Growing Up in Australia

The Longitudinal Study of Australian Children:

an Australian Government initiative

LSAC Technical paper No. 3

Wave 1 weighting and non-response

Carol Soloff, David Lawrence, Sebastian Misson and Robert Johnstone

May 2006

Australian Government

Australian Institute of
Family Studies

Contents

About the authors

Carol Soloff is the Project Manager for *Growing Up in Australia*, and previously worked for the Australian Bureau of Statistics. Carol has extensive experience across all aspects of the development, conduct and processing of household surveys.

Dr David Lawrence is Senior Statistician in the Centre for Developmental Health at Curtin University of Technology. David previously worked as a Survey Methodologist at the Australian Bureau of Statistics, and is currently involved in the Western Australian Aboriginal Child Health Survey.

Sebastian Misson is the Data Administrator for *Growing Up in Australia*. Sebastian has had prime responsibility for preparing the Wave 1 dataset for release. Sebastian has extensive experience with large-scale quantitative research at both the Australian Council for Education Research and the Australian Research Centre for Sex, Health and Society.

Robert Johnstone was data manager for the first wave of Growing Up in Australia. Robert has had extensive experience in data management, survey design and survey processing and has worked on a variety of projects at Deakin University, the Cancer Council of Victoria, the Epworth Hospital, and the Australian Bureau of Statistics. He is currently working on an evaluation of the Commonwealth Department of Family and Community Services' Stronger Families and Communities Strategy conducted by the Australian Institute of Family Studies and the University of New South Wales.

Acknowledgements

LSAC was initiated and is funded by the Australian Government Department of Family and Community Services.

Glossary

ABS - Australian Bureau of Statistics

ERP - Estimated Resident Population

HIC - Health Insurance Commission

LSAC - Longitudinal Study of Australian Children

Met/Exmet - Capital city statistical division/rest of state areas

TUD - Time-use diary

Introduction

This paper details the methodology used to calculate the weights for the Wave 1 sample of *Growing Up in Australia*, the Longitudinal Study of Australian Children (also known as LSAC). This study is funded by the Department of Family and Community Services as part of the Australian Government's *Stronger Families and Communities Strategy*, and is Australia's first national longitudinal study of children.

Growing Up in Australia is a broad, multi-disciplinary study that has been developed to examine the impact of Australia's unique social, economic and cultural environment on the next generation, particularly in regard to issues of policy relevance.

During 2004, the study recruited a nationally representative sample of 5,107 infants and 4,983 children aged 4-5 years selected from the Medicare enrolments database.

A two-stage clustered design was employed, first selecting postcodes then children, allowing analysis of children within communities and reducing the overall cost of the study. Children in both cohorts were selected from the same postcodes, with about 40 children per postcode usually invited to participate in the study in the larger states, and usually about 20 children per postcode in the smaller states and territories.

Stratification was used to ensure proportional geographic representation for states/territories and capital city statistical division ('met') /rest of state ('exmet') areas. The method of postcode selection accounted for the number of children in the postcode so all potential participants in the study Australia-wide had an approximately equal chance of selection (about one in 25). However, some remote postcodes were excluded from the design, and the population estimates have been adjusted accordingly.

The selection of children and corresponding fieldwork occurred in 4 phases. This was done to enable sample selection of children born across all months of the calendar year, to attempt to reduce the age range of children at interview, and also because some of the target population had not been born at the time of the first phase selection.

This paper should be read in conjunction with LSAC Technical Paper No. 1 "Sample Design" (Soloff et al, 2005) which outlines full details of the sample design.

The initial (or design) weights were derived from the probability of selecting each child in the sample. These design weights were then adjusted according to information collected about the child's family compared with characteristics for similar families at the time of the 2001 ABS Census of Population and Housing, and then further adjusted so that weighted estimates match population benchmark data provided by the Australian Bureau of Statistics (ABS).

Weighting principles

The weighting methodology was developed by Dr David Lawrence, in conjunction with the staff from the Australian Institute of Family Studies. The following broad principles guide the weighting for Wave 1 of the study:

- The main purpose of weights in this study will be to compensate for differences between the final sample and the national population;
- The weights should be considered as expansion factors permitting the scaling of the sample to the population. Hence the sum of the weights should accurately match appropriate population benchmarks;
- The weights will reflect both the design of the study (to allow for unequal probabilities of inclusion in the study that may result in sampling biases) and likelihood of response (those less likely to respond are given a higher weight and those more likely to respond are given a lower weight);
- To allow for any non-response effects, a post-stratification weighting system using appropriate population benchmarks should be used;
- The design of weighting should aim to give a dataset with broad application, while at the same time accepting that some variables will need to be treated as more important than others; and
- Care should be taken to match adjustment models and external constraints imposed on survey estimates to produce more accurate results. Excessively low or high weights, or marked clustering of weights about imposed bounds indicate that the weighting is being overstretched

Before describing the weighting process in detail, it is worth clarifying the scope and coverage of the study: the scope of the study is the population that the study purports to represent, and the coverage of the study is the population from which the sample was selected.

The study population

It was intended that the sample be nationally representative of two cohorts - children under 12 months and children aged 4 years. In practice, this translated to children born between March 2003 and February 2004, and between March 1999 and February 2000. Full details on the reasons for this decision are given in the technical paper on sample design (Soloff et al, 2005).

The scope is therefore all children born between the given months, and who are Australian citizens, permanent residents or applicants for permanent residence, with the exception of children living in some remote parts of Northern Territory, Western Australia, South Australia, Queensland and New South Wales.

The coverage is the children who were registered with Medicare at the time of the sample selection, excluding children in some postcodes where very few children lived. The main source of under-coverage is children born in January and February 2004, as many of these children had not been registered with Medicare at the time of the sample selection.

As discussed in this paper, the child weights have been adjusted so that the sum of the weights matches the population in scope. The in-scope population for benchmarking purposes was obtained from the ABS. State population estimates for male and female children aged 0 and 4 year at end March 2004 were used, with the distribution between capital city statistical division ("met")/rest of state ("exmet") based on the June 2003 ABS population estimates. HIC data was used to determine the number of children in remote areas that were excluded from the target population. HIC data were not used for the population benchmarks due to the undercount of the infants and an overcount of the 4-5 year old children, compared with ABS data.

It is important to note that the population estimates are of children, not of parents or families. Therefore in quoting the results from this survey, care should be taken to ensure this point is understood. Using the estimates to count families/parents will produce an over-count of the number of families/parents, due to the double/triple/etc counting of children from multiple births. Although this will make a relatively small difference to the actual numbers, it may be important in the interpretation of the information and in comparing data from other sources. Although it is possible to produce ‘family’ weights, this has not been attempted.

Overview of the weighting process

An initial sample of 9,259 children aged 0-1 years (infant cohort) and 10,275 children aged 4-5 years was selected by the Health Insurance Commission ((HIC) from its Medicare enrolments database. Ultimately 5,107 infants and 4,983 4-5 year old children were recruited to the study.

Once the initial sample has been selected, these children were matched against the 'fact-of-death' file and any child with a similar name to a child that had died was removed from the sample. Families of the remaining children were then sent a letter by HIC, inviting them to take part in the study, or to opt-out by either phoning a 1800 number or returning a reply paid form. HIC then removed any opt-outs or 'return-to-sender's from the sample, before passing the contact details to I-view, the wave 1 data collection agency. I-view then sent a letter to all these families, indicating that an interviewer would be visiting, and given another 1800 number to call for appointments or to opt-out of the study.

There are a number of places where sample loss occurred. The main sources of sample loss were:

- Name match with HIC 'fact-of-death' file (3.3% for infant and 3.4% for 4-5 year old cohort);
- Refusals – to HIC, I-view 1800 number, and interviewers (31.2% for infant and 35.0% for 4-5 year old cohort); and
- Non-contacts - mainly PO boxes and families who had moved (10.4% for infant and 14.2% for 4-5 year old cohort);

Analysis of the non-respondents was undertaken by comparing the characteristics of the sample children and their families with the characteristics of children of a corresponding age at the time of the 2001 ABS Census of Population and Housing, at the postcode level. The HIC database was not used as a comparison point since the ABS produces more accurate population estimates with better sociodemographic detail. Initially chi-squared analysis at the Australian level was used to identify possible variables that may be related to non-response. Non-response was most related to low level of school completion of mother and father, mother or father speaking a language other than English at home, the study child being identified as indigenous and single-parent (as opposed to dual parent) families.

Poisson regression was then undertaken to determine which of these variables were independently related to non-response. The results of this analysis showed 2 variables made significant independent contributions to non-response: mother who had not completed year 12 at school and mothers whose first language was not English.

Benchmark data for the production of population estimates was obtained from ABS for cohort by sex and state as at March 2004. ABS June 2003 population data were used to allocated the population to capital city ("met") and rest of the state ("exmet") areas, and HIC data from March 2005 were used to adjust the figures for children in the excluded remote areas.

Although the design for the sample was developed on the premise that all children should have an equal chance of selection, in practice this was not possible. Therefore a design weight was calculated for each child selected in the survey, the inverse of the probability of selection for each child.

Final weights were produced by adjusting the design weights to compensate for non-response. This was achieved using the technique of calibration on known marginal totals (Deville and Särndal, 1992). Effectively, respondents from categories with lowest participation rates are given higher weights in a procedure that ensures that weighted estimates from each cohort achieve the population benchmarks for each variable included in the process. Variables used were those identified by the non-response analysis: mother's level of schooling and whether the mother spoke a language other than English at home.

Wave 1 response

The following table details the sources of sample loss for each of the cohorts. The 4-5 year old cohort recruitment rate was lower, due to the higher proportion of out-of-date addresses and refusals compared with infants. There is no apparent difference in the reasons for refusals, but the higher numbers of refusals may be due to families with 4-5 year olds being, on average, busier, with more mothers working. In addition, the interview was a longer and more complex process for families with 4-5 year olds than infants.

Table 1 shows the distribution of the Wave 1 sample by state and part of state.

Table 1 Final Wave 1 sample by stratum

	Infant cohort			4-5 year old cohort			Total
State	Met	Exmet	Total	Met	Exmet	Total	
NSW	984	633	1,617	948	621	1,569	3,186
VIC	898	352	1,250	888	355	1,243	2,493
QLD	483	574	1,057	428	561	989	2,046
SA	255	90	345	256	84	340	685
WA	368	165	533	363	148	511	1,044
TAS	53	60	113	54	83	137	250
NT	48	38	86	44	38	82	168
ACT	106	0	106	112	0	112	218
Australia	3,195	1,912	5,107	3,093	1,890	4,983	10,090

Tables 2a and 2b indicate the distribution of the sample compared with what the sample distribution would be if it was perfectly representative of the target population (based on the adjusted May 2004 ABS population estimates).

As is common in national surveys, the rate of recruitment was lower for Sydney than other areas of Australia. However, considerably lower rates than expected were also found in exmet South Australia for both cohorts and exmet Tasmania, just for the infant cohort. These may be a function of ‘luck of the draw’ with the areas selected, however there may also be an interviewer effect.

Differences between the actual and the expected distribution were adjusted for in the weighting to population benchmarks.

