 0 years and 4 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 who contact records were available through Medicare, who could be contacted, and who agreed to participate in LSAC. For subsequent waves: The subset of Wave 1 coverage who could be contacted. This included tracking address changes and re-recruitment after missing waves where possible, including cases of temporarily moving overseas.
Stratification (division of population into cells from which sample was drawn)	Cells of State x Capital City/Balance of State x Large/Small Postcode
Selection Frame (from which children were selected and contact details obtained)	List Frame of Medicare Records for Children in Scope
Sample Design	Multi-Stage Cluster Sampling
Selection Unit(s)	Stage 1 Unit: Postcode Stage 2 Unit: 1 Cluster of Dwellings within Postcode Stage 3 Unit: Children in Dwellings in Cluster
Reporting Unit(s)	Parent 1, Parent 2, (when old enough) Child, Interviewer, Child care worker, Teacher
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.

General Approach to Waves 1-4

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 by state by part of state were drawn from ABS Estimated Resident Population as at March 2004, and benchmarks for households by language spoken at home and mother's education level within region were generated using proportions taken from 2001 Census.

Weights for Waves 2 to 4 were calculated adjusting previous wave weights for differential sample attrition in two stages, restoring additive alignment to population benchmarks (Sipthorp, Misson 2007; Sipthorp, Misson 2009; Sipthorp, Daraganova 2011). At the first stage a modelled response propensity factor was applied; at the second a post-stratum grossing factor was introduced. Extreme weights were capped as a form of outlier treatment to avoid any particular child representing many more than other children in the sample, because this can potentially lead to volatile statistics if any such child has unusual characteristics.

In each case, a population weight is calculated which adds up to the number of children in the population, and a sample weight is calculated which adds up to the number of children in the sample. The sample weights are therefore equal to the population weights multiplied by the sampling fraction.

Given that children who were out of contact during an earlier wave can be recruited back into the sample at a later stage, it has been the practice to calculate a distinct set of weights for every combination of response to maximise the sample available for analysis. This does however result in an exponentially growing number of weights with each wave requiring double the number of the previous wave. For example, Wave 4 included calculation of 16 weights: population and sample weights for each of the B and K cohorts for 4 combinations of response:

- Respondents to Waves 1, 2, 3 and 4 (the longitudinal weight, for respondents to all waves)
- Respondents to Waves 1, 2 and 4
- Respondents to Waves 1, 3 and 4
- Respondents to Waves 1 and 4 (the cross-sectional weight, for all respondents to most recent wave)

Following this pattern, Wave 5 would require 32 weights, increasing to 256 weights by Wave 8.

Approach to Wave 5

The substance of weighting methods and analysis for Wave 5 is the same as for Waves 2 to 4. In particular:

- Population and sample weights were calculated for the B and K cohorts;
- Wave 5 final weights were calculated by inflating previous wave weights by a factor reflecting the magnitude and distribution of non-response. This was done in the same manner as past waves, in two steps:
    - weight adjustment by response propensity modelling using logistic regression;
    - weight adjustment to align within post-stratum counts at the time of selection;
- Caps were applied to extreme weights occurring for a small number of units.

Some key changes were made to the manner in which weighting methods and analysis were carried out. In particular, a concise set of 8 weights were calculated for Wave 5 for two combinations of response:

- Respondents to Waves 1, 2, 3, 4 and 5 (the longitudinal weight, for respondents to all waves)
- Respondents to Waves 1 and 5 (the cross-sectional weight, for all respondents to most recent wave)

Future waves will be released with a set of 8 weights following this pattern.

Other changes include:

- To reduce the risk of overfitting, the logistic regression model used a smaller number of regression variables;
- The weighting method mathematics are reviewed in this paper with a view to clarifying and improving transparency;
- A systematic approach was taken to quality assurance and reporting for the weighting calculation process.

Weighting Methods and Quality

This section contains descriptions of the methods applied in calculating final Wave 5 weights, the choices involved in determining appropriate inputs, and the approach to quality assurance for these methods and the data they produce. Methods are essentially the same as applied for Waves 2, 3 and 4. Implementation of the post-stratum weight adjustment step and weight capping has been condensed, and a detailed description is provided in this section for transparency.

Table 2. Algebraic symbol definitions for weighting methods

c	Current wave c, which in this case is Wave 5
p	Previous wave p, which is either Wave 4 (longitudinal weights) or Wave 1 (cross-sectional weights)
U	LSAC scope population U of children i from which sample was drawn (for either B-cohort or K-cohort)
N	Total number of children i in population U (B-cohort or K-cohort)
s	Randomly selected LSAC sample of children i drawn from population U
n	Total number of children i in sample s (B-cohort of K-cohort)
π_i	Selection probability for child i at time of recruitment from population U
$r_{c\|p}$	Respondent group of children i (subset of s) for which data collected for current wave c, given that they responded to previous wave(s) p
$n_{c\|p}$	Number in respondent group of children i for current wave c, given that they responded to previous wave(s) p
y_i	Study variable y for child i (for some particular variable of interest in some specified wave)
Y_T	Aggregate statistic (total) of study variable y for all children i in population U
$\hat{Y}_T$	Estimate of aggregate statistic (total) of study variable y for responding children i in LSAC sample
$w_{ci}^{(I)}$	Initial sample weight for current wave c, child i, prior to response propensity weight adjustment step
$w_{ci}^{(RPA)}$	Response propensity adjusted sample weight for current wave c, child i, after response propensity weight adjustment step but before post-stratum weight adjustment step
$w_{ci}^{(F)}$	Final sample weight for current wave c, child i, after post-stratum weight adjustment step
$w_{pi}^{(F)}$	Final sample weight for a previous wave p, child i
$W_{ci}^{(F)}$	Final population weight for current wave c, child i
$\hat{\Phi}_{i,c\|p}$	Modelled response propensity for child i to respond in current wave c given past response in previous wave p
$g_{i,c\|p}$	Post-stratum weight adjustment factor g for child i

Survey estimation of descriptive statistics is done by multiplying study variables by weights and summing. For example, estimating for the total and mean of y (or the count and percentage if y is a membership variable) can be done via:

$$\hat{Y} = \sum_{i \in s} W_i y_i \qquad \hat{\bar{Y}} = \frac{1}{N} \sum_{i \in s} W_i y_i$$

Here the population weight W (capital) is defined as distinct from the sample weight w (lower case) and they have the following properties:

$$\sum_{i \in s} W_i = N \qquad \sum_{i \in s} w_i = n \qquad \frac{W_i}{N} = \frac{w_i}{n}$$

The population weight can be thought of as the number of children in the wider population that each sampled child i represents. LSAC has traditionally worked primarily with the sample weight which is by definition mean 1 and sometimes preferred for analysis. For example a mean (or percentage) could be estimated using the sample weight by:

$$\hat{\bar{Y}} = \frac{1}{n} \sum_{i \in s} w_i y_i$$

If the population weights are the inverse of the sample selection probabilities, then these estimators are forms of the Horvitz-Thompson estimator (demarked with a subscript π) and are unbiased estimators for each corresponding population quantity.

$$\hat{Y}_\pi = \sum_{i \in s} \pi_i^{-1} y_i \qquad \hat{\bar{Y}}_\pi = \frac{1}{N} \sum_{i \in s} \pi_i^{-1} y_i$$

However, in the presence of non-response, data are only available for a subset of selected sample s, and the weights are adjusted upward to compensate for the missing data in an attempt to mitigate bias that may be present. So for Wave c we would use:

$$\hat{\bar{Y}} = \frac{1}{n_{c|p}} \sum_{i \in r_{c|p}} w_{ci}^{(F)} y_i$$

Weight adjustment is an opportunity to adjust the relative contribution of respondents to improve the accuracy of survey estimates, both in terms of systematic bias from missing data and from random sampling variance.

Wave 1 had just one set of weights to account for non-response and frame undercoverage during the sample recruitment process. After Wave 1, LSAC weight adjustment has involved a sequence of weight adjustments to account for each case of sample attrition for each wave. Two types of weights are defined.

- 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

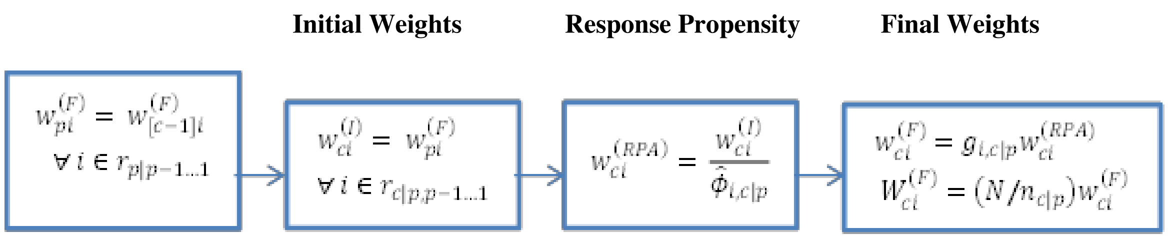
wave 1. Cross-sectional weights are most suitable for analysis that makes use only of the current data;

(All respondents to any wave were also respondents in Wave 1 by definition because non-respondents to Wave 1 were not recruited or followed up in later years.)

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 post-stratum adjustment is applied to re-align weight totals with known totals from the original sample. Both stages contribute to non-response bias reduction.

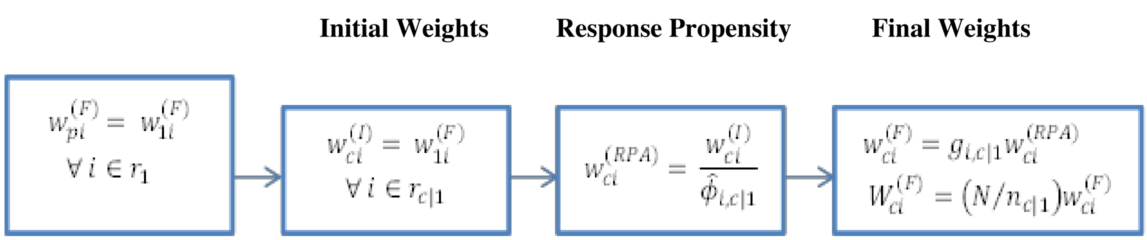
Calculating longitudinal weights from scratch would involve applying the weighting process c-1 times. Given previous wave weights are already calculated, the longitudinal weights can be defined inductively, starting with the longitudinal weight from the previous survey.

Figure 1. Longitudinal weighting schematic



Cross-sectional weights require one iteration of the weighting process, starting with the final weight from Wave 1.

