The Longitudinal Study of Australian Children:

an Australian Government initiative

LSAC Technical Paper No. 1

ISSN 1832-9918

Sample design

Carol Soloff, David Lawrence and Robert Johnstone

May 2005

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

Australian Government

Australian Institute of Family Studies

Contents

CONTENTS **2**

ABOUT THE AUTHORS **3**

CONSORTIUM ADVISORY GROUP MEMBERS 3
SAMPLING DESIGN TEAM MEMBERS 4
ACKNOWLEDGEMENTS 4

OVERVIEW **5**

INTRODUCTION **6**

SAMPLING FRAME **8**

BACKGROUND 8
MEDICARE ENROLMENT DATABASE 8
Scope *9*
Coverage *9*
Currency of address information *10*
Children on multiple Medicare enrolments *10*
Post Office boxes as address information *11*

SAMPLE DESIGN **12**

PRIMARY SAMPLE UNITS: POSTCODES 12
STRATIFICATION 12
Target population within postcodes *13*
Postcode exclusions *14*
CLUSTER SIZE 16

TARGET POPULATION **17**

TARGET POPULATION 17
EXCLUSIONS FROM SAMPLE SELECTION 17
Families with more than one target child *18*
Age range of children in the sample and sample selection phases *19*
Process for initial contact with selected families *20*

POSTCODE SELECTION **22**

SELECTION OF CHILDREN **24**

INITIAL SAMPLE SELECTION 24
HIC SELECTION OF CHILDREN 25
RESULTS OF SAMPLE SELECTION 26

SUMMARY **28**

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.

Robert Johnstone is the Data Manager for *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.

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.

Consortium Advisory Group members

Associate Professor Ann Sanson is the Project Director for *Growing Up in Australia*, and former Deputy Director Research and Acting Director of the Australian Institute of Family Studies. Ann is currently working in the Department of Psychology at the University of Melbourne.

Professor Stephen Zubrick is Chair of the Consortium Advisory Group. He is Head of the Division of Population Science at the TVW Telethon Institute for Child Health Research in Western Australia, and Co-Director of the Centre for Developmental Health at Curtin University.

Dr John Ainley is Deputy CEO and Research Director of the National and International Surveys Program at the Australian Council for Educational Research.

Dr Donna Berthelsen is a member of the Centre for Learning Innovation in the Faculty of Education, and Senior Lecturer in the Faculty of Early Childhood at the Queensland University of Technology.

Dr Michael Bittman is a Professorial Fellow at the Department of Sociology, School of Social Science, University of New England.

Dr Linda Harrison is a Senior Lecturer in Early Childhood in the Faculty of Education at Charles Sturt University.

Dr Jan Nicholson is a Senior Research Fellow in the Institute of Health and Biomedical Innovation and the Centre for Health Research (Public Health) at Queensland University of Technology.

Dr Bryan Rodgers is a Senior Fellow at the Centre for Mental Health Research, Australian National University.

Professor Michael Sawyer is Head, Department of Paediatrics and Associate Dean (Women's and Children's Health) at the University of Adelaide and Head, Research and

Evaluation Unit, Division on Mental health at the Child, Youth and Women's Health Service, Adelaide.

Professor Sven Silburn is Director of the Centre for Developmental Health at Curtin University and a Senior Research Fellow in the Division of Population Science at the TVW Telethon Institute for Child Health Research, Western Australia.

Dr Lyndall Strazdins is a Research Fellow at the National Centre for Epidemiology and Population Health, Australian National University.

Associate Professor Judy Ungerer is Head of the Division of Linguistics and Psychology at Macquarie University.

Professor Graham Vimpani is Professor of Paediatrics and Child Health at the University of Newcastle and Clinical Chair Kaleidoscope: Hunter Children's Health Network.

Dr Melissa Wake is Director of Research and Public Health in the Centre for Community Child Health, Murdoch Childrens Research Institute, Melbourne.