Table 2a Final Wave 1 sample distribution compared with intended distribution (Infant cohort)

	MET			EXMET		
	Achieved	Target	Ratio*	Achieved	Target	Ratio*
NSW	984	1,096	0.90	633	629	1.01
VIC	898	860	1.04	352	327	1.08
QLD	483	442	1.09	574	534	1.07
SA	255	249	1.02	90	100	0.90
WA	368	341	1.08	165	150	1.10
TAS	53	52	1.02	60	71	0.85
NT	48	36	1.33	38	37	1.03
ACT	106	80	1.33			
Total	3,195	3,156	1.01	1,912	1,848	1.03

*Ratio of achieved sample to target sample

Table 2b Final Wave 1 sample distribution compared with intended distribution (4-5 year old cohort)

	MET			EXMET		
	Achieved	Target	Ratio*	Achieved	Target	Ratio*
NSW	948	1051	0.90	621	648	0.96
VIC	888	856	1.04	355	359	0.99
QLD	428	431	0.99	561	537	1.04
SA	256	253	1.01	84	108	0.78
WA	363	343	1.06	148	150	0.99
TAS	54	48	1.13	83	74	1.12
NT	44	31	1.42	38	35	1.09
ACT	112	82	1.37			
Total	3093	3095	1.00	1890	1911	0.99

*Ratio of achieved sample to target sample

Design weights

The first factor to impact upon weights is the initial sample design that determines which children are invited to participate in the study in Wave 1.

In many surveys, sample members have been selected with known but unequal probabilities. In these cases, it will often be desirable to weight observations in order to produce unbiased estimates of population parameters. According to standard practice in such cases, observations are weighted inversely proportional to their probability of selection.

As indicated earlier, the aim of the sample design was to give all children in scope of the study approximately equal probability of selection. In practice, this was not achievable for several reasons:

- The selection of postcodes had to occur months before the actual selection of children and before some of the target population was even born. Thus the measure of size used for postcode selection often differed from the number of children actually in-scope for the study, particularly in postcodes where the birth rate was changing rapidly;
- In addition, because both cohorts were to be selected from the same postcodes, the postcodes were selected according to the total number of infants and 4-5 year olds in the population. Where there were approximately equal numbers of infants and 4-5 year olds in a postcode this had little effect on the weights. However, in postcodes that had significantly different numbers of infants and four year-olds this would have a noticeable impact on the weights; and
- More children were selected in met areas than in exmet areas to adjust for a greater level of non-response;

Additionally there were several exclusions made to reduce respondent burden, in particular not selecting more than one child from any multiple birth, or an infant and a 4-5 year old from the same family.

Therefore, a design weight has been attached to each selected child to account for the actual probability of selection as opposed to the intended equal probability of selection.

Details of the design weight calculations are given in Attachment A. As indicated above, the main source of variability in the design weights stems from the difference between the distribution of infants and 4-5 year olds at the time of the selection of postcodes, and the actual distribution at the time the children were selected, as well as the different distribution of infants and 4-5 year old children across postcodes. An example of this is provided in Attachment A.

Thought was given to whether the considerable range in the design weights reflect real differences between the sample and the population or whether the differences are mostly due to random variation. However, it was considered that it was possible that the features resulting in unequal weights could be related to measured outcomes from the study (for example, children in high growth areas had higher design weights than children in low growth area) and hence it was felt appropriate to factor in the design weight to the weighting calculation.

Non-response analysis

Non-response is a significant issue in this study. The most important part of the weighting strategy is to account for this non-response and any biases that could be introduced because the non-respondents are not a random sub-group of the selected sample.

At various stages in the survey, parents could refuse to participate. If the characteristics of the refusing individuals are the same as those who participate, this is not a significant problem. In practice, however, there will be relationships between the child/parent's characteristics and the probability of refusal. Hence, the weights should attempt to adjust for any such biases.

This is typically done by dividing the sample into response classes and adjusting weights to ensure that each response class has an appropriate final weight. For example, if non-response leads to single mother households being under represented in the sample, then those that are in the sample should be given higher weight.

This process relies on the findings of a non-response analysis. Depending on the results of the non-response analysis there are three possible approaches to the weighting strategy:

- If the non-response is judged to be random, then the design weights can simply be increased by a constant expansion factor to adjust for the proportion of non-response;
- If a small number of factors is found to be linked to participation in the study, the responding children can be post-stratified by these factors, and weights set within each post-stratum; or
- If the number of factors linked to survey participation means that the post-strata would be too small to be able to calculate reasonable, consistent weights not subject to random fluctuations, then the calibration approach of Deville and Särndal (1992) can be employed. In this approach, the design weight is taken as the starting point, and calibrated to give correct population totals for each factor included in the weighting, while minimising the deviation in weights from the original design weights. This strategy has been successfully employed in a number of surveys conducted by major statistical organisations such as the ABS and Statistics Canada.

The results of our analysis have indicated that the calibration approach is the most suitable. This procedure determines the set of weights that will sum to the correct benchmark population totals and minimise the difference between the final survey weights and the initial design weights. Calibration has several advantages over other approaches when it is desirable to account for several variables in non-response adjustment:

- Calibration only requires population benchmarks for each variable separately; there is no need for population benchmarks at the cross-classified level (ie for every combination of variables accounted for). To use post-stratification it is necessary to have one set of population benchmarks cross-classified for all variables being used in the non-response adjustment. These are not always available.

- Even if benchmarks for post-stratification are available, they can be unreliable due to small numbers in individual cells, and may result in instability in the weights.
- Under calibration it is also possible, within bounds, to constrain the overall range of the final weights, which can improve stability in the final estimates.

Approach to non-response analysis

The non-response to the first wave was analysed to identify any patterns that might suggest a non-response bias.

Non-response analysis is difficult because we generally know very little about the non-respondents. In the study some limited information is available about the non-respondents – the stage of the study at which they withdrew, the reason they did not wish to participate in the study, and the postcode in which they lived.

In addition to the limited amount of direct information about the individual non-respondents, it is also possible to infer extra information about them by making comparisons at an aggregate level. As the sample was chosen randomly, the distribution of the sample should only differ from census information about the same children and families by random chance (assuming the families of 0 and 4 year old children in the postcodes in August 2001 have similar characteristics to families with children of about these ages, 3 years later). It has been possible to compare the distribution of the sample with data from the 2001 ABS Census at the national and postcode level.

The analysis took the following steps:

- Collecting information about reasons for non-response, and analysing any geographic trends;
- Bivariate comparisons between respondents and comparable populations from ABS Population Census 2001 data; and
- Poisson regression modelling to predict response at the postcode level using the fewest possible variables.

Results

Reasons for non-response

Wherever possible, I-view interviewers and 1800 staff sought reasons for non-response (see Table 3). There appears to be minor difference between the 2 cohorts in the distribution of reasons. Interviewer debriefing indicated that many refusals were due to people who are very time pressured not being able to find time for one more thing in their lives (e.g. they felt the time the interview would take was time better spent with their children).

Table 3 Reasons for refusals to interviewers

	Infant	4-5 year old	Overall
Not interested / too busy	61.9%	62.7%	62.3%
Privacy / confidentiality	6.7%	4.1%	5.2%
Not capable	0.9%	0.7%	0.8%
Moving house/going overseas	3.2%	3.3%	3.3%
Illness / death	4.8%	4.8%	4.8%
Husband refused	4.8%	5.2%	5.0%
No reason	5.7%	6.4%	6.1%
Other	12.0%	12.9%	12.5%
TOTAL	1046	1363	2409

These reasons are the usual ones that might be expected. Demographic factors that may be related to the reasons for non-response include: family size (too busy), mother working (too busy), language spoken at home, family type (husband refused/ too busy for lone parent), and nature of occupancy (moving).

Geographic trends

Differences in response by state and met/exmet are shown in the previous section. There are clearly differences in response between strata. Since weights were calculated at the stratum level, these differences were automatically accounted for in the weighting procedure.

Bivariate analysis

The first step in considering the impact of non-response on a sample is to determine the size of the group of non-respondents. The second step is to consider how different the non-respondents are from the respondents. This is not possible to do directly as we have minimal, if any, information about the non-respondents. The usual approach, therefore, is to consider how different the respondent characteristics are from accepted population characteristics.

Use was made of the ABS 2001 Census of Population and Housing to compare the LSAC respondents with comparable populations (children aged 0 years and children aged 4 years). Characteristics that were compared were:

- Lone versus couple family – family type;
- Birthplace of parents;
- Language spoken at home;
- Level of school completion of parents;
- Family size;
- Gender of reference child;
- Indigenous status; and
- Nature of occupancy

Chi-squared test were performed to determine whether the differences between the observed proportions and the census proportions were statistically significant, or whether they were the result of random chance. The results from these analyses can be seen in Attachment B.

Poisson regression

A number of variables were identified as being related to non-response in the bivariate analysis. These included family structure (dual parent/sole parent), carers' main language spoken at home, parent's level of schooling, family income and indigenous status. However, these factors may themselves be inter-related. It is possible, for instance, that the association between response and family income may be a by-product of the association between response and family structure, given the strong relationship between family income and family structure.

In order to come up with a weighting strategy that would ameliorate as many of the differences in response as possible while weighting by the fewest possible variables, a Poisson regression was used. This technique allows the modelling of response rate at the postcode level against attributes of the postcode, based on 2001 Census data. All variables found to be strongly or weakly associated with non-response in the bivariate analyses were entered into the models, and the modelling procedure was used to determine the important independent predictors of response.

Analysis was undertaken on a file that contained variables indicating the number of children selected in the postcode, the number of these children that participated in the study and the proportion of 0 or 4 year olds living in each postcode with the following characteristics: mother and father with less than Year 12 at school, family income lower than $500 per week, family income over $1200 per week, mother speaks a language other than English in the home, mother born in Australia, dual parent families, living in a rented home and indigenous status. The results are shown in Attachment C.

Mother speaking a language other than English in the home and mother having completed Year 12 emerged as the major influences on response rate for both cohorts. For this reason it was decided to weight the sample by having a female parent in the home speak a language other than English and having a female parent in the home that has completed Year 12.

It was also possible to model 3 different types of non-response by postcode: refusals given to interviewers, opt-outs given at the time of initial contact by the HIC, and inability to contact the potential participant (e.g. families that had moved etc). See Attachment D for the results.

These results suggest that different factors underpin the various types of non-response. While the overall response rate in the LSAC is relatively low, exposing the survey sample to potentially significant biases, no one single factor has been found to account for all non-response. As separate mechanisms underpin families' decisions to opt-out of the survey at the initial approach, compared to refusing at the time of the approach by the interviewer and those families that could not be contacted for interview, the overall extent of bias in the sample is likely to be less than it otherwise may have been. This is also seen in the results of the bivariate analyses, where a number of significant differences were found, but overall differences were relatively small. Thus it appears unlikely that there is any one type of family that has not been represented in the sample at all. As the size of the non-response biases that have been identified is relatively small, considering the overall level of non-response, using a weighting strategy to adjust for potential non-response bias is likely to go a significant way towards addressing biases in the sample. It is also likely that weighted estimates will be representative of the overall population.

Design weight adjustments

To produce the final weights for each cohort, the design weights were adjusted using the process of calibration on marginal totals, using the two variables identified from the non-response analysis. The design weights were adjusted by a factor that was the inverse of the probability that the 'child' was likely to respond. Therefore the less likely a child was to respond, the higher the weight he/she was given. This may be thought of as increasing the contribution of children that have characteristics similar to other children who did not respond. Conversely, the more likely a child was to respond, the lower the weight boost it was given. A more detailed description of this process is given in Attachment E.