Figure 2. Longitudinal weighting schematic



In previous waves, weights were calculated for respondents to other combinations of waves, for example responding to waves 1, 3 and 4 but not 2. These were also calculated recursively as per Figures 1 and 2. These were referred to as cross-sectional weights. Explicit calculation of these weights was discontinued for Wave 5 given the exponentially increasing number of combinations.

Initial Weights

For initial weights for a wave weighting process, the final weights of a previous wave are carried forward.

For Wave 5 Longitudinal Weights (which apply to the subset of respondents to Waves 1, 2, 3, 4 and 5), the initial weight for Wave 5 is the final Longitudinal weight from Wave 4.

$$w_{5i}^{(I)} = w_{4i}^{(F)}$$

For Wave 5 Cross-sectional Weights (which apply to the larger subset of respondents to Waves 1 and 5, who may or may not have also responded to intervening waves), the initial weight for Wave 5 is the final weight from Wave 1.

$$w_{5i}^{(I)} = w_{1i}^{(F)}$$

Because the Wave 5 responding set is a strict subset of the previous wave responding cohort, this is achieved by copying over weights for the responding subset.

For quality checking, it is expected the sum of initial sample weights for the Wave 5 responding subset to be roughly equal to the size of the relevant Wave 5 responding sample, simply because sample weights are by definition mean 1.

Response Propensity Modelling

The purpose of this step is to adjust for differential non-response by particular demographic groups which 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 unit weight is inflated by 1/90%.

Mathematically, a logistic regression is conducted to model the response propensity $\hat{\phi}_{i,c|p}$ for each unit i via:

$$log\left(\frac{\hat{\phi}_{i,c|p}}{1-\hat{\phi}_{i,c|p}}\right) = \boldsymbol{x}_{pi}^{T}\boldsymbol{\beta}$$

where $\boldsymbol{x}_{pi}^{T}$ is a vector of response variables or demographics for each child i from the past wave p, and β is a vector of coefficients estimated by maximum likelihood estimation. The modelled response propensities for responding units are then derived directly by:

$$\hat{\phi}_{i,c|p} = \left(1 + exp\{-\boldsymbol{x}_{pi}^{T}\boldsymbol{\beta}\}\right)^{-1}$$

The response propensity adjusted weights are derived by estimating the total probability of response $\hat{\pi}_{i,c}$ for all specified waves up to and including the current wave.

$$\hat{\pi}_{i,c} = Pr(Response\ to\ current|Response\ to\ previous).Pr(Response\ to\ previous)$$
$$= \hat{\phi}_{i,c|p}\hat{\pi}_{i,p}$$

The desired weight is the inverse of this probability, so for each responding unit, the weight is then adjusted by the inverse of the modelled response propensity via:

$$w_{ci}^{(RPA)} = \frac{w_{ci}^{(I)}}{\hat{\phi}_{i,c|p}}$$

For longitudinal weights, the set of previous waves p is all past waves (in Wave 5 this is Waves 1, 2, 3 and 4), and for cross-sectional weights, the set of previous waves p is just Wave 1.

A systematic modelling approach was taken to select the suite of logistic regression model variables $\boldsymbol{x}_{pi}^{T}$ which have the strongest discrimination between response and non-response groups. This involved some degree of pre-processing for variable candidates, such as recoding and collapsing categories. In a small number of cases where missing values were found for variable candidates, standard techniques like Last Observation Carried Forward (LOCF) or modal imputation were employed. Generally speaking, variables with large counts of missing data were not good candidates and (even with imputation) do not tend to show significance in logistic regression given that basic imputation methods do not generally enhance explanatory power.

For quality checking , it is expected the sum of weights following this step to be approximately equal to the number of responding children in the previous wave, which should be the same as the number of responding children in the current wave divided by the average response propensity (ie the wave-on-wave response rate).

Selection of covariates for logistic regression non-response adjustment

Covariate selection is often characterised as a trade-off between model fit and parsimony. It is possible to fit an extensive suite of variables, but at the cost of introducing noise that makes predictions less reliable from a variance perspective (over-fitting).

The approach taken for Wave 5 was to err on the side of parsimony, selecting only a small number of covariates. Concepts in support of this approach include;

1. A reduction in model volatility.
2. Promotion of robustness in the underlying model.
3. Simplicity, aiding data users should they seek to replicate the process used within this paper.
4. Assertion that the subsequent calibration step will capture a bulk of the residual systematic non-response bias.

Further to this, priority was also given to consistency between waves of response and also the four different weights produced (sample-population pairs). That is to say, variables selected are typically seen in previous waves and/or are predictors of response for more than one Wave 5 weight.

In assessing the statistical significance of the covariates aforementioned, then were some considerations. Firstly, in the case where missing values were present, a crude imputation procedure was applied. For continue covariates, the median value at the cohort level was applied – median chosen to form a robust guard against skewness in the variable distribution. Categorical covariates were imputed using the cohort level mode. While there is risk in using crude imputation methods, the aim was to ensure that no record was removed from training the model. Any potential bias at the model parameter estimates should be washed out by the removal of non-response bias, and also through process attached to the subsequent calibration step.

Once the datasets were run through the pre-processing stage, more technical metrics were consulted in deciding upon variable inclusion. Type 3 Wald Chi-Square statistics are seen as the most informative metric, particularly for the case of categorical covariates (Engle, 1983). This is due to the ability to summarise covariate performance across the multiple indicator variables fitted to the assigned categorical levels. If one is to investigate the Wald Chi-Square statistics of the individual levels, then an informed decision is difficult to arrive at given the vast array of p-values presented to the analyst.

The R-Square and closely related Max-rescaled R-Square diagnostics were not used to assess model fit. Such a metric was designed for, and indeed is only valid for use in assessing the performance of linear models with an underlying normal response distribution. Once one leaves this realm and looks at non-normal models such as logistic regression, other measures are more appropriate. Such a measure includes the Deviance. It allows for formal testing of goodness-of-fit and thus comparison of model performance for different combinations of potential covariates. Another useful measure for the logistic model is the AUC and associated ROC curve. AUC measures around the 70% mark were common for all four weight pairs, a reasonable, albeit not ideal performance (Swets 1973).

Post-Stratum Weight Adjustment

The purpose of this step is to restore representativity of the sample for population inference by re-aligning 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 again. The original selection strata are cells of state by part-of-state by large/small postcode. Aligning the sample within each of these cells to known totals is called post-stratification (Holt & Smith 1979).

Similarly to stratification, post-stratification involves dividing the population and sample into mutually exclusive groups, but is done after selection and data collection and can use information collected during surveying to do so (hence ‘post’). In the case of LSAC, the post-strata is defined to be identical to the strata; no new information from data collection was used in defining post-strata for the purpose of weight adjustment. Post-stratum weight adjustment can be done only with demographic variables that are known for both the sample and the population.

Post-stratification can achieve several goals:

- Reduction of sampling variance, if the post-strata categories are correlated with survey variables;
- Non-response bias reduction, if in addition to a correlation there is differential non-response across post-strata;

- Alignment of weight totals to known population benchmark totals. (Post-stratification is sometimes called benchmarking.)

Post-stratification was used to perform non-response adjustment for Wave 1 by aligning population weights to known totals within strata, and further cross-classifying by two additional non-response variables from Census data. Estimated Resident Population benchmarks were obtained from the Demography section of the Australian Bureau of Statistics at the time of Wave 1 weighting.

The post-stratum adjustment step for subsequent waves – re-aligning back to the Wave 1 composition - is equivalent to direct post-stratification to the population of children (B and K cohorts) as at the time of Wave 1 sample recruitment in 2004, which remains the reference population for LSAC. In this sense, the post-stratum adjustment includes an additional non-response treatment for differential attrition across cells of state by part-of-state by large/small postcodes, over and above the non-response treatment handled by the previous response propensity modelling step.

One restriction on post-stratification is that it requires known population totals cross-classified in the same way as the post-strata, and hence this method cannot make use of data collected by respondents to past waves if it is not known for the population at large. The current configuration of two successive weight adjustment steps makes use of known population totals as well as past data collected from LSAC respondents when treating non-response.

Weighting processes for past waves have handled the post-stratum adjustment in a large number of steps, for example by aligning the total sum of sample weights separately from re-balancing the distribution of weights across strata, and then repeating a second time after capping weights. Here the process is handled in one compact step.

Standard post-stratification is done by dividing the sample s and population U into some number of post-strata k using categorical variables (ensuring each post-strata is not empty ie. $n_k > 0$), and then estimating via:

$$\hat{Y}_{PS} = \sum_k \frac{N_k}{\hat{N}_{k\pi}} \hat{Y}_{k\pi} = \sum_k \frac{N_k}{\sum_{i \in s_k} \pi_i^{-1}} \sum_{i \in s_k} \pi_i^{-1} y_i = \sum_k G_i \sum_{i \in s_k} \pi_i^{-1} y_i = \sum_{i \in s} G_i \pi_i^{-1} y_i$$

The adjustment factor G_i, called the g-weight, equals the ratio of the known population total to the estimate of the population total give the sample and weights within post-stratum k.

$$G_i = \frac{N_k}{\hat{N}_{k\pi}}$$

Hence if the sample has more (fewer) units compared to chance in post-stratum k then this factor adjusts the weights down (up) to compensate. In fact, it can be shown that the total adjusted weights add up exactly to the population total by estimating the population total itself by substituting N in place of Y.

$$\hat{N}_{PS} = \sum_k \frac{N_k}{\hat{N}_{k\pi}} \hat{N}_{k\pi} = \sum_k N_k = N$$

This also holds within each post-stratum, which is why post-stratification is sometimes called calibration or benchmarking (Sarndal, Swensson & Wretman 1992).

For the situation where there is non-response, the adjusted weights are substituted and the sum is taken over the responding sample, whence post-stratification also performs a non-response adjustment function to the extent that the post-strata k may be related to non-response.

$$\hat{Y}_{PS} = \sum_k \frac{N_k}{\sum_{i\in r_{c|p}} W_{ci}^{(RPA)}} \sum_{i\in r_{c|p,k}} W_{ci}^{(RPA)} y_i = \sum_{i\in r_{c|p}} G_{i,c|p} W_{ci}^{(RPA)} y_i$$

Given that LSAC weighting typically works with the sample weights rather than population weights, we can define a corresponding g-weight to apply to the sample weights via:

$$\hat{\hat{Y}}_{PS} = \frac{1}{n_{c|p}} \sum_k \frac{n_{c|p}}{N} \frac{N_k}{\sum_{i\in r_{c|p}} w_{ci}^{(RPA)}} \sum_{i\in r_{c|p,k}} w_{ci}^{(RPA)} y_i = \sum_{i\in r_{c|p}} g_{i,c|p} w_{ci}^{(RPA)} y_i$$

This also holds within each post-stratum, which is why post-stratification is sometimes called calibration or benchmarking (Sarndal, Swensson & Wretman 1992).