Sampling Design Team members

Ann Sanson, Steve Zubrick, Michael Sawyer, John Ainley, David Lawrence, Carol Soloff, plus

Professor John Carlin has appointments in the Departments of Paediatrics and Public Health, University of Melbourne, and is Director, Clinical Epidemiology and Biostatics Unit, Murdoch Childrens Research Institute.

Karen Wilson is Director of the Longitudinal Surveys Section at the Australian Government Department of Family and Community Services

Paula Chevalier is Assistant Director of the Longitudinal Surveys Section at the Australian Government Department of Family and Community Services

Acknowledgements

Initial work on the study design was undertaken by members of the *Growing Up in Australia* Consortium Advisory Group. Further development on the design has been undertaken by the principal authors, in partnership with the Consortium and with particular contributions from the Sampling Design Team. There has also been input from staff in the Health Insurance Commission and the Australian Bureau of Statistics.

Material has been used from the Longitudinal Study of Australian Children Proposal submitted by the Consortium to the Department of Family and Community Services, which included significant work undertaken by Jim Millwood, Consultant to the Consortium.

We are very grateful to Jan Nicholson for her useful comments on this paper and for the comments from Ann Sanson on early drafts of this paper.

Overview

Growing Up in Australia, the Longitudinal Study of Australian Children (also known as LSAC), 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. This paper outlines the full details of the sample design for this cross-sequential study comprising two 12-month age cohorts (infants and children aged 4-5 years olds).

With facilitation by the Australian Government Department of Family and Community Services, the Health Insurance Commission agreed that the sample could be selected from the Medicare database, the most comprehensive database of Australia's population. Data collection for Wave 1 of the study was undertaken by I-view, a social market research company, in conjunction with Colmar Brunton Social Research, a social research agency working in the government and not-for-profit sector.

A clustered design, based on postcodes, was chosen as it allows community level effects to be measured and analysed, and also allows for reasonably cost effective face-to-face interviewing. Every effort was made to ensure that the sample chosen would be as representative as possible of Australia's infants and 4-5 year olds.

A two-stage clustered design was employed, first selecting postcodes then children, allowing analysis of children within communities. Children in both cohorts were selected from the same 311postcodes. An average of 40 children per postcode in the larger states and 20 children per postcode in the smaller states and territories participating in the study.

Stratification was used to ensure proportional geographic representation for states/territories and capital city statistical division/rest of state areas. Postcodes were selected with probability proportional to size selection where possible, and with equal probability for small population postcodes.

Children were selected with approximately equal chance of selection for each child (about one in 25). Due to excessive data collection costs, 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.

The sample design was developed in collaboration with the LSAC Consortium's Sampling Design Team. The design posed a number of challenges and a brief rationale behind each design feature is included in this paper. For further information on the rationale for any design aspect, contact growingup@aifs.gov.au. In addition, any comment on this paper is welcome.

Introduction

Growing Up in Australia, the Longitudinal Study of Australian Children (LSAC), 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. The study is being funded by the Australian Government Department of Family and Community Services (FaCS) as part of the Australian Government's *Stronger Families and Communities Strategy*.

Growing Up in Australia will identify the developmental pathways that Australian children follow and the factors (both risk and resilience) that predict the course of these pathways. It is therefore important that the sample of children selected for the study is as representative as possible of Australian children, so that the results can be generalised to all Australian children.

The essential focus of the study design is on the early years of children's lives, and therefore defines 'the child' as the sampling unit of interest. The sample was intended to be representative of all 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.

By following two cohorts whose ages will overlap as the study progresses, the design is cross-sequential in nature. Cross-sequential designs have a number of advantages over simple single-cohort designs (see LSAC Discussion Paper No. 1, 'Introducing the Longitudinal Study of Australian Children').

The sample design requirements specified by FaCS were that:

- The minimum sample size of each cohort at the first data collection point should be one per cent of the population in each selected age cohort (about 2,500 children);
- The sample should be representative of all Australian children in each of the selected age cohorts, that is, proportional to the regional distribution of the Australian population;
- Study informants should include the child's parents and the child (when of an appropriate age); and
- Over-sampling of sub-populations is not required.