Outcome of weighting

Table 7 shows the effect of the weights on the estimates of census variables not used for weighting (see Attachment E for the effects on those used in the weighting). In general, the weighting procedure improved the estimation of these proportions, bringing the LSAC and census estimates up to 2% closer together in some cases. However for the infant cohort mother's country of birth and indigenous status of child produced estimates that were less accurate when weighted. For the 4-5 year old child cohort estimates of mother's country of birth, indigenous status of child, and number of people in the home were less accurate when weighted.

Table 7 Effect of weighting on estimates of variables shared by the 2001 census and LSAC

	INFANT			4-5 YEAR OLDS		
	Unweighted	Weighted	Census	Unweighted	Weighted	Census
Family type						
Dual parent	90.5%	89.4%	88.2%	85.9%	84.9%	82.1%
Single parent	9.5%	10.6	11.8%	14.1%	15.1%	18.0%
Mother's country of birth						
Australia	78.2%	76.7%	77.6%	75.0%	74.0%	74.7%
Other	21.9%	23.3%	22.4%	25.0%	26.0%	25.3%
Father's country of birth						
Australia	76.1%	74.4%	75.2%	73.3%	72.1%	72.2%
Other	23.9%	25.6%	24.8%	26.7%	27.9%	27.8%
Father's language spoken at home						
English only	86.5%	84.2%	82.8%	84.3%	82.3%	81.7%
Other	13.6%	15.8%	17.2%	15.7%	17.7%	18.4%
Father's highest school completion						
<Year 12	41.5%	43.3%	50.2%	47.3%	50.5%	54.7%
Year 12	58.5%	56.7%	49.8%	52.7%	49.6%	45.3%
Number of people in the family						
<5	77.7%	77.2%	74.4%	62.4%	61.7%	62.3%
5+	22.3%	22.8%	25.7%	37.6%	32.3%	37.7%
Indigenous status						
Non-indigenous	95.5%	95.1%	96.5%	96.3%	96.1%	96.5%
Indigenous	4.5%	4.9%	3.5%	3.8%	3.9%	3.5%
Nature of occupancy						
Renting	28.9%	30.8%	34.1%	26.7%	28.3%	31.3%
Other	71.1%	69.2%	65.9%	73.3%	71.7%	68.7%

Weights in the wave 1 dataset

Three weights have been included in the LSAC datasets:

- Child population weight – this weight would be used to produce population estimates based on the LSAC data (e.g. based on LSAC data there are approximately 22,464 infants in Australia that were never breastfed). The sum of the responding infant and 4-5 year old child population weights is 243,026 and 253,202, which is the ABS estimated resident population of children aged 0 and 4 years, respectively, at end March 2004, adjusted for the remote parts of Australia that were excluded from the study design;
- Child sample weight – this is the child population weight rescaled such that the sum of the weights matches the number of children in the sample (that is 5,107 infants and 4,983 4-5 year olds). This weight would be used in analyses that expect the weights to sum to the sample size rather than the population, particularly when tests of statistical significance are involved; and
- Day weight (time use dairy only) - this is the sample weight adjusted so that each day of the week receives equal weight in analyses of time use data.

Standard Errors and Design Effect

Design effect

The design effect associated with the *Growing Up in Australia* study is the loss in statistical precision that results from using clustered sample rather than a simple random sample of children from the HIC database.

Clustering achieves cost savings in the collection of the data, however a clustered sample may not give as precise estimates as a simple random sample of the same size, if variables being measured tend to be more similar within individual clusters. The design effect depends principally on two things: the degree to which responses tend to be similar within postcodes (measured by the intra-class correlation coefficient), and the size of the clusters (the number of participating children in each postcode).

The extent of such loss in statistical precision depends largely upon whether the issues that are the focus of the study are likely to have underlying geographic variations.

The potential effect of the sample design on the precision of estimates derived from a clustered sample is essentially related to the heterogeneity of the stratum population. If the members of a cluster are effectively no more like each other than they are to others within the stratum population, then the intra-cluster correlation is zero and there is no design effect. However, where regional clusters result in cluster members being more like each other and less like other members of the stratum population, then even where the intra-cluster correlation is quite small, there will be a design effect, the size of which is then dependent upon cluster size.

It should be noted that analysis of the relative precision of estimates derived from clustered and unclustered sample designs focuses only on sampling error and fails to adequately account for the more substantial reduction in 'non-sampling' error provided by a clustered sample design by freeing money for other purposes. This reduction in non-sampling error results from several factors, including:

- More valid measurement of specific issues that result from the more detailed and accurate information collected through face-to-face interviewing; and
- More reliable measurement of such issues through more efficient sample management, and consequent control against non-response bias resulting from sample loss through non-contacts, and ensuring greater sample retention over time.
- Reduced data collection costs per respondent mean that more respondents can be recruited to the study.

For this study, a clustered sample design offers a further advantage in providing for multiple observations within a community, increasing the capacity of the study to analyse community-level effects. The study design will result in there being sufficient children from a community to use community-level indicators in analysis (for example, various indices can be developed at the community level that can be used in analysis), and for comparisons to be made between children within the one community (for example, do parents within that community share similar or different views of that community).

Accuracy of estimates

Table 5 gives the 95% confidence intervals for different sample sizes assuming a design effect of 1.5 as an example. Approximately 90% of LSAC variables have a design effect lower than this figure.

Table 7 95 per cent confidence limits [(a)] for survey estimates of proportions

Sample size	Survey estimate of proportion									
	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%
5000	0.7%	1.0%	1.4%	1.6%	1.7%	1.7%	1.7%	1.6%	1.4%	1.0%
4000	0.8%	1.1%	1.5%	1.7%	1.9%	1.9%	1.9%	1.7%	1.5%	1.1%
3000	1.0%	1.3%	1.8%	2.0%	2.1%	2.2%	2.1%	2.0%	1.8%	1.3%
2500	1.0%	1.4%	1.9%	2.2%	2.4%	2.4%	2.4%	2.2%	1.9%	1.4%
1500	1.4%	1.9%	2.5%	2.8%	3.0%	3.1%	3.0%	2.8%	2.5%	1.9%
1000	1.7%	2.3%	3.0%	3.5%	3.7%	3.8%	3.7%	3.5%	3.0%	2.3%
500	2.3%	3.2%	4.3%	4.9%	5.3%	5.4%	5.3%	4.9%	4.3%	3.2%
250	3.3%	4.6%	6.1%	7.0%	7.4%	7.6%	7.4%	7.0%	6.1%	4.6%
100	5.2%	7.2%	9.6%	11.0%	11.8%	12.0%	11.8%	11.0%	9.6%	7.2%
50	7.4%	10.2%	13.6%	15.6%	16.6%	17.0%	16.6%	15.6%	13.6%	10.2%

(a) For example, for a (sub) sample size of 1000 and a variable that is estimated to be present in 50 per cent of the population, there is a 95 per cent chance that the true value is 50 per cent plus or minus 3.8 per cent - i.e. the true value is in the range 46.2-53.8 per cent.

The design effect may differ for each variable, as each variable could have a different intra-class correlation coefficient. In general, demographic variables often show the greatest intra-class correlation while health and wellbeing outcomes often show a lesser degree of clustering. The cluster size is reasonably large in this study, but as the clusters (postcodes) are themselves quite large and often heterogenous geographic units, the intra-class correlation coefficients for many variables will be small. Thus it is likely that the clustered design will allow greater flexibility in the analysis without significantly affecting the accuracy of the results. The design effect will also be influenced by the accuracy of the sampling frame and the differing response rates of different groups in the population.

Non-response from other study insstruments

There is interest in users of the data in how best to handle the non-response that has arisen from not all self-complete questionnaires from parents, teachers and carers being returned.

The following table shows the response from the other respondents in the study and to the other study materials. Response rates were highest for those materials that were or could be filled out by Parent 1 or Parent 2 (75-85% for Parent 1 and 2 self-completes and time use diaries (TUDs)). Response rates for teacher and carer questionnaires were lower, but given that these people hadn't consented to be part of the study prior to being sent a questionnaire, this is to be expected. While the response rates to the carer questionnaires and the Australian Early Development Index (AEDI) nested study is around the 50% mark, the teacher questionnaire had a high response rate of almost 70%.

Table 8 Final response for other study materials

	Possible	Placed	Total returned	Return rate (a)	Response rate (b)
			Infant		
Parent 1	5107	5024	4341	86.4%	85.0%
Parent 2	4630	4469	3696	82.7%	79.8%
TUD	5107	4983	4031	80.9%	78.9%
Carer	1219	1071	575	53.7%	46.3%
			4-5 year old		
Parent 1	4983	4907	4229	86.2%	84.9%
Parent 2	4286	4148	3388	81.7%	79.0%
TUD	4983	4873	3867	79.4%	77.6%
Teacher	4761	4667	3276	70.2%	68.8%
AEDI	1471	1366	721	52.8%	49.0%

(a) Return rate refers to the number of questionnaires returned when these were given to potential respondents

(b) Response rate refers to the number of questionnaires returned for all study participants who were eligible to have a questionnaire placed

Tables 9a to f refer to the characteristics of those that responded to the other study materials compared to the full sample. Comments on the other tables are given after the tables.

Table 9 Respondents compared with non-respondents for other study material

9a PARENT 1 SELF-COMPLETE	INFANT			4-5 YEAR OLD		
	LSAC	Census	P1SC	LSAC	Census	P1SC
Family Type						
2 resident parents/guardians:	90.5%	88.2%	91.8%	85.9%	82.1%	86.9%
1 resident parent/guardian:	9.5%	11.8%	8.2%	14.1%	17.9%	13.2%
Siblings						
Only child	39.5%	36.3%	40.0%	11.5%	12.2%	11.1%
One sibling	36.8%	35.8%	36.8%	48.4%	46.2%	49.7%
Two or more siblings	23.7%	27.9%	23.2%	40.1%	41.6%	39.2%
Ethnicity						
Aboriginal or Torres Strait Islander	4.5%	3.5%	3.7%	3.8%	3.5%	3.2%
Mother speaks a language other than English at home	14.5%	16.8%	13.0%	15.7%	17.6%	13.9%
Work Status						
Both parents or lone parent work	47.9%	na	48.8%	55.5%	na	56.1%
One parent works (in couple family)	40.8%	na	41.4%	32.8%	na	33.7%
No parent works	11.3%	na	9.8%	11.6%	na	10.2%
Educational Status						
Mother completed Year 12	66.9%	56.6%	68.9%	58.6%	48.3%	60.2%
Father completed Year 12	58.5%	50.2%	59.7%	52.7%	45.3%	53.0%
Parents' Income						
Less than $800 per week	31.7%	na	29.4%	29.2%	na	27.5%
$800-1499 per week	41.0%	na	42.1%	37.2%	na	37.9%
$1500 or more per week	27.3%	na	28.5%	33.6%	na	34.7%
State/Territory						
NSW	31.6%	34.8%	30.8%	31.6%	33.7%	31.3%
VIC	24.5%	24.1%	24.4%	25.0%	23.8%	24.7%
QLD	20.6%	19.1%	21.1%	19.8%	19.7%	20.5%
SA	6.8%	7.0%	6.5%	6.8%	7.2%	6.5%
WA	10.4%	9.6%	10.9%	10.2%	10.1%	10.3%
TAS	2.2%	2.3%	2.3%	2.7%	2.5%	3.0%
NT	1.7%	1.6%	1.7%	1.7%	1.6%	1.4%
ACT	2.1%	1.5%	2.4%	2.3%	1.3%	2.3%
Region						
Met	62.5%	65.1%	62.8%	62.1%	61.9%	61.6%
Exmet	37.5%	34.9%	37.3%	37.9%	38.1%	38.4%
Gender						
Male	51.2%	51.3%	51.5%	50.9%	51.3%	51.2%
Female	48.8%	48.7%	48.5%	49.1%	48.7%	48.9%
Total	5107		4339	4983		4229