Where

$$g_{i,c|p} = \frac{n_{c|p}}{N} \frac{N_k}{\sum_{i\in r_{c|p}} w_{ci}^{(RPA)}}$$

Hence we have the final weight as

$$w_{ci}^{(F)} = g_{i,c|p} w_{ci}^{(RPA)}$$

It is straightforward to demonstrate that the weight sums align with intended totals:

$$\sum_{i\in r_{c|p,k}} w_{ci}^{(F)} = \frac{n_{c|p}}{N} N_k \qquad \sum_{i\in r_{c|p,k}} W_{ci}^{(F)} = N_k \qquad \frac{W_{ci}^{(F)}}{N} = \frac{w_{ci}^{(F)}}{n_{c|p}}$$

There are no method decisions to make in implementing the post-stratum adjustment step. The post-strata k are the same as the selection strata set as part of the sample design in Wave 1.

Weight Capping

Weight capping is the process of limiting extreme large 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. Weight capping is therefore 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 loss of representativity for some respondent groups.

The weight capping process was similar to that used in previous waves, but has been simplified to one single step occurring during the stratum adjustment. The weight capping is done iteratively, with the weights of other units adjusted upward to compensate, and this process is repeated if other unit's weights would exceed caps. (See Appendix F for a description of this process as implemented in the ABS GregWT macro.)

Quality Assurance for Final Weights

The following section details a raft of quality checks that ensure the published weights meet reasonable standards of quality. It's also an opportunity to highlight key qualities in the weights that the analyst would find desirable when forming survey estimates, or fitting models to the data.

Micro checks

Micro checks seek assess the data at the record level, investigating peculiarities that may lead to distortion at an aggregate level. The identification of an unusual result at the micro level may not affect meaning analysis in practice, but may form evidence to support an incorrect implementation of the weighting process.

Distribution of estimated response propensities

Response propensities were estimated using the following covariates:

- family member's age last birthday
- SEIFA economic resources
- highest level of school completed by mother
- P2 in the home
- Mother first spoke a language other than English

The specific data items used for the four pairs of weights were;

Table 3. Data items used in logistic regression model

Weight pair	Continuous covariates used	Categorical covariates used
B cohort – longitudinal	DF03DP1 DCNFSER	DFD08M1 DP2SCD DFD11M2
B cohort – cross sectional	AF03M3 AF03M2	AFD08M1 AP1SCD AFD11M2
K cohort – longitudinal	FF03FP1 FCNFSER	FFD08M1 FFD11M2 FP2SCD
K cohort – cross sectional	CF03M2	CFD08M1 CFD11M2 CP1SCD

In order to validate the logistic regression non-response adjustment procedure, the estimated response propensities were plotted. One is interested in ensuring that desirable properties are retained within the probabilities. This ensures that the subsequent calibration step is commencing with weights 'bumped up' by an appropriate child-level grossing factor.

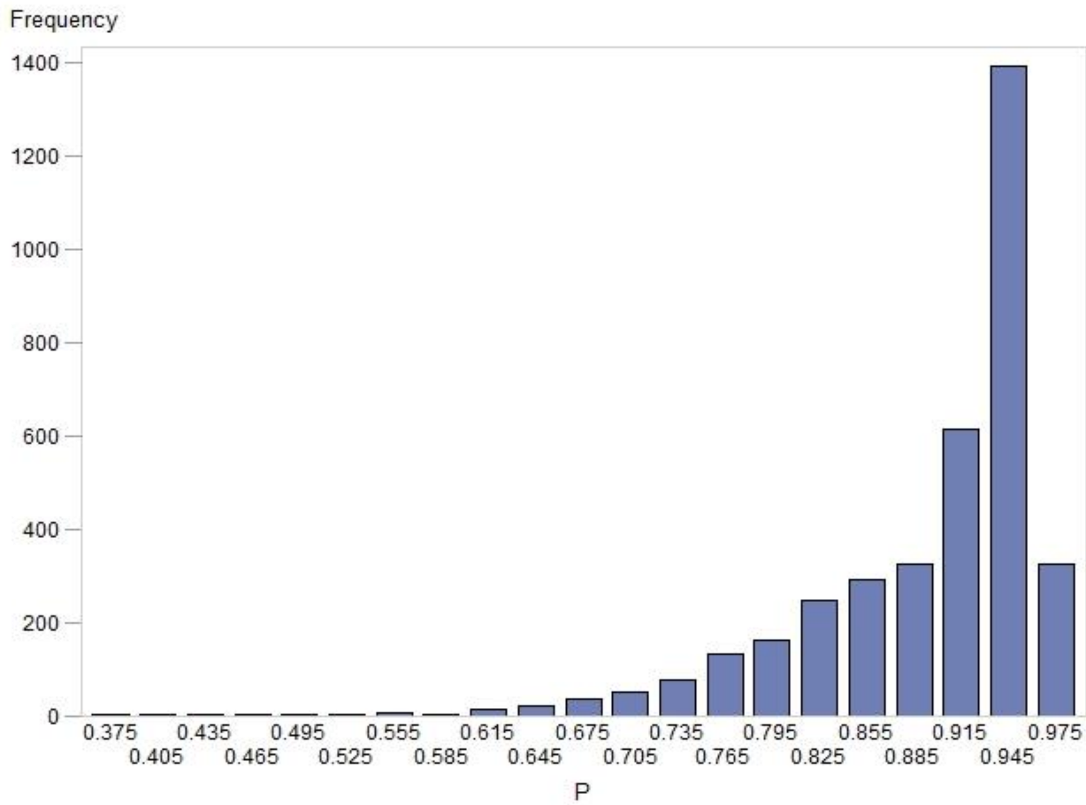


Figure 3. Distribution of estimated response propensities – B cohort longitudinal sample weight

Table 4. Analysis of variable P – B cohort longitudinal sample weight

Analysis Variable: P							
Mean	**Standard Deviation**	**Minimum**	**Maximum**	**Modal Range**	**Range**	**Sum**	**Count**
0.89	0.0786	0.39	0.99	0.70-0.80	0.60	3360	3758

Table 5. Observed response propensity properties – B cohort longitudinal sample weight

Response propensity property	**Observed?**
Left skewness in the distribution of propensities	YES
Relatively small number of propensities lying outside the bounds of precedent from previous waves (the majority of propensities have historically been between 0.70 and 0.98)	YES
Modal propensity lies around the 0.90 mark, reflecting the Australia-level sample retention rate (see section 3.1 for details)	YES
Propensities should typically be higher than that of the cross sectional equivalent for the cohort (time period between response is shorted)	YES

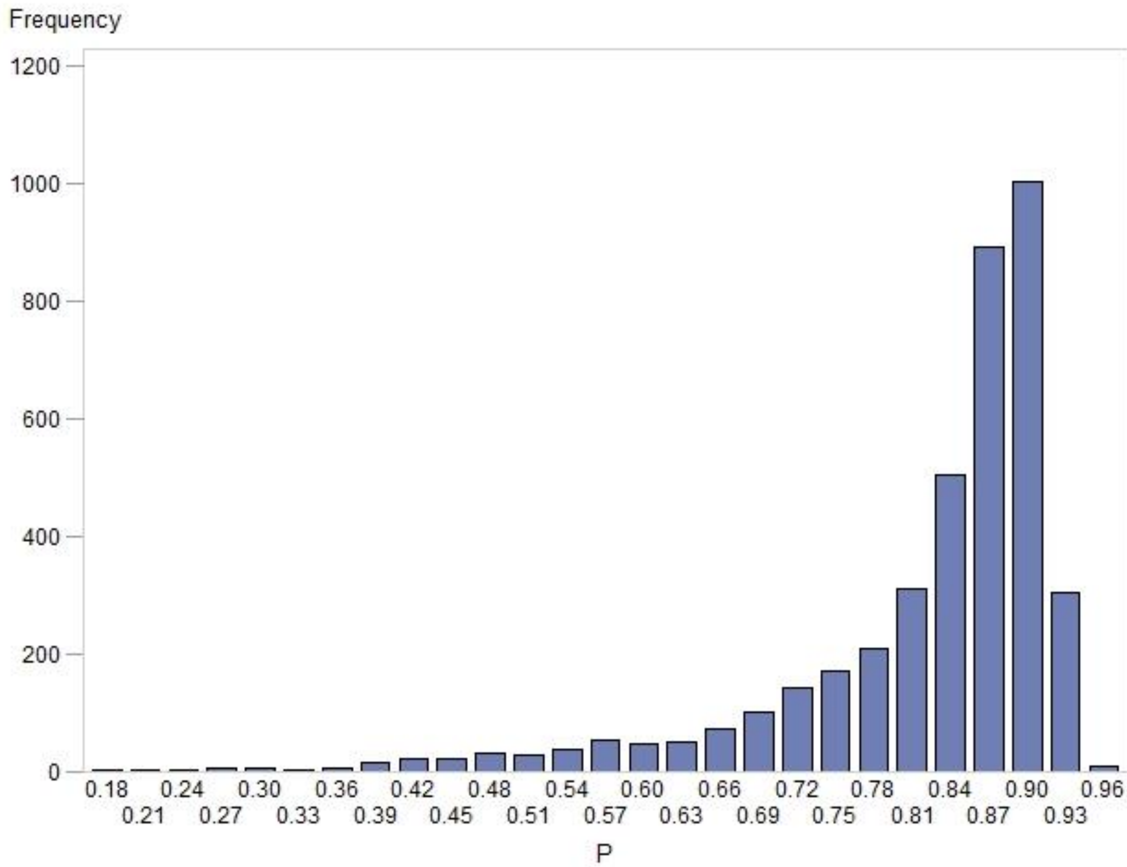


Figure 4. Distribution of estimated response propensities – B cohort cross sectional sample weight

Table 6. Analysis of variable P – B cohort cross sectional sample weight

Analysis Variable: P							
Mean	**Standard Deviation**	**Minimum**	**Maximum**	**Modal Range**	**Range**	**Sum**	**Count**
0.82	0.1135	0.19	0.96	0.80-0.90	0.77	3361	4085

Table 7. Observed response propensity properties – B cohort cross sectional sample weight

Response propensity property	Observed?
Left skewness in the distribution of propensities	YES
Relatively small number of propensities lying outside the bounds of precedent from previous waves (the majority of propensities have historically been between 0.70 and 0.98)	YES
Modal propensity lies around the 0.90 mark, reflecting the Australia-level sample retention rate (see section 3.1 for details)	YES
Propensities should typically be lower than that of the longitudinal equivalent for the cohort (time period between response is longer)	YES