The sample design was developed in collaboration with FaCS and the LSAC Consortium's Sampling Design Team, which comprised members with statistical and practical sample design experience in the social sciences. There was extensive discussion with stakeholders about the optimal composition of the sample.

One sample design that was considered, was the proposal that *Growing Up in Australia* should be over-sampled for children with particular characteristics (for example, children from Indigenous or culturally diverse families or children with disabilities of various sorts) as these are groups of particular interest for policy development. However, a major strength of a study like *Growing Up in Australia* is the large and nationally representative nature of its sample.

In addition, oversampling for small subgroups of the population tends to only give limited improvement in the statistical precision of the aggregated estimates. While a significant increase in the numbers sampled in such subgroups can increase the statistical precision of the low prevalence estimates, it would not be an efficient use of the sample. Greater statistical precision of aggregated estimates can be achieved by distributing sample across all of the population.

It was therefore concluded that more intensive studies of subgroups are better conducted as separate studies, perhaps nested within or linked to *Growing Up in Australia* (Nicholson, Sanson and the LSAC Research Consortium, 2003).

Consideration was also given to ensuring a minimum sample size in each state and territory (for example, 1000 children per cohort for each state and territory). However, this would add significantly to the costs of the study since it would involve a large increase in the sample size. An alternative is to increase the sample in the smaller states through the transfer of sample from the larger states. However, this would result in a less efficient sample at the Australian level, without a significant improvement in the accuracy of state level estimates.

It was therefore decided that funds would be best spent on ensuring high quality comprehensive information from a nationally proportional sample with at least 5,000 children and their families from each cohort recruited to the study. A sample size of 5,000 was chosen as it provides a large enough initial sample to ensure that there will still be sufficient sample after a number of years to allow for detailed analyses to be undertaken.

Sampling frame

Background

Locating the *Growing Up in Australia* target populations of children aged less than 12 months and children aged 4 years was not a straightforward task. These populations are relatively rare in the Australian population – they each make up just over one per cent of the total Australian population. Based on Australian Bureau of Statistics Census of Population and Housing figures[1], about one in 15 Australian households at any one time has a child aged under 12 months or aged four years.

Locating rare populations through methods such as area sampling or telephoning households are expensive and inefficient, even though efficiency can be increased through over-sampling from areas with likely higher concentrations of the target population. Use of administrative sources provides a much more efficient means for identifying a rare population. However, administrative sources also have limitations, principally with the extent of coverage of the target population and the currency of information.

A number of possibilities for the sampling frame were considered in earlier stages of the development of this study. These included using reverse telephone directory CD-ROM or random digit dialling to help locate in-scope families, or making use of Birth Registry or Family Tax Benefits recipient listings. The latter options were not investigated in any detail because these listings were not able to be used as the source for the study sample (due to legislative requirements). In addition, all of the methods investigated had other significant problems and/or costs associated with them (see LSAC Discussion Papers 1 and 2 for further information).

With facilitation by FaCS, the sampling frame was extracted from the Medicare enrolment database held by the Health Insurance Commission (HIC). Medicare records contain data on date of birth and contact address, and hence provide a direct way of locating the required target population. The Medicare enrolment database was the only administrative source that could supply reasonably current information on both cohorts for all of Australia for *Growing Up in Australia*.

Medicare enrolment database

The major advantage of the Medicare enrolment database over any other available sampling frame is that both cohorts of children can be directly identified from this source, through use of the date of birth field on the database. This is a cost efficient search method for finding the *Growing Up in Australia* target population and it meant that selected families could be contacted using a personal pre-approach letter, rather than cold-calling.

---

[1] There were about 7 million households in Australia and about 450,000 of these have children aged under 12 months or 4 years - Australian Bureau of Statistics 2002, Census of Population and Housing: Selected Social and Housing Characteristics, Australia (catalogue no. 2015.0) plus unpublished Census data.