Note: LSAC=Full LSAC sample, Census=ABS Census figures for families of 0 and 4 year olds, P1SC=Sub-sample of LSAC completing the Parent 1 Self-Complete Questionnaire, na=Comparison with census data not applicable

9b PARENT 2 SELF-COMPLETE	INFANT			4-5 YEAR OLD		
	LSAC	Couple	P2SC	LSAC	Couple	P2SC
Siblings						
Only child	39.5%	39.0%	40.0%	11.5%	8.9%	8.7%
One sibling	36.8%	37.5%	37.4%	48.4%	49.9%	51.7%
Two or more siblings	23.7%	23.5%	22.6%	40.1%	41.2%	39.6%
Ethnicity						
Aboriginal or Torres Strait Islander	4.5%	3.2%	2.4%	3.8%	2.9%	2.2%
Mother speaks a language other than English at home	14.5%	14.5%	12.4%	15.7%	16.2%	14.0%
Work Status						
Both parents or lone parent work	47.9%	50.4%	51.2%	55.5%	57.6%	58.4%
One parent works (in couple family)	40.8%	45.0%	44.9%	32.8%	38.2%	38.2%
No parent works	11.3%	4.6%	3.9%	11.6%	4.2%	3.4%
Educational Status						
Mother completed Year 12	66.9%	69.9%	72.5%	58.6%	62.1%	64.3%
Father completed Year 12	58.5%	58.5%	60.1%	52.7%	53.0%	54.6%
Parents' Income						
Less than $800 per week	31.7%	25.1%	23.0%	29.2%	19.5%	18.0%
$800-1499 per week	41.0%	44.9%	45.8%	37.2%	41.7%	42.2%
$1500 or more per week	27.3%	30.1%	31.2%	33.6%	38.8%	39.8%
State/Terrritory						
NSW	31.6%	31.5%	30.6%	31.6%	31.9%	31.1%
VIC	24.5%	24.6%	24.6%	25.0%	25.2%	24.9%
QLD	20.6%	20.2%	20.6%	19.8%	19.4%	20.3%
SA	6.8%	6.9%	6.7%	6.8%	6.7%	6.6%
WA	10.4%	10.7%	10.9%	10.2%	10.2%	10.3%
TAS	2.2%	2.2%	2.4%	2.7%	2.7%	3.1%
NT	1.7%	1.8%	1.7%	1.7%	1.6%	1.4%
ACT	2.1%	2.2%	2.5%	2.3%	2.4%	2.4%
Region						
Met	62.5%	63.3%	63.3%	62.1%	63.7%	62.7%
Exmet	37.5%	36.7%	36.7%	37.9%	36.3%	37.3%
Gender						
Male	51.2%	51.2%	51.4%	50.9%	50.6%	50.8%
Female	48.8%	48.8%	48.6%	49.1%	49.4%	49.2%
Total	5107	4630	3696	4983	4286	3388

Note: LSAC=Full LSAC sample, Couple=Sub-sample of LSAC with two resident parents, P2SC=Sub-sample of LSAC completing the Parent 2 Self-Complete Questionnaire

9c TIME USE DIARY	INFANT			4-5 YEAR OLD		
	LSAC	Census	TUD	LSAC	Census	TUD
Family Type						
2 resident parents/guardians:	90.5%	88.2%	92.4%	85.9%	82.1%	88.2%
1 resident parent/guardian:	9.5%	11.8%	7.6%	14.1%	17.9%	11.8%
Siblings						
Only child	39.5%	36.3%	40.5%	11.5%	12.2%	10.9%
One sibling	36.8%	35.8%	36.9%	48.4%	46.2%	50.5%
Two or more siblings	23.7%	27.9%	22.6%	40.1%	41.6%	38.6%
Ethnicity						
Aboriginal or Torres Strait Islander	4.5%	3.5%	3.1%	3.8%	3.5%	2.6%
Mother speaks a language other than English at home	14.5%	16.8%	12.0%	15.7%	17.6%	13.3%
Work Status						
Both parents or lone parent work	47.9%	na	49.8%	55.5%	na	56.6%
One parent works (in couple family)	40.8%	na	41.6%	32.8%	na	34.2%
No parent works	11.3%	na	8.6%	11.6%	na	9.2%
Educational Status						
Mother completed Year 12	66.9%	56.6%	71.0%	58.6%	48.3%	61.9%
Father completed Year 12	58.5%	50.2%	60.4%	52.7%	45.3%	53.9%
Parents' Income						
Less than $800 per week	31.7%	na	28.4%	29.2%	na	25.6%
$800-1499 per week	41.0%	na	42.2%	37.2%	na	38.3%
$1500 or more per week	27.3%	na	29.4%	33.6%	na	36.1%
State/Terrritory						
NSW	31.6%	34.8%	30.0%	31.6%	33.7%	30.8%
VIC	24.5%	24.1%	24.6%	25.0%	23.8%	25.0%
QLD	20.6%	19.1%	21.2%	19.8%	19.7%	20.5%
SA	6.8%	7.0%	6.4%	6.8%	7.2%	6.3%
WA	10.4%	9.6%	11.3%	10.2%	10.1%	10.7%
TAS	2.2%	2.3%	2.4%	2.7%	2.5%	3.0%
NT	1.7%	1.6%	1.8%	1.7%	1.6%	1.5%
ACT	2.1%	1.5%	2.4%	2.3%	1.3%	2.3%
Region						
Met	62.5%	65.1%	63.0%	62.1%	61.9%	61.6%
Exmet	37.5%	34.9%	37.0%	37.9%	38.1%	38.4%
Gender						
Male	51.2%	51.3%	51.6%	50.9%	51.3%	51.4%
Female	48.8%	48.7%	48.4%	49.1%	48.7%	48.6%
Total	5107		7780	4983		7449

Note: LSAC=Full LSAC sample, Census=ABS Census figures for families of 0 and 4 year olds, TUD=Sub-sample of LSAC completing the Time Use Diary (each case counted twice if completed two diaries), na=Comparison with census data not applicable

9e Carer self-complete	LSAC	Has HBC	HBC data	Has CBC	CBC data
Family Type					
2 resident parents/guardians:	90.5%	91.2%	93.6%	91.6%	92.7%
1 resident parent/guardian:	9.5%	8.8%	6.4%	8.4%	7.3%
Siblings					
Only child	39.5%	50.6%	54.1%	44.0%	44.6%
One sibling	36.8%	34.5%	33.9%	41.0%	39.5%
Two or more siblings	23.7%	14.9%	12.0%	15.0%	15.9%
Ethnicity					
Aboriginal or Torres Strait Islander	4.5%	2.3%	1.2%	3.5%	3.0%
Mother speaks a language other than English at home	14.5%	15.8%	9.7%	8.2%	9.9%
Work Status					
Both parents or lone parent work	47.9%	86.6%	89.2%	83.8%	84.5%
One parent works (in couple family)	40.8%	8.5%	9.1%	11.7%	11.6%
No parent works	11.3%	4.9%	1.8%	4.5%	3.9%
Educational Status					
Mother completed Year 12	66.9%	76.3%	78.4%	77.4%	78.8%
Father completed Year 12	58.5%	59.9%	59.3%	66.7%	67.1%
Parents' Income					
Less than $800 per week	31.7%	19.1%	15.9%	15.6%	15.4%
$800-1499 per week	41.0%	40.6%	42.5%	39.0%	39.5%
$1500 or more per week	27.3%	40.2%	41.6%	45.4%	45.2%
State/Terrritory					
NSW	31.6%	38.8%	35.7%	24.8%	23.2%
VIC	24.5%	25.5%	26.0%	23.9%	24.5%
QLD	20.6%	17.7%	22.2%	27.2%	26.6%
SA	6.8%	6.4%	6.7%	6.8%	9.4%
WA	10.4%	6.1%	4.1%	7.7%	7.3%
TAS	2.2%	2.2%	3.2%	2.1%	2.2%
NT	1.7%	1.8%	0.6%	3.3%	3.0%
ACT	2.1%	1.6%	1.5%	4.2%	3.9%
Region					
Met	62.5%	62.4%	59.7%	70.0%	67.4%
Exmet	37.5%	37.6%	40.4%	30.0%	32.6%
Gender					
Male	51.2%	53.3%	56.4%	48.7%	50.2%
Female	48.8%	46.7%	43.6%	51.3%	49.8%
Total	5107	792	342	427	233

Note: LSAC=Full LSAC sample, Has HBC=Sub-sample of LSAC attending home-based carer 8+ hours/week, HBC data=Sub-sample of LSAC with Home-Based Carer Questionnaire data, Has CBC=Sub-sample of LSAC attending centre-based carer 8+ hours/week, CBC data=Sub-sample of LSAC with Centre-Based Carer Questionnaire data

9f Teacher	LSAC	Has Teacher	Teacher data
Family Type			
2 resident parents/guardians:	85.9%	86.4%	87.3%
1 resident parent/guardian:	14.1%	13.6%	12.7%
Siblings			
Only child	11.5%	11.7%	10.6%
One sibling	48.4%	49.0%	48.9%
Two or more siblings	40.1%	39.3%	40.5%
Ethnicity			
Aboriginal or Torres Strait Islander	3.8%	3.5%	2.8%
Mother speaks a language other than English at home	15.7%	15.1%	13.9%
Work Status			
Both parents or lone parent work	55.5%	56.9%	58.0%
One parent works (in couple family)	32.8%	32.5%	32.8%
No parent works	11.6%	10.5%	9.3%
Educational Status			
Mother completed Year 12	58.6%	59.6%	60.3%
Father completed Year 12	52.7%	53.2%	54.0%
Parents' Income			
Less than $800 per week	29.2%	28.2%	26.7%
$800-1499 per week	37.2%	37.4%	37.9%
$1500 or more per week	33.6%	34.5%	35.4%
State/Terrritory			
NSW	31.6%	30.5%	29.9%
VIC	25.0%	25.5%	26.2%
QLD	19.8%	19.8%	19.3%
SA	6.8%	7.1%	6.8%
WA	10.2%	10.5%	10.7%
TAS	2.7%	2.7%	3.3%
NT	1.7%	1.7%	1.7%
ACT	2.3%	2.3%	2.1%
Region			
Met	62.1%	62.4%	62.9%
Exmet	37.9%	37.6%	37.2%
Gender			
Male	50.9%	50.9%	50.8%
Female	49.1%	49.1%	49.2%
Total	4983	4761	3276

Note: LSAC=Full LSAC sample, Has Teacher=Sub-sample of LSAC attending school, pre-school or a day care centre, Teacher data=Sub-sample of LSAC with Teacher Questionnaire data

Parent 1 self-complete

Table 9a shows the characteristics of those that completed the Parent 1 self-complete. In general, respondents to the Parent 1 self-complete had a similar distribution of characteristics to that of the full sample, although where differences were observed between the LSAC sample and the census (e.g. family type, mother's education, mother speaking a language other than English) these tended to be very slightly amplified. The exception to this pattern was ATSI status with children from an ATSI background being less represented in the Parent 1 self-complete sample compared with the full LSAC sample, bringing the proportion of ATSI children more in line with the census figures.