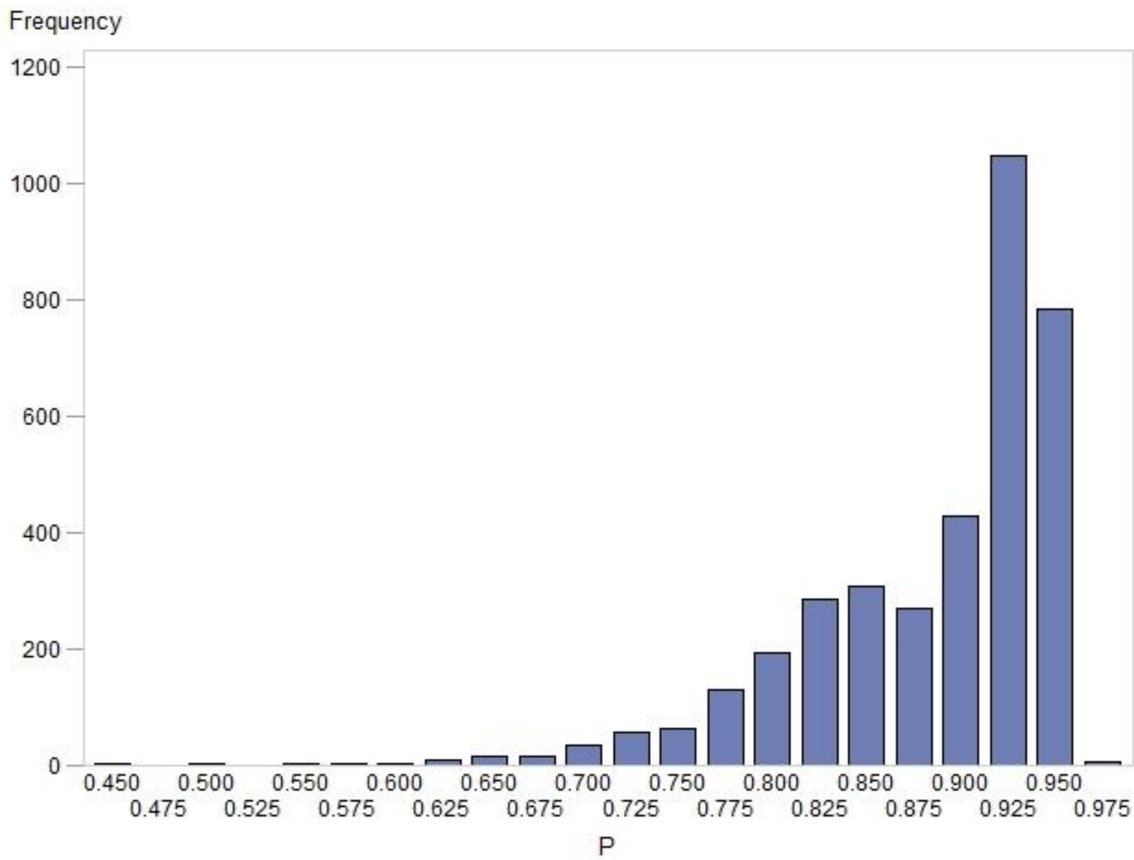


Figure 5. Distribution of estimated response propensities – K cohort longitudinal sample weight

Table 8. Analysis of variable P – K cohort longitudinal sample weight

Analysis Variable: P							
Mean	**Standard Deviation**	**Minimum**	**Maximum**	**Modal Range**	**Range**	**Sum**	**Count**
0.89	0.0669	0.45	0.98	0.70-0.80	0.53	3275	3682

Table 9. Observed response propensity properties – K cohort longitudinal sample weight

Response propensity property	**Observed?**
Left skewness in the distribution of propensities	YES
Relatively small number of propensities lying outside the bounds of precedent from previous waves (the majority of propensities have historically been between 0.70 and 0.98)	YES
Modal propensity lies around the 0.90 mark, reflecting the Australia-level sample retention rate (see section 3.1 for details)	YES
Propensities should typically be higher than that of the cross sectional equivalent for the cohort (time period between response is shorted)	YES

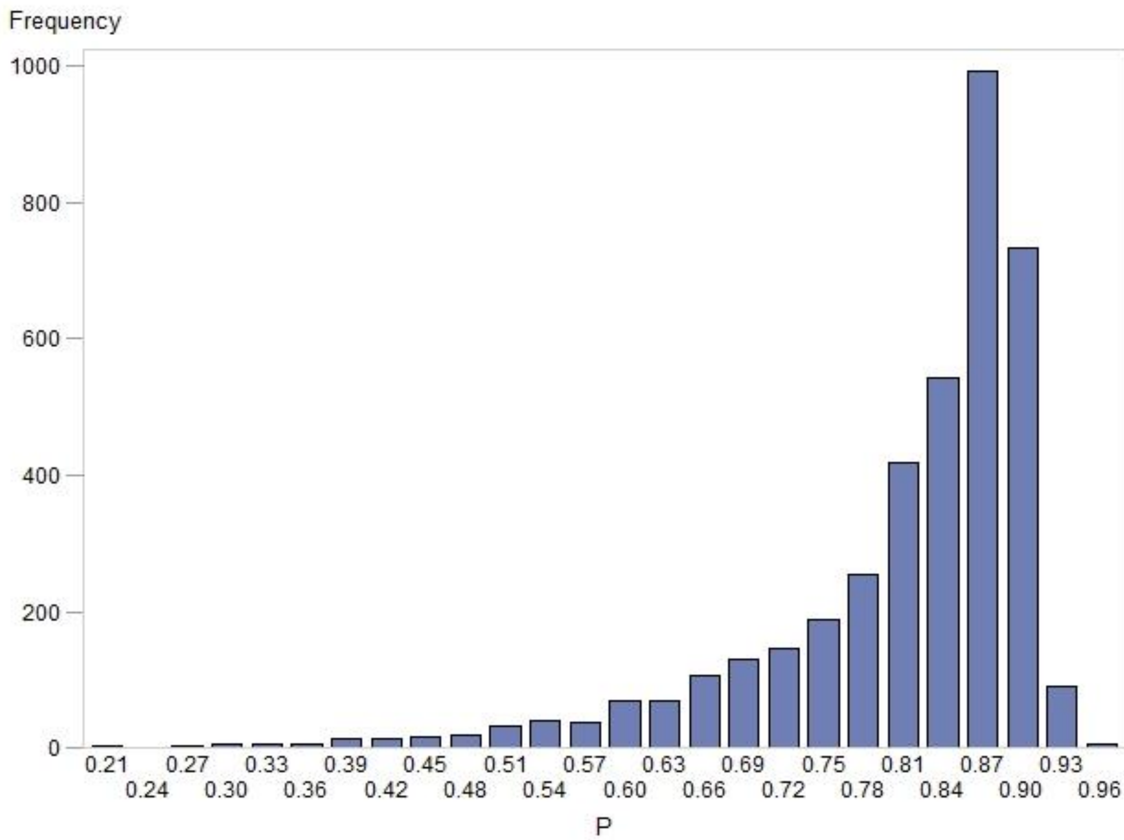


Figure 6. Distribution of estimated response propensities – K cohort cross sectional sample weight

Table 10 Analysis of variable P – K cohort cross sectional sample weight

Analysis Variable: P							
Mean	**Standard Deviation**	**Minimum**	**Maximum**	**Modal Range**	**Range**	**Sum**	**Count**
0.81	0.1043	0.22	0.96	0.70-0.80	0.74	3215	3956

Table 11. Observed response propensity properties – K cohort cross sectional sample weight

Response propensity property	**Observed?**
Left skewness in the distribution of propensities	YES
Relatively small number of propensities lying outside the bounds of precedent from previous waves (the majority of propensities have historically been between 0.70 and 0.98)	YES
Modal propensity lies around the 0.90 mark, reflecting the Australia-level sample retention rate (see section 3.1 for details)	YES
Propensities should typically be lower than that of the longitudinal equivalent for the cohort (time period between response is longer)	YES

Distribution of weights, and outliers

In order to assess the general behaviour of the weights produced, a set of plots are shown below. One is seeking to check that useful properties in the weights are retained in order to validate population inference as a methodologically sound form of analysis.

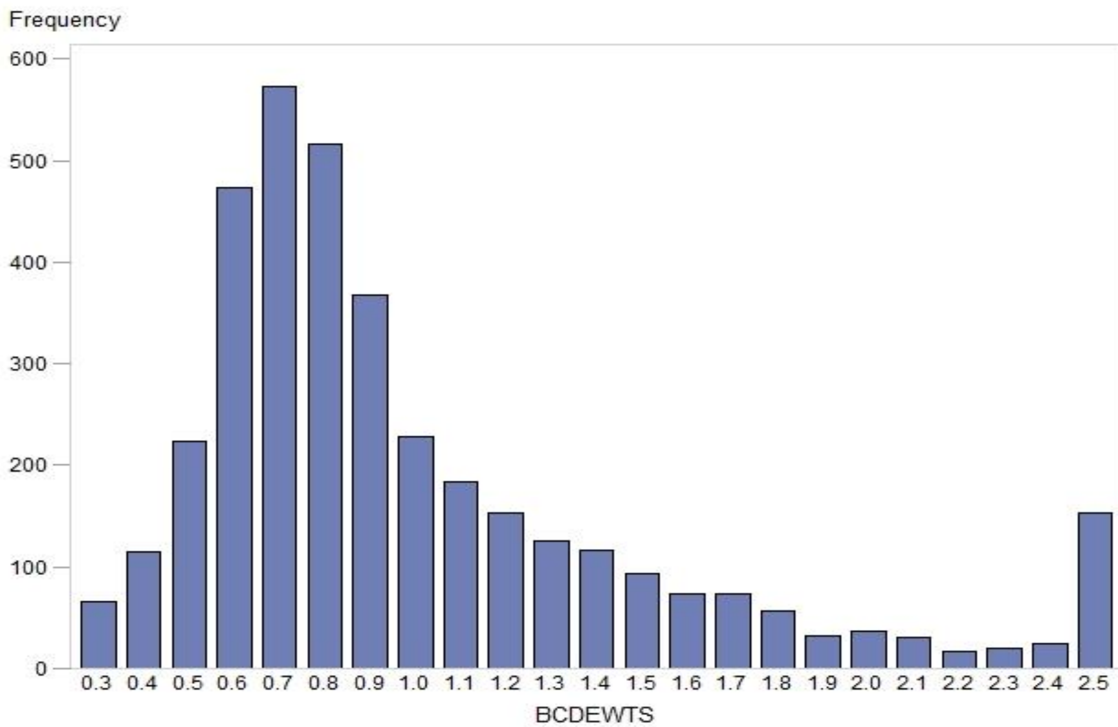


Figure 7 Distribution of BCDEWTS – B cohort longitudinal sample weight

Table 12. Analysis of variable BCDEWTS – B cohort longitudinal sample weight

Analysis Variable: BCDEWTS							
Mean	**Standard Deviation**	**Minimum**	**Maximum**	**Modal Range**	**Range**	**Sum**	**Count**
1.00	0.5112	0.33	2.50	0.93-0.96	2.17	3758	3758