In addition, use can be made of the age and sex information of people listed on a Medicare card to provide some level of non-response analysis.

The following sections detail some of the issues associated with the use of Medicare enrolment database as the sampling frame.

Scope

Theoretically, Medicare includes all Australian residents. In practice, it can exclude some Australian residents who have access to alternative health services such as some defence force personnel, prisoners, and persons eligible for Department of Veteran's Affairs' Health Services. This does not affect the *Growing Up in Australia* target population, as any dependent children of these persons still need to be enrolled for Medicare.

Conversely, Medicare's population base is expected to exceed the ideal population base (i.e. all Australian residents), as it includes an unknown number of international visitors and former Australian residents who have died or are now permanently resident overseas but have not yet been removed from the database.

Medicare also includes some non-permanent residents who have access to Medicare (such as some temporary residents, and people from countries with Reciprocal Health Arrangements). These people were excluded from the sample through the specification of the required enrolment type.

Children with an end-date, indicating that the child has either died or been cancelled from that enrolment for some reason, were able to be excluded.

Coverage

The Health Insurance Commission reported that coverage of children by the Medicare register, especially the recently born and toddlers, has been enhanced by the introduction of the Australian Childhood Immunisation Register (ACIR) – a subset of the Medicare enrolment file that contains information on children aged 7 years and under.

It was expected that the coverage provided by this database for 4 year old children would be very good. When compared with the Australian Bureau of Statistics data at December 2001[2], the number of Medicare enrolments of children aged 4 years was 101.5 per cent of the Australian Bureau of Statistics' estimated resident population figures.

However, coverage for children aged less than 12 months was expected to be incomplete due to the time lag in registering babies with Medicare. Information from the last evaluation report of the Australian Childhood Immunisation Register (Hull et al 2002, p5) indicated that once missing data were excluded, about 80 per cent of children were registered on Medicare by 2 months, just over 90 per cent by 4 months and 98 per cent by 12 months (Hull at al 2001, p46). When compared with Australian Bureau of

[2] Correspondence with the Australian Bureau of Statistics.

Statistics estimated resident population figures, this means that only 88.5 per cent of all children aged under 12 months were on the HIC frame.

Given only recent births were under-represented in the HIC database (and this was verified by comparing HIC and ABS birth month data), some consideration was given to only including children who were at least 4 months old at the time the HIC sample would be selected. For reasons outlined in the later section on 'Age range of children in sample', it was decided that these younger children still needed to be included in the sample.

Currency of address information

The Health Insurance Commission is likely to be notified of a change of address through cardholder contact with a Medicare Branch with regard to patient claims, replacement for a lost, stolen or expired card, or through its card replacement program every seven years. There is no opportunity for address details to be checked when people are bulkbilled, which accounted for 68 per cent of claims in 2003-2004 (data from the HIC 2004 Annual Report) However, current facilities now give Medicare cardholders the ability to update their address across a range of government services and make it easier for people to lodge address changes over the telephone or the internet.

It was expected that address information for families with young children should be reasonably current, especially for those with infants who have recently been registered with Medicare.

At the time of designing the sample, there was no good source of information on how current the address information was. HIC staff estimated that 5-10 per cent of addresses may not be current. Despite the sample loss that would result from this, the HIC database still offered the best source for cost effectively locating the required sample.

Children on multiple Medicare enrolments

It is possible for a child to be registered on more that one Medicare card. Where parents each have their own Medicare number, children are generally listed on each parent's card. This can occur for both intact and separated families.

Unpublished Medicare data provided to the Institute showed that no child under 6 years of age is on 3 or more enrolments, with about 2 per cent of children under 12 months and 8 per cent of children aged 4 years on two enrolments. Children on two enrolments were only given one chance of selection, using the following selection rules:

- For children who had a Medicare claim or ACIR service in the last 12 months (for 4 year olds) or 6 months (for infants), contact details were taken from the card which was used for this last service;
- For children without a claim in this period, the card with an adult female was selected;
- For children without an adult female on either card, the 'primary' card (the card that the child was first registered on) was selected.