Parent 2 self-complete

The table for the Parent 2 self-complete looks at characteristics of this sample compared to the full sample and two-parent families within the sample. Again, respondents to this questionnaire were remarkably similar in characteristics to the sample from which they were drawn. However children from an ATSI background, those with mothers who use a language other than English, those whose parents didn't complete Year 12, and those from low-income families are slightly underrepresented in the Parent 2 self-complete data.

Time Use Diary

As for the Parent 1 self-complete, the children that form the time use diary sample have similar characteristics to the full sample, but where differences occur they tend to correspond with differences between the full sample and the census estimates (with ATSI status again being the exception). Although the pattern was similar, observed differences between the TUD sample and the full sample were slightly larger than those between the full sample and the Parent 1 self-complete sample.

Carers/teachers

Given the lower sample sizes obtained for these questionnaires we would expect somewhat greater percentage differences between those invited to have a questionnaire filled out and the sample obtained. However, particularly for the centre-based carer questionnaire, there seems to be little observable non-response bias. The same applies for the teacher questionnaires, where once again little non-response bias is observable.

Conclusion

Given the minimal bias that can be detected between the characteristics of all the children in the sample and the characteristics of those children for whom questionnaires were obtained from other survey informants, it does not appear necessary to provide separate weights for these informants. An initial trial of simply calculated weights to adjust for non-response for the Parent 1 self-complete showed very little adjustment to the estimates produced (<.5%).

After Wave 1

At each subsequent wave it is possible to calculate cross-sectional weights for that wave following the procedures outlined above with two additional modifications:

- The pattern of response can also be analysed in terms of the distribution of responses for items collected in previous waves. This may yield further insights into the mechanisms underpinning the non-response that could be incorporated into the weighting strategy; and
- If calibration is used for determining weights, the cross-sectional weight from the previous wave can be used as a starting point rather than the original design weight.

When calculating cross-sectional weights for the subsequent waves there are two additional issues to consider:

- Should population benchmarks be adjusted for migration in and out of the sampling frame? Children who were not living in Australia at the time of the sample selection, but who moved to Australia and are in the age range will be represented in population benchmarks at that time. An assessment would have to be made as to whether children who fall into this category could be assumed to be similar to other children selected in the survey, or whether they are systematically different. However, although these children are likely to be different, they are unlikely to be sufficient in number to make a difference to the estimates for this survey. If at later stages it is felt important to include children who did migrate to Australia at later ages, then an appropriate process for doing so would need to be determined.
- Children and families who migrate interstate, for example, and thus move out of the stratum they were originally selected in may need to be re-weighted with regard to the stratum they are actually residing in.

In addition to calculating cross-sectional weights for each wave, a set of longitudinal weights may need to be calculated. Rather than trying to represent an ever changing population at each point in time, longitudinal weights can take a cohort approach. The longitudinal weights would be based on the cross-sectional weights for Wave 1, but would be adjusted for attrition over time.

The development of longitudinal weights will depend on the pattern of sample attrition and the consequent pattern of missing data, and will not be assessed until attrition rates are known. For instance, if there was no attrition, or attrition occurred entirely at random, there would be little need to have separate cross-sectional and longitudinal weights.

Attachment A Design weight calculation

Stratification

The Primary Sampling Unit was a postcode area. The postcode area was either:

- A single residential postcode;
- A combination of a residential postcode and a post office box only postcode where the post office was located in the residential postcode area; and
- A combination of adjacent residential postcodes, and post office box only postcodes, where relevant.

For ease, postcode areas will be referred to as postcodes in further discussion.

Postcodes were stratified by state and capital city statistical division ("met")/rest of state ("exmet"). In addition, postcodes had to be divided into 2 size strata, as some postcodes had fewer children than the required cluster size (ie the number of children to be selected from the postcode) and it was not feasible to amalgamate these postcodes with others so that the minimum size was obtained. Group 1 postcodes had at least the minimum cluster size (see table A1). Group 2 comprised postcodes with less than the minimum cluster size. NT had a different design and details are available in LSAC Technical paper no. 1.

The number of children in a postcode with Medicare activity was based on the sum of the number of children born between March 1998 and February 1999 (4-5 year olds), and between March 2002 and February 2003 (infants), that HIC had registered on the Medicare enrolment database at the time of the statistical extract in March 2003, and who had had Medicare activity in the previous 12 months (4-5 year olds) or previous 6 months (infants).

Cluster sizes

For postcodes in size stratum 1, the probability that they would be selected was proportional to the number of eligible children that lived in the postcode. The following table indicates the cluster size that was used to determine the number of postcodes that were to be selected in each stratum.

Table A1 Minimum cluster size for determining number of postcodes for each stratum

	NSW	Vic	Qld	SA	WA	Tas	ACT
Met 1	80	80	80	40	80	40	40
Xmet 1	80	80	80	80	80	40	

For postcodes in size stratum 2, postcodes were selected with equal probability of selection.

Number of postcodes selected

The number of postcodes to be selected was determined using the ABS Estimated Resident Population (ERP) for children aged less than 12 months and children aged 4 years at June 2002.

The following table indicates these ERP estimates.

Table A2 ERP estimates for each state by region stratum

	NSW	Vic	Qld	SA	WA	Tas	NT	ACT
Met	107,638	88,742	44,550	25,705	34,557	4,869	3,479	8,120
Exmet	61,621	34,201	53,261	10,333	14,766	7,029	3,674	0

The HIC extract as at March 2003 was used to determine how the sample should be split size stratum 1 and 2. Tables A3 and A4 show the HIC estimates of the population distribution of eligible children as frequencies and as the proportion size stratum 1.

Table A3 HIC estimates for each state by region by size stratum

	NSW	Vic	Qld	SA	WA	Tas	NT	ACT
Met 1	108,832	85,199	45,672	25,315	33,424	5,188	3,385	8,203
Met 2	312	1,956	20	8	1,332	6	25	4
Exmet 1	56,292	26,733	48,090	6,891	10,244	6,487	3,562	0
Exmet 2	5,884	6,960	4,651	3,423	3,894	112	216	0

Table A4 Proportion of each state by region stratum in size stratum 1 according to HIC data

	NSW	Vic	Qld	SA	WA	Tas	NT	ACT
Met	99.7%	97.8%	100.0%	100.0%	96.2%	99.9%	99.3%	100.0%
Exmet	90.5%	79.3%	91.2%	66.8%	72.5%	98.3%	94.3%	

The proportions in Table A4 were then applied to the counts in Table A2 to estimate the numbers of eligible children in each state by region by size stratum based on the ERP. The results can be seen in Table A5. For example, for NSW Met the HIC estimated that 99.71% of the population was in size stratum 1. Applying this ratio to the ERP estimate, it was estimated that there are 107,638*.997141=107,330 in-scope children in NSW Met size stratum 1.

Table A5 ERP estimates for each state by region by size stratum

	NSW	Vic	Qld	SA	WA	Tas	NT	ACT
Met 1	107,330	86,750	44,530	25,697	33,233	4,863	3,453	8,116
Met 2	308	1,992	20	8	1324	6	26	4
Exmet 1	55,790	27,136	48,564	6,904	10,699	6,910	3,464	0
Exmet 2	5,831	7,065	4,697	3,429	4,067	119	210	0

Where the number of eligible children in size stratum 2 would have meant that either no or only 1 postcode would be selected, no selections were made from that stratum and the whole sample was selected from the size 1 stratum and the population numbers were adjusted accordingly. Table A6 shows how the population was finally distributed for the purposes of determining the number of postcodes selected. Note that at this stage the population was not adjusted for the remote exclusions.

Table A6 Final estimates of population distribution used to determine sampling fractions

	NSW	Vic	Qld	SA	WA	Tas	NT	ACT
Met 1	107638	86750	44550	25705	33232	4869	3479	8120
Met 2		1992			1325			
Exmet 1	55790	27136	48565	6904	10698	7029	3674	
Exmet 2	5831	7065	4696	3429	4068			

Due to characteristics of the sample design the number of children actually approached to participate in the study could not be known until sample selection was actually complete. Therefore postcode selection was performed to achieve a sample size of approximately 20000 children. The required sampling fraction was:

$$\begin{aligned} SF &= \frac{T}{E} \\ &= \frac{20{,}000}{502{,}545} \\ &= .039797 \end{aligned}$$

Where: SF is the sampling fraction

T is the target selected sample size

E is the Estimated Resident Population

Therefore, for size 1 strata, the number of postcodes to be selected in each stratum was:

$$S_1 = \frac{N_e \times SF}{C}$$

Where: S_1 is the number of postcodes selected in the stratum

N_e is the total number of children in the stratum according to the ERP

SF is the sampling fraction

C is the cluster size for the stratum

For example, for NSW met the formula was:

$$\begin{aligned} S_1 &= \frac{107638 \times .039797}{80} \\ &= 54 \end{aligned}$$

For size 2 strata, the number of postcodes to be selected in each stratum was further complicated by the fact that it was considered cost ineffective to select postcodes with fewer than 20 children in them. However, it was considered that these children would not differ significantly from other children in the size 2 strata, so replacement with children from the larger size 2 postcodes would be appropriate. Thus, these children were included in the ERP population of the stratum when determining how many children needed to be selected for the stratum, but excluded when calculating the number of postcodes needed to achieve this sample. Therefore, the number of postcodes selected in the size 2 strata was:

$$S_2 = \frac{N_e \times SF}{A}$$

Where: S_2 is the number of postcodes selected in the stratum

N_e is the total number of children in the stratum according to the ERP

SF is the sampling fraction

A is the average number of children per postcode in the stratum for postcodes with >20 children

For example, for NSW xmet size stratum 2 where the average number of children per postcode with 20+ children was 37.4, the formula was:

$$S_2 = \frac{5831 \times .039797}{37.4}$$
$$= 6$$

Table A7 indicates the number of postcodes that were selected in each stratum.

Table A7 Number of postcodes selected in each state by region by size stratun

	NSW	Vic	Qld	SA	WA	Tas	NT	ACT
Met 1	54	43	22	26	17	5	3	8
Met 2	0	2	0	0	2	0	0	0
Xmet 1	28	13	24	3	5	7	3	0
Xmet 2	6	8	5	4	5	0	0	0

Probability of selecting a postcode

The calculation of the probability of selecting a postcode was complicated by several postcodes being excluded from any chance of selection. Postcodes could be excluded from LSAC for two reasons both related to the cost-effectiveness of data collection: a) the postcode was in a remote area of Australia; b) the postcode had fewer than 20 children in-scope children at the time of selection.

For determining the probability of selection of postcodes in size 1 strata the following formula was used:

$$P_{s1} = \frac{n \times S_1}{N_r}$$

Where: P_{s1} is the probability of a size 1 stratum postcode being selected

n is the number of children in the postcode

N_r is the total number of children in the stratum minus those in remote postcodes

S_1 is the number of postcodes selected in the stratum

For the postcodes in size 2 strata, the selection probability is:

$$P_{s2} = \frac{S_2}{E}$$

Where: P_{s2} is the probability of a size 2 stratum postcode being selected

S_2 is the number of postcodes selected in the stratum

E is the number of postcodes eligible for selection

Selection of children

The children were selected from the most recent data available from the HIC to ensure that the address details were as current as possible and to allow as much time as was possible to ensure the younger infants were registered with HIC. To assist with this, the children were selected in 4 phases and the probability of selection was calculated for children in each phase.