Table 13. Observed weighting properties – B cohort longitudinal sample weight

Weighting property	Observed?
Right skewness in the distribution of weights	YES
Relatively small number of weights hitting the upper threshold of 2.50	YES
Modal weight less than one (by definition if the mean is one)	YES
Distribution can be regarded as approximately smooth, with an absence of volatile spikes	YES
Generally comparable with weights in previous waves with respect to shape/distribution, modal weight and proportion at upper threshold	YES

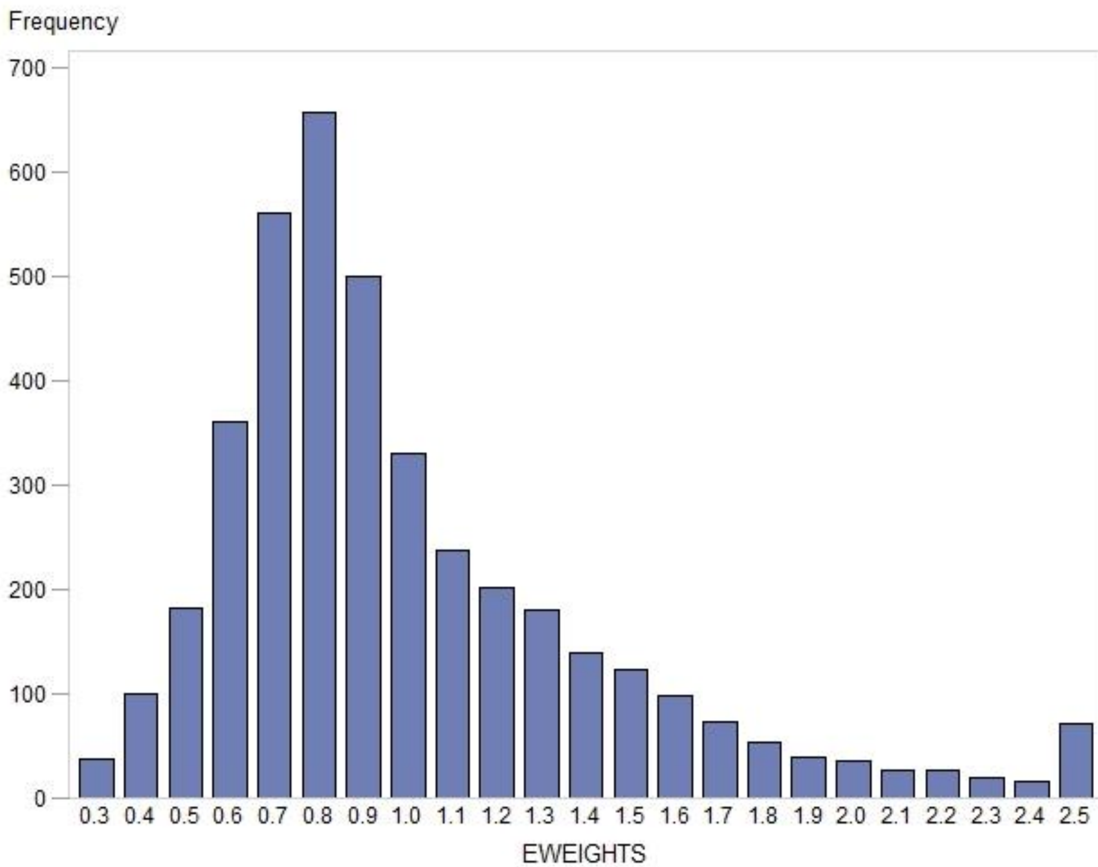


Figure 8. Distribution of EWEIGHTS – B cohort cross sectional sample weight

Table 14. Analysis of variable EWEIGHTS – B cohort cross sectional sample weight

Analysis Variable: EWEIGHTS							
Mean	**Standard Deviation**	**Minimum**	**Maximum**	**Modal Range**	**Range**	**Sum**	**Count**
1.00	0.4366	0.33	2.50	0.89-0.92	2.17	4085	4085

Table 15. Observed weighting properties – B cohort cross sectional sample weight

Weighting property	**Observed?**
Right skewness in the distribution of weights	YES
Relatively small number of weights hitting the upper threshold of 2.50	YES
Modal weight less than one (by definition if the mean is one)	YES
Distribution can be regarded as approximately smooth, with an absence of volatile spikes	YES
Generally comparable with weights in previous waves with respect to shape/distribution, modal weight and proportion at upper threshold	YES

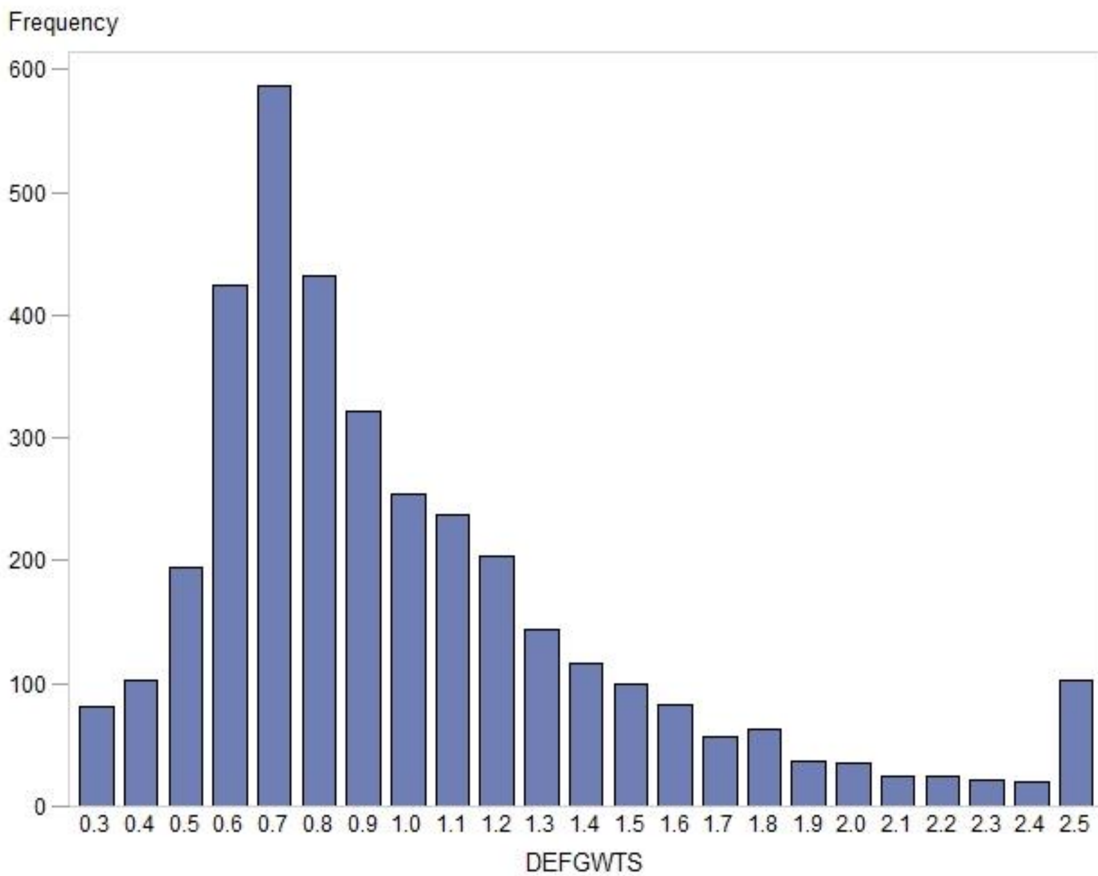


Figure 9. Distribution of DEFGWTS – K cohort longitudinal sample weight

Table 16. Analysis of variable DEFGWTS – K cohort longitudinal sample weight

Analysis Variable: DEFGWTS							
Mean	**Standard Deviation**	**Minimum**	**Maximum**	**Modal Range**	**Range**	**Sum**	**Count**
1.00	0.4822	0.33	2.50	0.91-0.93	2.17	3682	3682

Table 17. Observed weighting properties – K cohort longitudinal sample weight

Weighting property	Observed?
Right skewness in the distribution of weights	YES
Relatively small number of weights hitting the upper threshold of 2.50	YES
Modal weight less than one (by definition if the mean is one)	YES
Distribution can be regarded as approximately smooth, with an absence of volatile spikes	YES
Generally comparable with weights in previous waves with respect to shape/distribution, modal weight and proportion at upper threshold	YES