These selection rules were adopted after considerable work was undertaken to determine the optimal process for deciding which card to associate the child with, that would maximise the probability that the cardholder would be the child's primary caregiver. The selection rules also needed to be relatively straightforward to apply.

Post Office boxes as address information

About 7 per cent of *Growing Up in Australia* target children are likely to have a post office box as the contact address (based on unpublished data provided by the Health Insurance Commission). The proportion of families who use post office box numbers is particularly high in the Northern Territory (about 30 per cent). For these families to be interviewed, a residential address needed to be obtained.

Sample design

Primary sample units: postcodes

A clustered (by area) sample design is desirable for two reasons: it provides the opportunity to gather multiple observations within a community, increasing the capacity of the study to analyse community level effects; and it offers the opportunity to cost-effectively conduct face-to-face interviews.

The geographic indicator available through Medicare is postcode. This has some challenges for sample design purposes when interviewing is to be conducted face-to-face at the child's home. A postcode can cover a wide geographic area and one postcode can include urban, rural and remote areas. The possibility of coding the 500,000 addresses of the target population to Census Collector Districts was investigated but was not an operationally feasible option. Automated coding programs will at best code up to 80 per cent of records, meaning that about 100,000 records would have to be manually coded. This process would have needed to be undertaken by the Health Insurance Commission and was not possible in the time frame. It would also have been a very expensive exercise. In addition, despite their limitations, postcodes do offer a degree of clustering that would not have been available through some telephone contact methods.

Stratification

Postcodes were stratified by state/ territory and by capital city statistical division vs. rest of state ('met'/exmet') to ensure the sample was distributed across strata in the same proportions as children in the target populations.

Postcodes were allocated to a stratum using the following process:

- All postcodes in a HIC March 2003 statistical extract were matched to the Australia Post file of all postcodes. This identified which postcodes were residential and which were post office box or large volume receiver (LVR).
- All residential postcodes were then assigned to a state and to met/exmet using the postcode to statistical local area (SLA) concordance, as at June 2002. Mismatches were resolved through consultation with ABS. For the purposes of this design, the postcode was matched to the SLA that contained the greatest proportion of the population in that postcode, as indicated in the concordance.
- Non-residential postcodes were manually assigned to met or exmet.

In addition, in order to identify postcodes in remote areas, a concordance to the ABS Remoteness Area was used for residential postcodes, and a number of these were subsequently excluded from the sampling frame (as described in the later section 'Postcode exclusions').

Target population within postcodes

The number of postcodes within Australia listed by Australia Post at the time of selection of postcodes is shown in Table 1.

Table 1 Number of postcodes by state and type

Category	ACT	NSW	NT	QLD	SA	TAS	VIC	WA	Australia
Residential area	23	598	25	433	336	117	694	384	2610
LVR	9	309	1	25	20	44	29	13	450
Post office boxes	6	96	12	4	4	3	43	97	265
Total	38	1003	38	462	360	164	766	494	3325

Source: Australia Post postcode file, available on internet http://www1.auspost.com.au/postcodes/

Of these postcodes, 577 were not listed in the Medicare enrolment database, as shown in Table 2. Postcodes not appearing in the Medicare database were mainly LVR and some post office boxes, but also some residential areas did not have any target population. Table 2 shows the number of postcodes in the database by met/exmet status.

Table 2 Distribution of postcodes with at least one infant or 4 year old children in the postcode, HIC unpublished data[(a)]

Category	ACT	NSW	NT	QLD	SA	TAS	VIC	WA	Australia
Met	29	260	19	132	126	39	269	117	991
Exmet	1	416	19	300	207	83	430	301	1757
Total	30	676	38	432	333	122	699	418	2748

[(a)] Statistical extract taken in March 2003 for children born Mar 2002 – Feb 2003, and Mar 1998 – Feb 1999.