The probability of selecting a child once their postcode had been selected for stratum size 1 postcodes:

$$P_{p1} = \frac{C}{n}$$

Where: P_{p1} is the probability of a child being selected once their postcode has been selected if in size stratum 1

C is the cluster size for the stratum

n is the number of children in the postcode

For stratum size 2 postcodes, all children were selected, therefore

$$P_{p2} = 1$$

Where: P_{p2} is the probability of a child being selected once their postcode has been selected if in size stratum 2

In all the above calculations, the number of children in the postcode is the number before any children were excluded for any reason (eg because their family had already been selected in the study) and the number of children selected in a postcode is the number selected before children were removed due to fact-of-death matching or because children from the same family (eg multiple births) had been selected.

The probability of selecting the 4-5 year old children could have been adjusted to allow for the fact that any 4-5 year-old children in families where infants had already been selected were excluded from a chance of selection. However, this was a complex process and required extra information to be supplied from HIC that it was not possible to obtain in the time frame. More importantly, the impact of this adjustment on the overall weight was likely to be minimal, and therefore judged to be not worth the effort it would require and the additional complexity it would introduce.

As indicated earlier, a few other children were excluded from having a chance of selection, because a sibling had already been included in earlier phases of the study. Again, no adjustment has been made for this in the design weights, for similar reasons to the above - the complexity of the adjustment which would require calculating probabilities separately for each phase for both infants and four year-olds that did not have a sibling who could be selected in another phase, and those that do. Also, the results are likely to be numerically unstable, while the impact of these exclusions (as there were so few of them) is substantively small.

The number of children that needed to be selected per postcode was initially calculated based on likely response rates given the Dress Rehearsal experience (56% response for infants; 53% response for 4-5 year olds). Phase 1 and 2 selections followed this design, but adjustments were made in phases 3 and 4 to increase the sample selected. This was to compensate for higher than expected non-response and so that a final sample of around 5,000 per cohort would be obtained.

Overall selection probability

Overall the chance of selection of each child, in each postcode and in each phase, is:

$$P_c = P_{si} * P_{pi}$$

Where: P_c is the probability of a child being selected

P_{si} is the probability of a postcode being selected in a stratum

P_{pi} is the probability of a child being selected once their postcode has been selected in a stratum

$_i$ is the size strata of the child

Design weights

In size stratum 1 the design weight for a child was then calculated as the inverse of the probability of selecting that child, ie:

$$D_1 = \frac{1}{P_c}$$

Where: D_1 is the design weight for a child in size stratum 1

P_c is the probability of a child being selected

However, in size stratum 2, given that extra cases were selected from the postcodes to make up for the exclusion of those with less than 20 children, using the inverse of the probability would lead to underweighting. Thus it was decided to use the inverse of the stratum sampling fraction as the design weight, that is:

$$D_2 = \frac{N_r}{ns}$$

Where: D_2 is the design weight for a child in size stratum 2

N_r is the total number of children in the strata minus those in remote postcodes

ns is the number of children selected in the stratum

No adjustment was made at this stage to ensure that the design weights for a cohort would sum to the ERP or another estimate for that cohort, rather than for half the total number of children each cohort. Given that the design weights would be substantially adjusted to allow for non-response, it was deemed more appropriate to make such an adjustment in the final stages of the weighting process.

In size stratum 1, the main source of variability in the design weights stems from the difference between the number of infants and 4-5 year olds at the time of the selection of postcodes, and the actual numbers at the time the children were selected. For example, a postcode in phase 1 was given a design weight of 40 for the infant cohort. At the time it was selected this postcode had 97 in-scope children living within it. It was assumed in the selection process that half of these children would be infants, and half 4-5 year olds, and half would have birthdays making them eligible for selection in phase 1 and half would be eligible in phase 3. Therefore this postcode was selected assuming that there would be approximately 24 infants eligible for selection in Phase 1. However, when it became to time to select the infants this postcode had 34 infants eligible for selection, meaning that each infant had less chance of selection than that which was assumed, causing the design weight to inflate by a factor of 34/24.

In size stratum 2 most of the variation in design weights comes from failing to exclude remote postcodes when determining the number of postcodes in each selection. This meant that more postcodes were selected than were theoretically required. For example, the design weight for children in WA exmet size stratum 2 was 10. This was because the number of postcodes used was determined based on there being 3938 children in the stratum. However, after the remote postcodes were excluded, there were only 1820 children in the stratum that were actually considered part of the target population, hence the design weight was substantially lowered.

Attachment B Chi-squared analysis

In order to identify demographic variables important in determining non-response a number of variables were identified for comparison with census variables. The population estimates of these variables for 0 and 4 year old children were obtained from the ABS from the 2001 Census of Population and Housing and these were compared to the responses from the LSAC Parent 1 interview.

The comparisons can be seen in Table B1. As would be expected with a sample of this size, many of the differences were statistically significant, but often represented a difference of only a few percentage points. In general, cultural factors such as country of birth and language seem not to be too important in explaining non-response, whereas education produced greater discrepancies.

Table B1 Simple bivariate comparisons between Census and LSAC data

		Census		LSAC		χ^2	p
		Freq	%	Freq	%		
Family type							
Infant	Dual parent	186453	88.2	4625	90.6	27.54	<.001
	Single parent	24963	11.8	482	9.4		
4-5 yo	Dual parent	195200	82.1	4283	86.0	51.44	<.001
	Single parent	42694	18.0	700	14.0		
Mother's country of birth							
Infant	Australia	161387	77.6	3987	78.1	0.85	.36
	Other	46651	22.4	1117	21.9		
4-5 yo	Australia	172767	74.7	3706	74.9	0.20	.65
	Other	58619	25.3	1239	25.1		
Father's country of birth							
Infant	Australia	137120	75.2	3523	76.1	1.96	.16
	Other	45133	24.8	1105	23.9		
4-5 yo	Australia	138274	72.2	3163	73.2	2.28	.13
	Other	53275	27.8	1157	26.8		
Mother's language spoken at home							
Infant	English only	172838	87.4	4364	85.5	18.95	<.001
	Other	34838	16.8	740	14.5		
4-5 yo	English only	190240	82.4	4168	84.3	12.30	<.001
	Other	40719	17.6	778	15.7		
Father's language spoken at home							
Infant	English only	150553	82.8	4002	86.5	44.25	<.001
	Other	31358	17.2	627	13.5		
4-5 yo	English only	156121	81.7	3644	84.4	21.00	<.001

	Other	35079	18.4	676	15.6		
Mother’s highest school completion							
Infant	<Year 12	89213	43.4	1688	33.1	220.98	<.001
	Year 12	116202	56.6	3410	66.9		
4-5 yo	<Year 12	117404	51.7	2044	41.4	209.26	<.001
	Year 12	109810	48.3	2895	58.6		
Father’s highest school completion							
Infant	<Year 12	89081	50.2	1887	41.5	124.27	<.001
	Year 12	89811	49.8	2657	58.5		
4-5 yo	<Year 12	102193	54.7	2013	47.3	93.94	<.001
	Year 12	84694	45.3	2244	52.7		
Number of people in the family							
Infant	<5	157185	74.4	3969	77.7	30.38	<.001
	5+	54231	25.7	1138	22.3		
4-5 yo	<5	148286	62.3	3114	62.5	0.05	0.82
	5+	89608	37.7	1869	37.5		
Study child gender							
Infant	Male	126757	51.3	2614	51.2	0.01	.92
	Female	120568	48.7	2493	48.8		
4-5 yo	Male	132617	51.3	2537	50.9	0.26	.61
	Female	126024	48.7	2446	49.1		
Indigenous status							
Infant	Non-indigenous	195795	96.5	4880	95.6	13.22	<.001
	Indigenous	7120	3.5	227	4.4		
4-5 yo	Non-indigenous	224299	96.5	4795	96.3	0.82	.36
	Indigenous	8130	3.5	186	3.7		
Nature of occupancy							
Infant	Renting	72704	34.1	1599	31.4	16.95	<.001
	Other	140594	65.9	3501	68.7		
4-5 yo	Renting	75067	31.3	1387	27.9	27.02	<.001
	Other	164742	68.7	3587	72.1		

Attachment C Poisson Regression analysis

The variables found to be associated with non-response from the analysis of overall census distributions compared with the survey sample distributions are likely to be inter-related. Accounting for one of these variables in the weighting adjustment could have the effect of correcting for several others. In order to identify a minimal set of variables responsible for non-response to be used in weighting response rate for a postcode was modelled by the proportion of families with children of the appropriate ages with certain characteristics according to the 2001 census using Poisson regression.

The predictors entered into the model had to meet two criteria to be included: a) data had to be collected as part of LSAC that matched closely with the way the census collected similar information (e.g. employment status was excluded as LSAC collected this data differently), and b) that the characteristic would cause similar people to cluster in a postcode (e.g. gender of the study child was excluded for this reason). As can be seen from Table C1, for both the infant and child cohorts the best predictors of response rate were the education level of mothers and whether mothers spoke a language other than English in the home.

Table C1 Results of Poisson regression modelling response rate for postcode by socio-demographic characteristics

	Estimate	Standard error	Chi-square	p
Infants				
Intercept	-.2535	.3003	0.71	.3985
Mother speaks LOTE	-.0065	.0016	17.24	<.0001
Mother Australian born	-.0028	.0018	2.33	.1273
Dual parent family	-.0025	.0027	0.88	.3478
Renting home	-.0006	.0013	0.23	.6320
Child Indigenous	-.0019	.0027	0.53	.4670
Mother completed Year 12	.0042	.0012	11.45	.0007
4-5 year olds				
Intercept	-.9394	.2588	13.18	.0003
Mother speaks LOTE	-.0046	.0016	8.52	.0035
Mother Australian born	.0017	.0018	0.84	.3606
Dual parent family	.0007	.0023	0.09	.7665
Renting home	-.0005	.0015	0.10	.7531
Child Indigenous	-.0010	.0027	0.13	.7146
Mother completed Year 12	.0027	.0011	5.82	.0158

Attachment D Poisson regression by type of non-response

As well as performing the Poisson regression to determine which variables to weight by, a second set of Poisson regressions were performed to identify the factors important in producing different types of non-response. Table D1 shows the results of this regression for rates of non-response due to refusals to interviewers. For both the infant and 4-5 year old samples, the proportion of families refusing to interviewers in a postcode was found to be linked to a higher proportion of families where the mother speaks a language other than English in the home and a lower proportion of families with mothers that had completed Year 12. Additionally for the infants a higher proportion of mothers born in Australia contributed significantly, while for the 4-5-year-olds a lower proportion of indigenous 4-year-olds in the area was also a significant predictor of higher refusal rates.