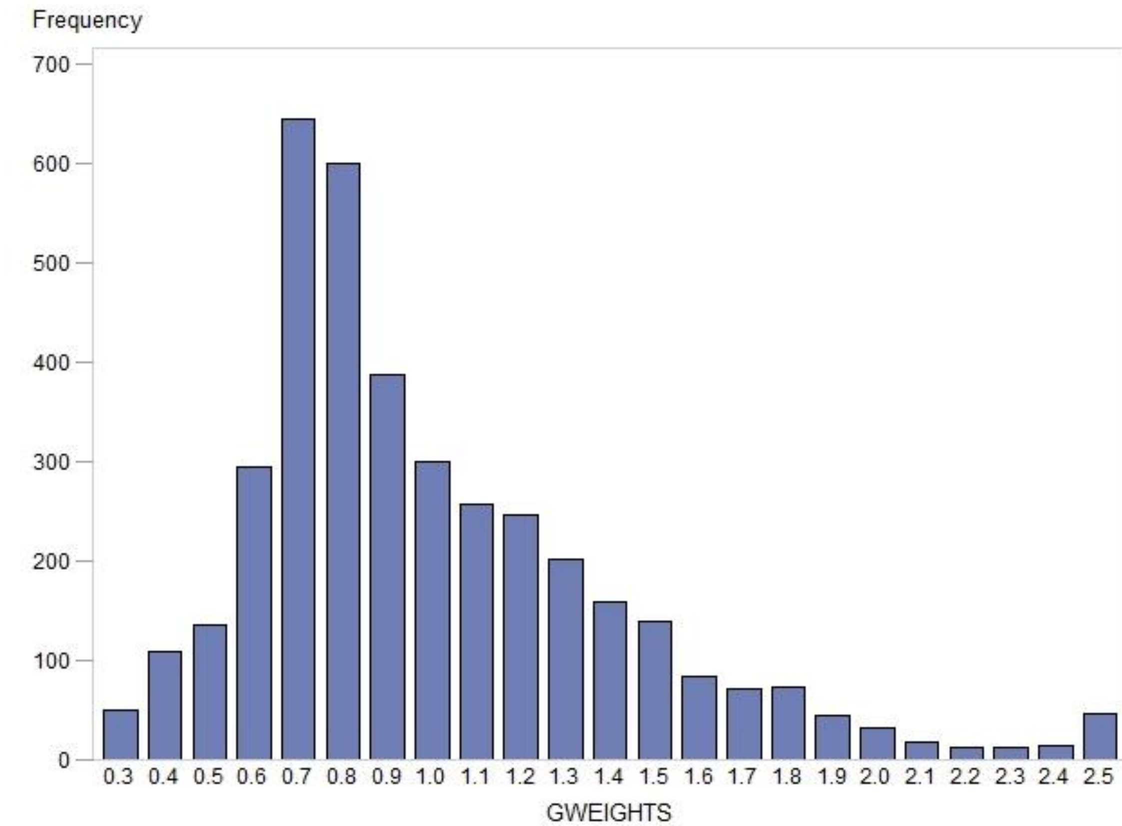


Figure 10. Distribution of GWEIGHTS – K cohort cross sectional sample weight

Table 18. Analysis of variable GWEIGHTS – K cohort cross sectional sample weight

Analysis Variable: GWEIGHTS							
Mean	**Standard Deviation**	**Minimum**	**Maximum**	**Modal Range**	**Range**	**Sum**	**Count**
1.00	0.4160	0.33	2.50	0.86-0.89	2.17	3956	3956

Table 19. Observed weighting properties – K cohort cross sectional sample weight

WEIGHTING PROPERTY	OBSERVED?
Right skewness in the distribution of weights	YES
Relatively small number of weights hitting the upper threshold of 2.50	YES
Modal weight less than one (by definition if the mean is one)	YES
Distribution can be regarded as approximately smooth, with an absence of volatile spikes	YES
Generally comparable with weights in previous waves with respect to shape/distribution, modal weight and proportion at upper threshold	YES

Weight capping

Weight capping was seen as important in seeking to retain essential properties that exist in previous wave releases. In order to achieve these caps, a small amount of collapsing of post-stratum was required.

Note that longitudinal surveys fundamentally will encounter the situation mentioned above after many waves. Sample attrition sees the fully responding sample decrease over time, assuming the absence of a top-up mechanism. Iterative model-driven calibration algorithms rely on sufficient post-stratum sample being available to ensure weight caps are applied correctly subject to benchmark constraints. It is clear that the responding sample in the Wave 5 cohort has reached the stage whereby convergence isn't able to be achieved at the finer level. It is common for Northern Territory to have small cell counts and receive high weights compared to other states due to a relatively small population in non-remote areas, and difficulties obtaining responses.

The table below presents counts of capped weights for Wave 5 i.e. all sample weights equal to 2.50.

Table 20. Counts of capped weights for Wave 5

WAVE 5 WEIGHTS				
	B COHORT		K COHORT	
STATE	LONGITUDINAL	CROSS-SECTIONAL	LONGITUDINAL	CROSS-SECTIONAL
AUS	149	66	92	40
NSW	57	25	45	17
VIC	33	15	25	17
QLD	22	12	12	4
SA	12	7	6	1
WA	14	5	2	1
TAS	6	1	2	0
NT	0	0	0	0
ACT	5	1	0	0

Macro checks

Macro checks assess the data at an aggregate level, seeking to identify issues that arise once records are pooled together into meaningful summary measures ie. estimates, variances, counts. Often these checks are quite intuitive as then can more directly affect statistical analysis and dissemination procedures.

Sums and checking alignment

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

Table 21. Sample counts for the B cohort

Wave	1	2	3	4	5
Cross Sectional Response	5107	4606	4386	4242	4085
Longitudinal Response	-	4606	4253	3997	3758
Cross Sectional Attrition Rate (%)	-	9.81%	14.12%	16.94%	20.01%
Longitudinal Attrition Rate (%)	-	9.81%	7.66%	6.02%	5.98%

Table 22. Sample counts for the K cohort

Wave	1	2	3	4	5
Cross Sectional Response	4983	4464	4331	4169	3956
Longitudinal Response	-	4464	4196	3940	3682
Cross Sectional Attrition Rate (%)	-	10.42%	13.08%	16.34%	20.61%
Longitudinal Attrition Rate (%)	-	10.42%	6.00%	6.10%	6.55%

- *Cross Sectional Response* – counts of those who responded to at least the current wave in question.
- *Longitudinal Response* – counts of those who responded to the current wave in question PLUS all previous waves ie. complete trajectory information for these children since the time of selection.
- *Cross Sectional Attrition Rate (%)* – those not responding to the current wave in question 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.

The following checks were conducted to ensure weights were of an appropriate quality for use by LSAC stakeholders.

Table 23. Status of weighting quality

QUALITY CHECK	STATUS
Cohort B and K longitudinal weights are allocated to all longitudinal respondents for Wave 5.	PASSED
Cohort B and K longitudinal sample weights average to one.	PASSED
Cohort B and K longitudinal population weights sum to 243026 and 253202 respectively, the Australia-level benchmark totals at the time of selection, or alternatively, scaled by the sample fraction.	PASSED
Cohort B and K cross sectional weights are allocated to all cross sectional respondents for Wave 5.	PASSED
Cohort B and K cross sectional sample weights average to one.	PASSED
Cohort B and K cross sectional population weights sum to 243026 and 253202 respectively, the Australia-level benchmark totals at the time of selection, or alternatively, scaled by the sample fraction.	PASSED

Reacquisition of sample from previous waves

In this context, the reacquisition of sample refers to gaining a full response from a participant who had been lost in a previous wave. Consider the following acquisition figures for Wave 5.

For the B cohort, out of **865** that didn't respond to Wave 4, **129** responded to Wave 5. Out of the **1110** that didn't respond to at least one of waves 2, 3 or 4, **327** responded to Wave 5

For the K cohort, out of **814** that didn't respond to Wave 4, **94** responded to Wave 5. Out of the **1043** that didn't respond to at least one of waves 2, 3 or 4, **274** responded to Wave 5

The tale below shows those whom have historically been reacquired having not fully responded to the previous wave of questioning.

Table 24. Sample Re-acquisition 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
B	133	135	129
K	135	119	94

Test estimates of variance

For reference, test estimates of variance (standard error or relative standard error as appropriate) are reproduced here for comparative purposes. These were compared to variances from past waves for validation purposes during the production of weights.

Table 25. Parent one employment status (Australia level) – Standard error estimates

	B COHORT		K COHORT	
EMPLOYMENT STATUS	LONG	CROSS	LONG	CROSS
EMPLOYED	0.91%	0.96%	0.85%	0.88%
UNEMPLOYED	0.32%	0.33%	0.31%	0.33%
NILF	0.89%	0.88%	0.84%	0.82%

Table 26. Parent one usual income (State level) – Relative standard error

	B COHORT		K COHORT	
STATE	LONG	CROSS	LONG	CROSS
NSW	5.7%	4.7%	4.9%	5.0%
VIC	5.3%	5.4%	4.6%	4.3%
QLD	6.6%	7.7%	5.1%	4.5%
SA	8.9%	11.4%	8.0%	8.0%
WA	9.6%	8.5%	7.9%	8.3%
TAS	19.2%	14.9%	12.0%	11.6%
NT	18.3%	17.0%	11.3%	11.0%
ACT	15.3%	12.5%	8.7%	8.4%

Certify calibration diagnostics

Rather than step the reader through the details of calibration checks, this section will simply highlight issues considered, and indicate whether the calibration satisfied that consideration. Details of the checks are housed in the appendix for reference. Anyone seeking to run their own checks are encouraged to obtain a copy of the GregWt macro from the ABS, as it presents highly intuitive diagnostics.

Table 27. Observed calibration quality checks

CALIBRATION QUALITY CHECK	OBSERVED?
All records are read into the algorithm	YES
Ensure all weights satisfy capping requirements	YES
Post-stratum totals match with benchmark counts	YES
All replicate groups converge (for stability)	YES

Bibliography

Little, R.J.A. Rubin, DB. (1987). *Statistical analysis with missing data*. Vol. 539. New York: Wiley.

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix A: Logistic Regression Type 3 Analysis of Effects

Table A1: B Cohort – Longitudinal Weights

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
DF03DP1	1	44.20	<0.0001
DCNFSER	1	3.73	0.0534
DFD08M1	5	15.61	0.0080
DP2SCD	1	109.77	<0.0001
DFD11M2	4	19.33	0.0007

Table A2: B Cohort – Cross Sectional Weights

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
AF03M3	1	51.26	<0.0001
AF03M2	1	39.74	<0.0001
AFD08M1	5	61.08	<0.0001
AP1SCD	1	153.84	<0.0001
AFD11M2	4	62.67	<0.0001

Table A3: K Cohort – Longitudinal Weights

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
FF03FP1	1	28.85	<0.0001
FCNFSER	1	4.45	0.0326
FFD08M1	4	28.3	<0.0001
FFD11M2	4	20.40	0.0004
FP2SCD	4	98.27	<0.0001

Table A4: K Cohort – Cross Sectional Weights

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
CF03M2	1	68.51	<0.0001
CFD08M1	5	100.04	<0.0001
CFD11M2	4	64.09	<0.0001
CP1SCD	1	183.64	<0.0001

Appendix B: Description of Wave 5 weights

Table B1: Description of Wave 5 weights

Variable name	**Cohort**	**Type**	**Waves cases responded to**	**Used for**
eweight	B	Population	1 & 5	Wave 5 cross-sectional analyses Wave 1 & 5 longitudinal analyses
eweights	B	Sample	1 & 5	Wave 5 cross-sectional analyses Wave 1 & 5 longitudinal analyses
bcdewt	B	Population	1, 2, 3, 4 & 5	Wave 1, 2, 3, 4 & 5 longitudinal analyses
bcdewts	B	Sample	1, 2, 3, 4 & 5	Wave 1, 2, 3, 4 & 5 longitudinal analyses
gweight	K	Population	1 & 5	Wave 5 cross-sectional analyses Wave 1 & 5 longitudinal analyses
gweights	K	Sample	1 & 5	Wave 5 cross-sectional analyses Wave 1 & 5 longitudinal analyses
defgwts	K	Population	1, 2, 3, 4 & 5	Wave 1, 2, 3, 4 & 5 longitudinal analyses
defgwt	K	Sample	1, 2, 3, 4 & 5	Wave 1, 2, 3, 4 & 5 longitudinal analyses