Table D1 Results of Poisson regression modelling non-response due to refusals to interviewers for postcode by socio-demographic characteristics

	Estimate	Standard error	Chi-square	p
Infants				
Intercept	-1.8313	.5377	11.60	.0007
Mother speaks LOTE	.0124	.0028	19.95	<.0001
Mother Australian born	.0073	.0035	4.48	.0343
Dual parent family	-.0041	.0048	0.71	.3980
Renting home	-.0032	.0024	1.67	.1969
Child Indigenous	-.0051	.0048	1.14	.2866
Mother completed Year 12	-.0051	.0023	5.11	.0237
4-5 year-olds				
Intercept	-1.6530	.4308	14.72	<.0001
Mother speaks LOTE	.0057	.0025	5.24	.0221
Mother Australian born	-.0002	.0030	0.01	.9433
Dual parent family	.0040	.0038	1.09	.2972
Renting home	-.0011	.0024	0.20	.6529
Child Indigenous	-.0144	.0053	7.36	.0067
Mother completed Year 12	-.0082	.0019	19.50	<.0001

Table D2 shows the results of the regressions predicting non-response due to opting out after the initial contact by the HIC. For the infant sample none of the predictors were statistically significant, while for the 4-5 year old sample, the only significant predictor of higher opt-out rate was a higher proportion of families renting their home

Table D2 Results of Poisson regression modelling non-response due to opting out after the HIC letter for postcode using socio-demographic characteristics

	Estimate	Standard error	Chi-square	p
Infants				
Intercept	-2.2137	.5404	16.78	<.0001
Mother speaks LOTE	.0039	.0028	1.94	.1631
Mother Australian born	.0018	.0034	0.27	.6022
Dual parent family	.0042	.0049	0.73	.3918
Renting home	-.0014	.0024	0.35	.5526
Child Indigenous	-.0021	.0048	0.19	.6624
Mother completed Year 12	-.0020	.0022	0.79	.3749
4-5 year-olds				
Intercept	-1.4382	.4238	11.52	.0007
Mother speaks LOTE	-.0002	.0024	0.01	.9357
Mother Australian born	-.0020	.0029	0.46	.4955
Dual parent family	.0009	.0038	0.06	.8024
Renting home	-.0070	.0024	8.34	.0039
Child Indigenous	.0005	.0047	0.01	.9212
Mother completed Year 12	.0010	.0018	0.29	.5926

Table D3 shows the regression results for predicting non-response due to not being able to contact the family. For the infants higher non-contact rates were significantly predicted by a higher proportion of families that were renting their home, a higher proportion of families with an indigenous infant and a lower proportion with a mother that had completed Year 12. For the 4-5 year old sample, a lower proportion of dual parent families, a higher proportion of families renting their home and a lower proportion of families with an indigenous 4-year-old were all significant predictors of a higher non-contact rate.

Table D3 Results of Poisson regression modelling non-response due to not being able to contact the potential participant for postcode using socio-demographic characteristics

	Estimate	Standard error	Chi-square	p
Infants				
Intercept	-2.5225	.5703	19.56	<.0001
Mother speaks LOTE	.0031	.0032	0.95	.3296
Mother Australian born	-.0009	.0039	0.05	.8218
Dual parent family	.0067	.0051	1.71	.1909
Renting home	.0080	.0027	8.63	.0033
Child Indigenous	.0102	.0042	5.91	.0151
Mother completed Year 12	-.0085	.0026	11.05	.0009
4-5 year-olds				
Intercept	-1.5799	.4699	11.31	.0008
Mother speaks LOTE	.0051	.0027	3.51	.0610
Mother Australian born	-.0022	.0034	0.43	.5141
Dual parent family	-.0082	.0041	3.90	.0481
Renting home	.0103	.0025	16.52	<.0001
Child Indigenous	.0113	.0038	9.04	.0026
Mother completed Year 12	.0002	.0021	0.01	.9312

Attachment E Weighting method

Differential non-response can affect the validity of survey results if not corrected for. Rather than using the inverse of the original selection probabilities as survey weights, these design weights have been adjusted for different response rates within different groups of the population. The result is that a higher weight is given to children in categories where lower response proportions were achieved.

To adjust for differential non-response, a variation of post-stratification weighting was employed. The weights were calculated using the generalised raking procedure of Deville and Särndal (1992). This method of producing survey weights in household surveys was pioneered by the French National Statistical Agency INSEE, and is commonly used in large-scale household surveys conducted by INSEE, Statistics Canada and the Australian Bureau of Statistics.

The generalised raking procedure sets out to determine the set of weights that will sum to the correct benchmark population totals for each separate benchmark variable that minimise the *difference* between the final survey weights and the initial survey weights. Initial survey weights were taken from the sampling design, and the distance function used was:

$$G = \sum_{i=1}^{n_h} \frac{1}{2}\left(\frac{w_i}{d_i} - 1\right)^2$$

Where : n_h is the number of responding children in stratum *h*,

d_i is the original design weight for the *i*th child

w_i is the final weight for the *i*th child as determined by the calibration procedure.

The distance function was minimised subject to boundary constraints that the final weight be in the range 10 to 100. However, in order to find a solution it was necessary to relax this constraint in Tasmania where final weights were constrained to be in the range 5 to 110. The distance function was minimised using the conjugate gradient method as described by Beale (1972), and the calculations were undertaken using SAS/IML software, based on a suggested algorithm of Deville, Särndal and Sautory (1993).

Within survey strata (state by part of state), weights were calculated to sum to marginal totals by sex of child, whether the mother's main language spoken at home was English, and mother's education level. A separate set of weights was calculated for each cohort.

The design weights were used as the initial point for the weighting procedure. In each region, the design weights were inflated by a factor equal to the overall response proportion in that region. As a result of the weighting adjustments, the overall distribution of weights is broader than the distribution of design weights.

In the Northern Territory Ex-Met, due to the small number of postcodes available for selection outside the Darwin area, and the number of postcodes that were excluded as being too remote, the design weight was set to one for all children. As a result the final weight in this region was based purely on the adjustment for non-response.

Benchmarks

Benchmark totals by age (for infants and 4 year-olds) and sex, for each state were provided by the ABS from the Estimated Resident Population (ERP) series, as at 31 March 2004. Since part of state figures were not available as at March 2004, Figures from the ERP for 30 June 2003 were provided by the ABS at the level of state by part of state (Met, Exmet). These figures were used to estimate the proportion of the population within each part of state region and these proportions were applied to the March 2004 ERP figures to estimate population benchmarks at the stratum level (for the purposes of the calculation of final weights, the small and large size strata were combined within each state by part of state group as the impact of the two methods of selection is accounted for in the calculation of the design weights).

Counts of enrolled children on the HIC database were obtained from the HIC split by state, part of state, and whether the child was living in a postcode that was included in the survey sampling frame. These counts were used to estimate the proportion of children in-scope of the survey frame in each state. These proportions were applied to the 2004 ERP figures to estimate the benchmarks that were ultimately used in the weighting programme.

Data was obtained from the 2001 Census of Population and Housing on the distribution of infants and four year-olds by whether the mother's language spoken at home was English, and mother's education level. These data were used to produce the marginal totals for use in the weighting algorithm.

Final weights were constrained to sum to the estimated benchmark total in each state by part of state region, and constrained to match the census proportions by language spoken at home and by mother's education level within each region.

Distribution of weights

The following graphs show the distribution of design weights, initial weights and final weights for each cohort. It can be seen that the adjustment for non-response broadens the overall distribution of the weights, and gives the distribution a slight skew.