Appendix C: Sums of population weights by state, and alignment to population benchmarks

Table C1. Sums of population weights across waves and cohorts

Cohort	B					K				
Selection State/Wave	1	2	3	4	5	1	2	3	4	5
All	243,026	243,026	243,026	243,026	243,026	253,202	253,202	253,202	253,202	253,202
1	82,232	82,232	82,232	82,232	82,232	86,636	86,636	86,636	86,636	86,636
2	61,707	61,707	61,707	61,707	61,707	61,876	61,876	61,876	61,876	61,876
3	45,072	45,072	45,072	45,072	45,072	48,856	48,856	48,856	48,856	48,856
4	18,122	18,122	18,122	18,122	18,122	18,815	18,815	18,815	18,815	18,815
5	23,316	23,316	23,316	23,316	23,316	24,420	24,420	24,420	24,420	24,420
6	5,892	5,892	5,892	5,892	5,892	6,236	6,236	6,236	6,236	6,236
7	2,483	2,483	2,483	2,483	2,483	2,226	2,226	2,226	2,226	2,226
8	4,202	4,202	4,202	4,202	4,202	4,128	4,128	4,128	4,128	4,128

Appendix D: Schematic of SAS Weighting Process in Australian Bureau of Statistics corporate environment

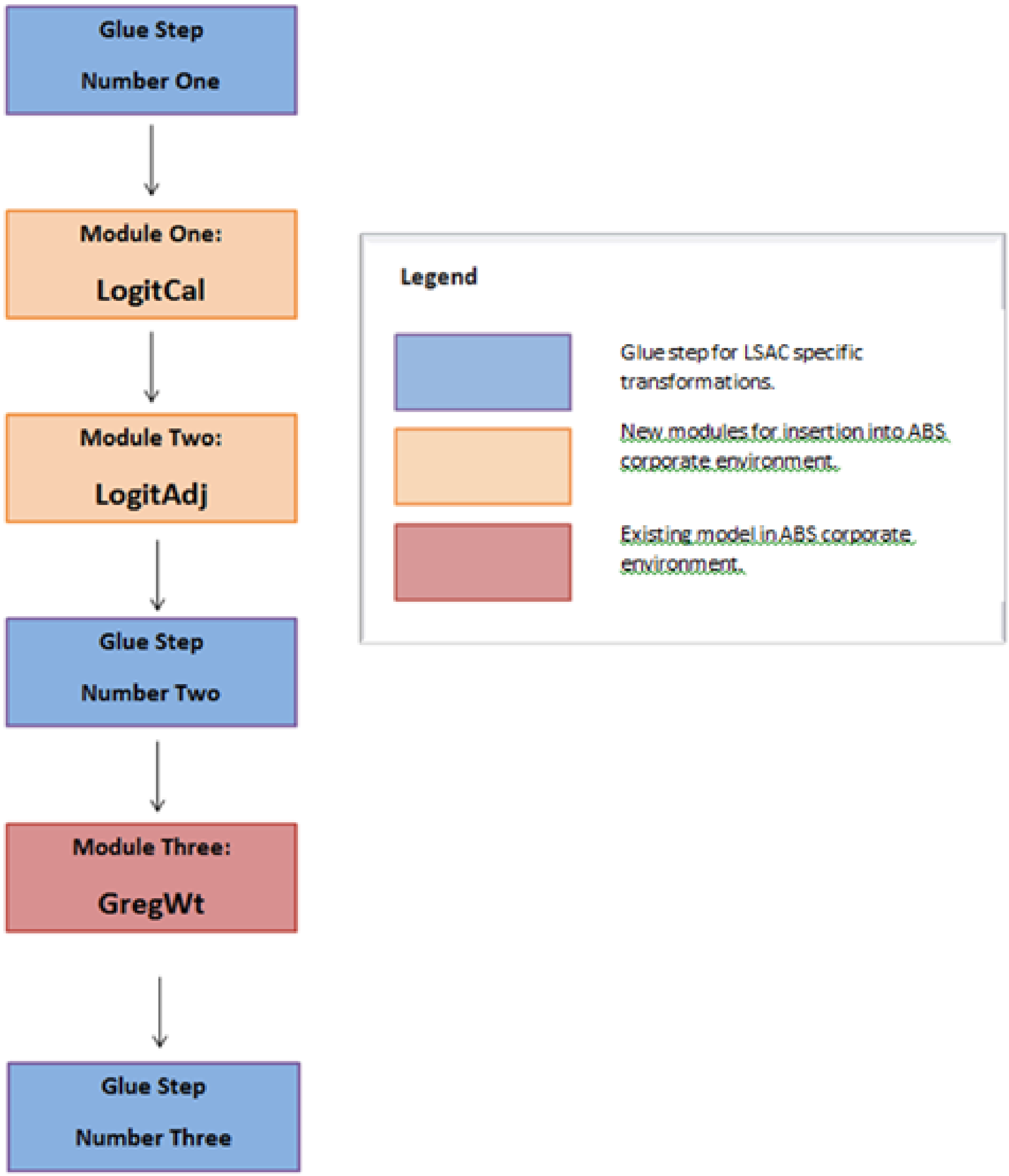


Figure C1. Schematic of SAS Weighting Process

Modularised weighting process

This section details the specific modules, including their inputs, outputs and general functionality. Note that intermediary glue steps exist within the process in order to conduct LSAC-specific transformation of datasets. These are included to coherently map the entire industrialisation.

a) Glue Step One: Survey data file(s) preparation

Inputs - survey data file (sdf) of current and previous waves

Outputs - clean model covariates with no missing values, response indicator variable

Method - simple mean impute for continuous covariates and mode impute for categorical covariates, merge two sdf's and check whether a household satisfied the given response criterion and flag accordingly

b) Module One: LogitCal

Inputs - merged sdf's with response indicator and cleaned covariates, previous wave's sample weight

Outputs - survey data file with estimated response probability

Method - model (and predict) response probabilities using logistic regression with pre-defined numeric and categoric covariates

c) Module Two: LogitAdj

Inputs - data file with unadjusted (previous wave's) weights and estimated response probabilities

Outputs - non response adjusted weights merged onto survey data file for current wave

Method - divide unadjusted weights by the estimated response probability for those satisfying the response condition

d) Glue Step Two: Benchmark file preparation

Inputs - survey data file of current and previous waves

Outputs - GregWt compatible survey data file for current wave, benchmark file

Method - sums reference wave's weights to generate benchmark file, identifies replicate groups using standard calibration glue (upper caps on these weights are available)

e) Module Three: GregWt (existing module in ABS corporate environment)

Inputs - survey data file (with non-response adjusted weight and replicate group identifier), benchmark totals, desired weight caps

Outputs - calibrated weight, replicate weights

Method - generalised regression estimation adjusted the inputs weights so that they sum to pre-determined benchmark totals

f) Glue Step Three: Derivation of population weights

Inputs - non-response adjusted/calibrated sample weights for current wave, population weights for previous wave

Outputs - population weights for the current wave

Method - scales all sample weights up by the same factor to form population weights

Appendix E: Variables used in weighting

The weighting is split into two methods, categorised by the following steps:

1. Logistic regression non-response adjustment of weights.
2. Calibration, using generalised regression, of weights to known post-stratum totals.

Logistic regression step

- family member's age last birthday
- SEIFA economic resources
- highest level of school completed by mother
- P2 in the home
- P1 first spoke a language other than English

Calibration step

- State (of selection)
- Metropolitan/Ex-metropolitan split (of selection)
- Post code (of selection)

Appendix F: The ABS GREGWT Macro

The calibration method applied makes use of the ABS standard GregWt SAS macro. This program achieves calibration through an iterative approach whereby weights are adjusted to ensure benchmark totals are aligned to, subject to a distance function. The algorithm will seek to minimise the ‘change’ in the pre-calibrated to post-calibrated weights, as defined by the distance function, all while meeting benchmark totals, and incorporating weight caps (see section 5.5). In this case, a relative quadratic distance function was applied. See below.

$$\sum_{i \epsilon s} \frac{(w_i g_i - w_i)^2}{w_i}$$

Where, for the case of LSAC, w_i is the pre-calibrated weight that exists after applying the logistic regression non-response adjustment, and g_i is the grossing factor that translates the w_i to calibrated weights.

$$\sum_{i \epsilon s} g_i w_i x_{ij} = \sum_{i \epsilon U} x_{ij} = X_{jT}$$

Full details for the GREGWT macro can be found in (Bell, 2000)

Appendix G: Non-response to Instruments

Table G1. Non-response to instruments

	Eligible	Responding	%Wave 1	Response rate %
			B cohort	
Wave 5 (issued sample = 4658)				
Interview	4085	4085	80.0	100.0
P1CASI	4077	4010	78.5	98.4
P2SC	3512	2444	na	69.6
PLECATI	537	404	na	75.2
TEACH	4021	3490	na	86.8
CSRB	4026	4014	78.6	99.7
Wave 4 (issued sample = 4929)				
Interview	4242	4242	83.1	100.0
P1CASI	4242	4240	83.0	100.0
P2SC	3706	2677	na	72.2
PLECATI	573	378	na	66.0
TEACH	4225	3427	na	81.1
CSRB	4242	4181	81.9	98.6
			K cohort	
Wave 5 (issued sample = 4551)				
Interview	3956	3956	77.5	100.0
P1CASI	3952	3857	77.4	97.6
P2SC	3277	2333	na	71.2
PLECATI	614	464	na	75.6
TEACH	3857	3225	na	83.6
ACASI	3873	3844	77.1	99.3
CSRK	3872	3850	77.3	99.4
TUD	3871	3649	73.2	94.3
Wave 4 (issued sample = 4774)				
Interview	4169	4169	83.7	100.0
P1CASI	4169	4117	82.6	98.8
P2SC	3512	2645	na	75.3
PLECATI	734	493	na	67.2
TEACH	4144	3352	na	80.9
ACASI	4169	4099	82.3	98.3
TUD	4169	3994	80.2	95.8

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
na	Not appropriate to compare with Wave 1;

Parent 1 CASI

Of the families interviewed in Wave 5, 2% of Parent 1's did not complete the P1 CASI.