Figure E1: Infants — Distribution of design weights

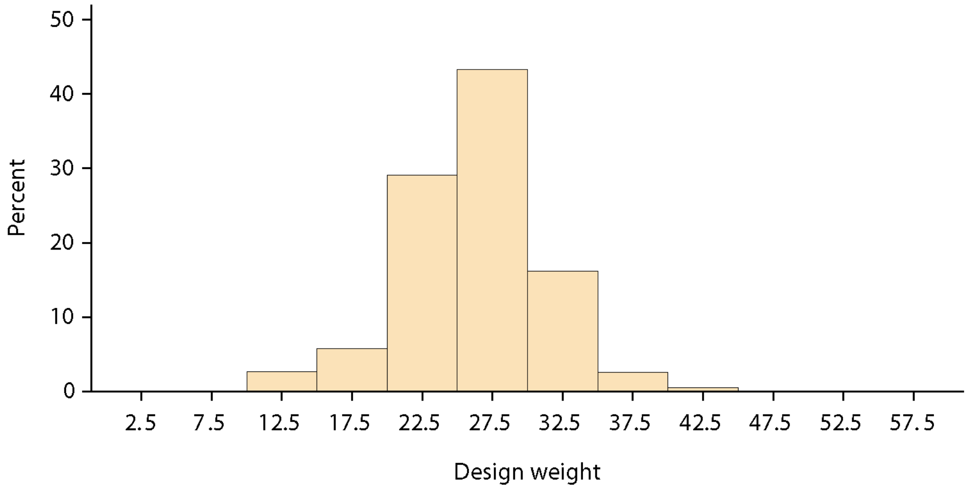

Figure E2: Infants — Distribution of initial weights

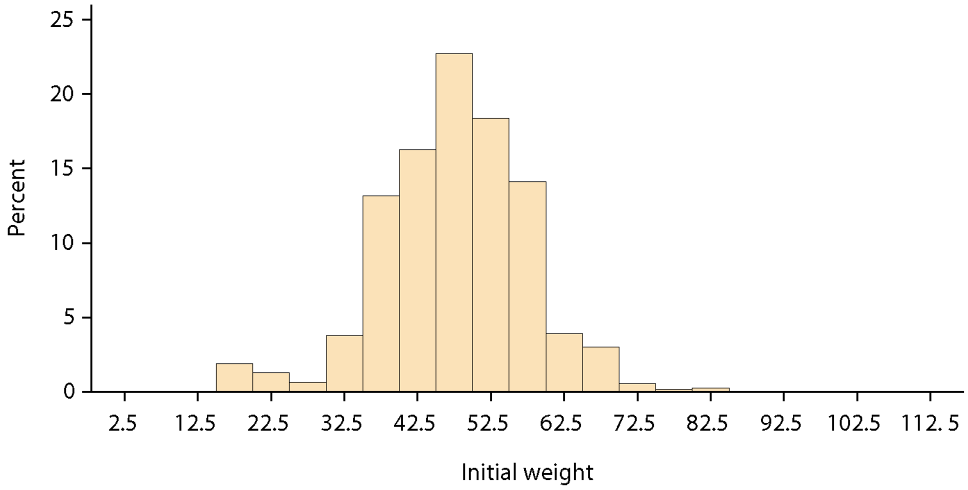

Figure E3: Infants — Distribution of final weights

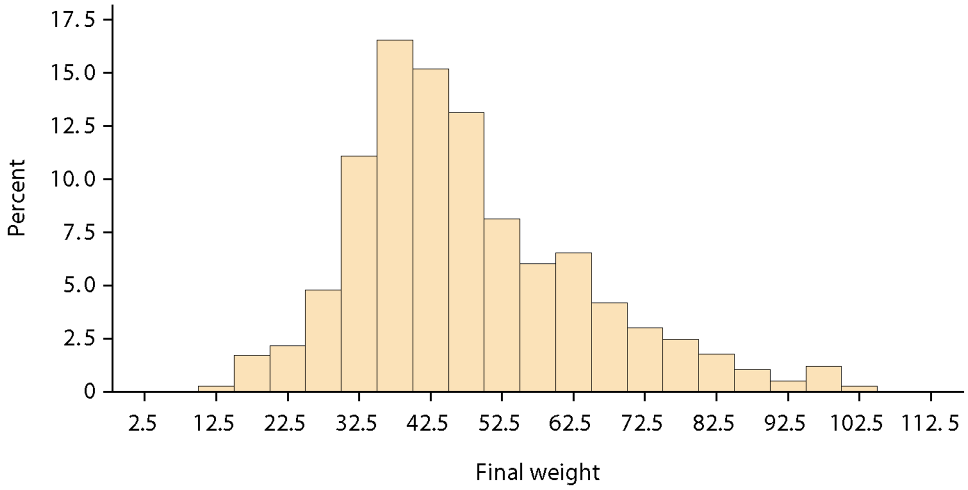

Figure E4: 4-5 year olds — Distribution of design weights

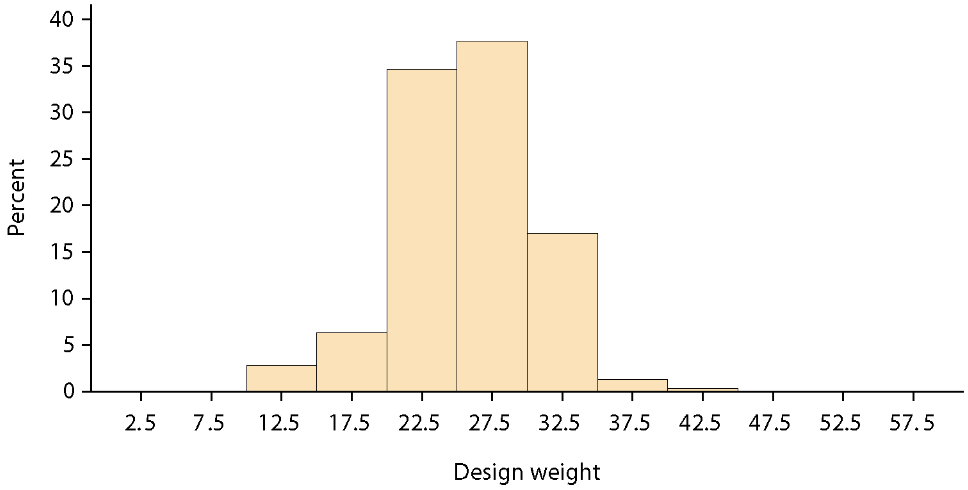

Figure 5: 4-5 year olds — Distribution of initial weights

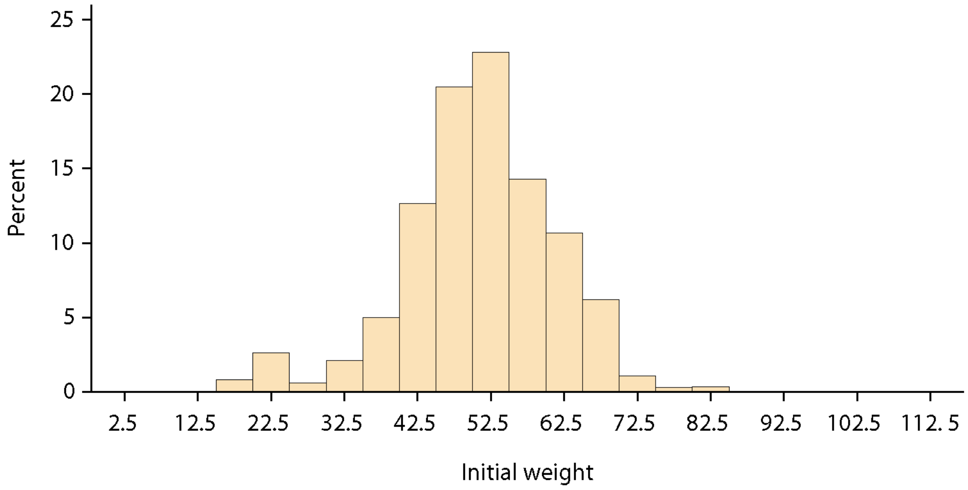

Figure 6: 4-5 year olds — Distribution of final weights

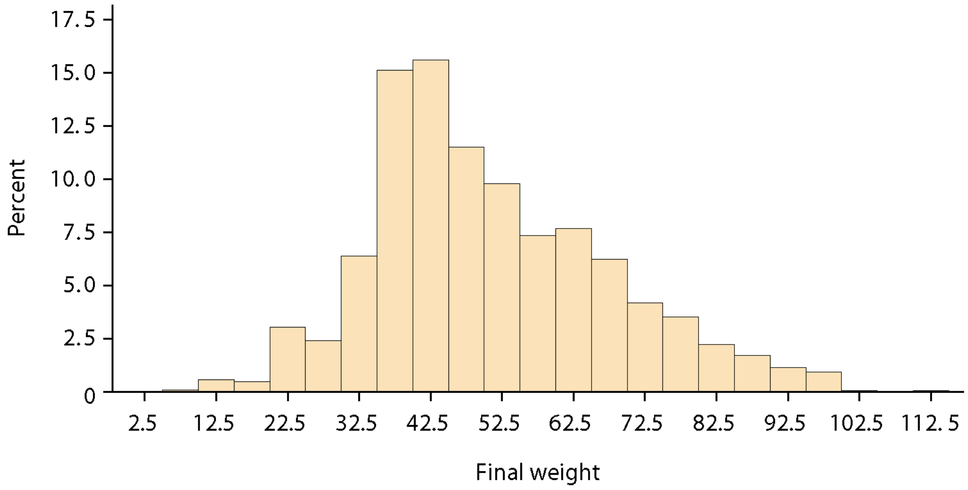

Table E1: Infants: Estimated numbers by state and part of state compared with ERP benchmarks

	Males				Females			
	ERP		LSAC		ERP		LSAC	
	Freq	%	Freq	%	Freq	%	Freq	%
State								
New South Wales	42,278	33.9	42,278	33.9	39,954	33.7	39,954	33.7
Victoria	31,644	25.4	31,644	25.4	30,063	25.4	30,063	25.4
Queensland	23,612	19.0	23,612	19.0	22,510	19.0	22,510	19.0
South Australia	8,690	7.0	8,690	7.0	8,382	7.1	8,382	7.1
Western Australia	11,908	9.6	11,908	9.6	11,408	9.6	11,408	9.6
Tasmania	3,034	2.4	3,034	2.4	2,858	2.4	2,858	2.4
Northern Territory	1,268	1.0	1,268	1.0	1,215	1.0	1,215	1.0
Australian Capital Territory	2,126	1.7	2,126	1.7	2,076	1.8	2,076	1.8
Part of state								
Metropolitan area	82,887	66.5	82,887	66.5	78,677	66.4	78,677	66.4
Rest of state	41,673	33.5	41,673	33.5	39,789	33.6	39,789	33.6
Australia	124,560	100.0	124,560	100.0	118,466	100.0	118,466	100.0

Note: ERP=ABS Estimated Resident Population of 0 and 4 year olds at March 2004

Table E2: 4 year-olds: Estimated numbers by state and part of state compared with ERP benchmarks

	Males				Females			
	ERP		LSAC		ERP		LSAC	
	Number	%	Number	%	Number	%	Number	%
State								
New South Wales	44,540	34.4	44,540	34.4	42,096	34.1	42,096	34.1
Victoria	31,467	24.3	31,467	24.3	30,409	24.6	30,409	24.6
Queensland	25,626	19.8	25,626	19.8	24,074	19.5	24,074	19.5
South Australia	9,225	7.1	9,225	7.1	8,755	7.1	8,755	7.1
Western Australia	12,379	9.5	12,379	9.5	12,041	9.7	12,041	9.7
Tasmania	3,206	2.5	3,206	2.5	3,030	2.5	3,030	2.5
Northern Territory	1,147	0.9	1,147	0.9	1,079	0.9	1,079	0.9
Australian Capital Territory	2,074	1.6	2,074	1.6	2,054	1.7	2,054	1.7
Part of state								
Metropolitan area	82,424	63.6	82,424	63.6	78,619	63.6	78,619	63.6
Rest of state	47,240	36.4	47,240	36.4	44,919	36.4	44,919	36.4
Australia	129,664	100.0	129,664	100.0	123,538	100.0	123,538	100.0

Note: ERP=ABS Estimated Resident Population of 0 and 4 year olds at March 2004

The above tables show weighted estimates of the numbers of infants and four year-olds by state or region compared with ERP benchmarks. As expected, the weighted estimates from each LSAC cohort exactly match the ERP benchmarks at the state and part of state level (Tables E1 and 2).

Proportions of survey children whose mother speaks a language other than English at home match census proportions at the state by part of state level, as they are constrained at this level. However, there is a slight difference between census proportions and survey estimates at higher levels, as shown in Tables E3 and 4. This is caused by the difference between census figures and the ERP figures used as benchmarks varying by region. This mainly reflects differential growth rates in different regions since the 2001 census. The overall weights are benchmarked to ERP figures and not census figures, so census proportions will only be matched exactly at the state by part of state level. However, the differences are minor (all less than 1%).

Table E3: Infants: Whether mother speaks a language other than English at home, LSAC estimates compared with census counts, by state

	Mother speaks English only at home				Mother speaks another language			
	Census		LSAC		Census		LSAC	
	Number	%	Number	%	Number	%	Number	%
New South Wales	55,411	77.9	63,515	77.2	15,753	22.1	18,717	22.8
Victoria	40,292	78.2	48,127	78.0	11,258	21.8	13,580	22.0
Queensland	37,255	92.7	42,741	92.7	2,954	7.3	3,381	7.3
South Australia	12,981	88.4	15,242	89.3	1,708	11.6	1,830	10.7
Western Australia	17,374	88.6	20,656	88.6	2,230	11.4	2,660	11.4
Tasmania	4,927	96.9	5,711	96.9	157	3.1	181	3.1
Northern Territory	1,588	84.8	2,110	85.0	284	15.2	373	15.0
Australian Capital Territory	3,010	85.9	3,609	85.9	494	14.1	593	14.1
Australia	172,838	83.2	201,709	83.0	34,838	16.8	41,317	17.0

Note: Census=2001 ABS Census figures for families of 0 and 4 year olds

Table E4: 4 year-olds: Whether mother speaks a language other than English at home, LSAC estimates compared with census counts, by state

	Mother speaks English only at home				Mother speaks another language			
	Census		LSAC		Census		LSAC	
	Number	%	Number	%	Number	%	Number	%
New South Wales	60,491	76.5	66,182	76.4	18,585	23.5	20,454	23.6
Victoria	44,788	77.6	47,916	77.4	12,934	22.4	13,960	22.6
Queensland	40,296	92.0	45,762	92.1	3,504	8.0	3,938	7.9
South Australia	14,800	87.9	15,814	88.0	2,035	12.1	2,166	12.0
Western Australia	19,352	87.9	21,500	88.0	2,655	12.1	2,920	12.0
Tasmania	5,660	97.4	6,077	97.4	149	2.6	159	2.6
Northern Territory	1,591	85.4	1,904	85.5	271	14.6	322	14.5
Australian Capital Territory	3,262	84.8	3,498	84.7	586	15.2	630	15.3
Australia	190,240	82.4	208,653	82.4	40,719	17.6	44,549	17.6

Note: Census=2001 ABS Census figures for families of 0 and 4 year olds

References

Beale EML (1972) A Derivation of Conjugate Gradients, in *Numerical Methods for Nonlinear Optimization*, Lootsma FA (ed.), London: Academic Press.

Deville, J.C. and Särndal, C.E. (1992), "Calibration estimators in survey sampling", *Journal of the American Statistical Association,* vol.87, pp. 376-382.

Deville, J.C., Särndal, C.E. and Sautory O. (1993), "Generalized raking procedures in survey sampling", *Journal of the American Statistical Association,* vol.88, pp. 1013-1020.