Parent 2 self-complete forms

The response rate for Wave 5 Parent 2's was around 70%. This was 3 - 4% lower than the response rate for Parent 2's in Wave 4.

Parent Living Elsewhere (PLE) Instrument

Of the eligible PLE's that interviewers attempted to contact 75% responded.

Teacher self-complete form

The teacher forms continue to achieve good response rates (over 80%). When compared to Wave 4 rates there was an increase in response rate for the teacher forms across both cohorts; over 5% in the B cohort and 2% in the K cohort. In Wave 5 teacher forms for the B cohort were again sent to the study child's main classroom teacher. However due to majority of the K cohort children attending high school in Wave 5, 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 TUD remains high at 94%. This represented a drop in the TUD response rate of almost 2% when compared with Wave 4.

Instrument response rate by characteristics of families

Based on Wave 1 characteristics, the response rates to the instruments in Wave 5 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 8 - 34% 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 7% lower than the full sample. Where Parent 1 spoke a language other than English at home, Parent 2 and the PLE had response rates 8 - 10% lower than the full sample.
- When Parent 1 had an income of at least $1000pw, Parent 2 was 8% more likely and the PLE was 18% more likely to take part in an interview then when the Parent 1 had an income below $1000pw.
- Similarly, where Parent 1 was employed Parent 2 was 8% more likely and the PLE was 12% more likely to take part in an interview compared to where Parent 1 was not employed.
- The Northern Territory had the highest response rate to the Parent 2 form (82%); the lowest was in New South Wales (67%).
- The highest response rate to the teacher questionnaire was in Queensland (89%); teachers in Tasmania had the lowest response rate (83%).

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 9 - 23% lower than the non- Aboriginal and Torres Strait Islander sample.
- Indigenous children also had a lower response rate to the TUD (82%) when compared to the non-Indigenous sample (94%).
- 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 9% lower than the full sample. Where Parent 1 spoke a language other than English at home, Parent 2 and the PLE had response rates 10% lower than families where Parent 1 spoke only English.
- When Parent 1 had an income of at least $1000pw, Parent 2 and the PLE were 11-12% more likely to take part in an interview then when the Parent 1 had an income below $1000pw.
- Where Parent 1 was employed Parent 2 and the PLE were 7 - 8% more likely to respond compared to when Parent 1 was not employed.

- Where families did not live in a capital city, Parent 2's were 13% less likely to return a questionnaire.
- Western Australia had the highest response rate to the P2 form (75%); Tasmania had the lowest (66%).
- The highest response rate to the teacher questionnaire was from Tasmania (87%); the lowest was from the Northern Territory (74%).
- Study children from the ACT had the highest response rate to the TUD (98%), while those from Tasmania had the lowest (85%).

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

Table H1: B cohort non response to forms

Response rate % (N)	F2F	P1CASI	P2SC	PLE CATI	TEACH	CSRB
Full sample	**87.4 (4658)**	**98.1 (4070)**	**69.6 (3537)**	**60.5 (669)**	**86.2 (4055)**	**99.4 (4032)**
Study child Indigenous	74.7 (182)	92.6 (136)	45.8 (96)	28.9 (45)	77.9 (136)	98.5 (135)
Study child non-Indigenous	87.9 (4476)	98.3 (3934)	69.9 (3441)	62.9 (623)	86.4 (3919)	99.4 (3897)
Parent 1 LOTE spoken	80.5 (622)	93.8 (501)	59.3 (455)	52.1 (48)	82.6 (500)	99.2 (497)
Parent 1 English only	88.4 (4036)	98.7 (3569)	70.7 (3082)	61.3 (620)	86.7 (3555)	99.3 (3535)
Parent 1 Employed	90.5 (2378)	98.9 (2151)	72.7 (1907)	67.2 (305)	86.6 (2141)	99.2 (2132)
Parent 1 Not Employed	84.1 (2272)	97.2 (1911)	65.2 (1622)	55.0 (362)	85.8 (1906)	99.4 (1892)
Parental income <$1000	83.9 (2203)	97.6 (1848)	64.8 (1515)	56.5 (402)	85.2 (1840)	99.3 (1830)
Parental income >=$1000	90.9 (2197)	98.8 (1997)	73.1 (1810)	74.6 (224)	87.1 (1991)	99.3 (1982)
NSW	87.8 (1464)	98.4 (1285)	66.5 (1129)	59.4 (202)	84.1 (1279)	98.8 (1279)
VIC	84.3 (1135)	96.4 (957)	71.3 (856)	63.0 (138)	87.0 (957)	99.5 (949)
QLD	88.5 (948)	99.2 (839)	68.5 (712)	59.1 (159)	89.1 (834)	99.5 (831)
SA	87.7 (349)	98.0 (306)	69.6 (250)	71.9 (57)	85.8 (303)	99.3 (300)
WA	87.2 (478)	98.8 (417)	71.8 (369)	57.4 (61)	85.6 (416)	99.8 (408)
TAS	96.3 (107)	97.1 (103)	68.5 (89)	58.8 (17)	82.5 (103)	100.0 (102)

Response rate % (N)	F2F	P1CASI	P2SC	PLE CATI	TEACH	CSRB
Full sample	**87.4 (4658)**	**98.1 (4070)**	**69.6 (3537)**	**60.5 (669)**	**86.2 (4055)**	**99.4 (4032)**
ACT	92.2 (103)	85.3 (95)	74.1 (81)	35.7 (14)	88.4 (95)	100.0 (95)
NT	91.9 (74)	100.0 (68)	82.0 (50)	65.0 (20)	88.1 (67)	100.0 (68)
Capital City	87.0 (2922)	98.0 (2542)	70.1 (2223)	62.2 (389)	86.7 (2533)	99.2 (2517)
Rest Of State	88.0 (1736)	98.3 (1528)	67.8 (1314)	58.4 (279)	85.4 (1522)	99.5 (1515)
Study child male	87.6 (2389)	98.3 (2092)	70.0 (1806)	60.6 (353)	84.6 (2083)	99.2 (2070)
Study child female	87.2 (2269)	97.9 (1978)	68.5 (1731)	60.6 (315)	87.9 (1972)	99.5 (1962)

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

Table I 1: K cohort non response to forms

Response rate % (N)	F2F	P1 CASI	P2SC	PLE CATI	TEACH	ACASI	CSRK	TUD
Full sample	**86.4 (4551)**	**97.8 (3933)**	**71.2 (3302)**	**61.0 (761)**	**83.1 (3900)**	**98.4 (3902)**	**98.5 (3902)**	**93.3 (3902)**
Study child Indigenous	72.5 (153)	89.2 (111)	52.4 (82)	38.9 (36)	74.8 (111)	97.2 (106)	99.1 (106)	82.1 (106)
Study child non-Indigenous	86.9 (4396)	98.0 (3820)	71.4 (3219)	62.2 (724)	83.4 (3787)	98.3 (3794)	98.4 (3794)	93.6 (3794)
Parent 1 LOTE spoken	77.5 (659)	93.9 (511)	62.6 (449)	51.7 (58)	78.9 (507)	98.4 (507)	98.6 (507)	92.5 (507)
Parent 1 English only	87.9 (3892)	98.3 (3422)	72.3 (2853)	61.7 (703)	83.8 (3393)	98.2 (3395)	98.4 (3395)	93.3 (3395)
Parent 1 Employed	88.7 (2661)	98.8 (2359)	73.8 (2022)	64.3 (429)	83.8 (2344)	98.7 (2346)	98.7 (2346)	94.8 (2346)
Parent 1 Not Employed	83.4 (1884)	96.2 (1571)	66.4 (1278)	56.4 (330)	82.2 (1553)	97.6 (1553)	98.0 (1553)	90.8 (1553)
Parental income <$1000	81.8 (1896)	96.7 (1551)	63.7 (1160)	56.3 (437)	83.0 (1537)	97.5 (1541)	97.9 (1541)	91.1 (1541)
Parental income >=$1000	90.3 (2330)	98.5 (2105)	75.6 (1899)	67.7 (279)	83.4 (2089)	98.8 (2086)	98.7 (2086)	95.0 (2086)
NSW	85.4 (1434)	98.1 (1224)	71.0 (1026)	62.2 (230)	82.5 (1218)	98.4 (1223)	98.6 (1223)	92.5 (1223)
VIC	84.2 (1127)	95.8 (949)	69.7 (796)	66.5 (167)	86.3 (941)	98.2 (942)	98.1 (942)	93.6 (942)
QLD	87.9 (885)	98.6 (778)	72.0 (661)	59.4 (170)	80.9 (768)	97.8 (769)	98.4 (769)	92.8 (769)
SA	88.1 (327)	99.0 (288)	70.8 (240)	61.7 (60)	83.6 (286)	98.9 (282)	99.3 (282)	96.1 (282)
WA	87.7 (462)	98.8 (405)	75.0 (348)	47.2 (72)	82.6 (402)	98.3 (401)	98.0 (401)	95.0 (401)

Response rate % (N)	**F2F**	**P1 CASI**	**P2SC**	**PLE CATI**	**TEACH**	**ACASI**	**CSRK**	**TUD**
Full sample	**86.4 (4551)**	**97.8 (3933)**	**71.2 (3302)**	**61.0 (761)**	**83.1 (3900)**	**98.4 (3902)**	**98.5 (3902)**	**93.3 (3902)**
TAS	92.2 (129)	96.6 (119)	65.6 (93)	59.3 (27)	87.1 (116)	95.8 (118)	96.6 (118)	84.7 (118)
ACT	92.7 (109)	100.0 (101)	67.1 (85)	55.6 (18)	81.2 (101)	100.0 (101)	100.0 (101)	98.0 (101)
NT	88.5 (78)	97.1 (69)	66.0 (53)	70.6 (17)	73.5 (68)	100.0 (66)	98.5 (66)	90.9 (66)
Capital City	85.6 (2822)	97.6 (2416)	75.8 (2063)	58.3 (412)	82.6 (2386)	98.6 (2402)	98.6 (2402)	94.3 (2402)
Rest Of State	87.7 (1723)	98.0 (1517)	62.9 (1239)	64.2 (349)	84.0 (1514)	97.7 (1500)	98.1 (1500)	91.6 (1500)
Study child male	86.3 (2326)	98.0 (2008)	70.7 (1691)	59.6 (391)	82.5 (1990)	97.7 (1994)	98.0 (1994)	92.0 (1994)
Study child female	86.5 (2225)	97.6 (1925)	71.2 (1611)	62.4 (370)	83.8 (1910)	98.8 (1908)	98.9 (1908)	94.5 (1908)