Of the postcodes listed in the Medicare enrolments database, over 900 contained very few children from the target populations, as shown in Table 3.

Table 3 Distribution of postcodes by number of infants plus 4 year old children, HIC unpublished data[(a)]

	Met	ACT	NSW	NT	QLD	SA	TAS	VIC	WA	Australia
80+ active(b) children	Met	19	222	10	109	92	18	194	82	746
	Exmet		161	11	116	31	17	88	31	455
	Total	19	383	21	225	123	35	282	113	1201
20-79 active children	Met	6	28	7	14	19	9	50	21	154
	Exmet		114	6	86	59	26	122	78	491
	Total	6	142	13	100	78	35	172	99	645
<20 active children	Met	4	10	2	9	15	12	25	14	91
	Exmet	1	141	2	98	117	40	220	192	811
	Total	5	151	4	107	132	52	245	206	902
Total		30	676	38	432	333	122	699	418	2748

[(a)] Statistical data were not supplied by HIC for postcodes containing 1-5 children, due to confidentiality requirements. Values for these postcodes were imputed. [(b)] See later section for definition of active children.

To enable the sample selection process to include children from areas with a low target population, whilst balancing the practical costs of data collection, a number of postcodes with very few children from the target populations were combined into 'postcode areas'. The following approach was used[3]:

- In metropolitan areas, residential postcodes with small numbers of children were combined with adjacent postcodes. This was a manual process. As far as possible, postcodes of similar SEIFA and within the same SLA were combined. This was done for 118 postcodes and resulted in 51 combined postcode areas.
- In non-metropolitan areas, where it was sensible to do so, residential postcodes with small numbers of children were also combined with adjacent postcodes. This was done for 44 postcodes and resulted in 16 combined postcode areas.
- Non residential postcodes with (an estimated) 10 or more children in the target population were combined with an appropriate residential postcode. This was also a manual process and resulted in 123 post office postcodes being combined into 54 combined postcode areas.

Postcode exclusions

In general, any remaining postcode with less than 20 active children in the target population was excluded from the selection process. This involved 605 postcodes and an estimated 1 per cent of the target population. It was decided to exclude these postcodes as these only contained a small proportion of the target population but would have been likely to incur disproportionately high per child data collection costs if any of these postcodes were selected.

This resulted in 1979 postcode areas being available for selection.

Remote locations

A number of children in remote locations were excluded because they are in postcodes that have very few children, as outlined above. In addition, there were some very remote locations where the benefits of obtaining data were not sufficient to justify the expense of data collection. It was agreed that postcodes with the following characteristics would be excluded:

- Postcodes that were very large in area (more than 100,000 sq. kms).
- Postcodes that were on Indigenous land (these also were very large in area).
- Postcodes with more than 50 per cent of children aged 4 years at the time of the 2001 ABS Census who were of Aboriginal or Torres Strait origin[4].
- Postcodes with 80 or fewer children with Medicare activity.

---

[3] For the purposes of working with these data, values were imputed for postcodes containing 1-5 children. For each state, the number of children in these postcodes was derived by subtracting the number of children in postcodes with 6 or more children from the state total. Then the average number of children per postcode with 1-5 children was calculated and used for each postcode.

[4] 'Proportion of children aged 4 years' at the time of the ABS Census of Population and Housing was used as an indicator of areas that have a high Indigenous population, where it was expected that very low recruitment rates would be obtained for high expense in data collection.

ABS Census data indicate that 4.3% of children aged 0 and 4 years are of Aboriginal or Torres Strait origin. After exclusion of the above postcodes, the Census data indicates that 3.5% of the remaining population was of Aboriginal or Torres Strait origin.

This exclusion of postcodes resulted in a further 268 postcode areas being excluded from selection, which excluded about 40 per cent of children living in remote areas from the chance of being selected in the sample.

The following tables show the final distribution of postcode areas[5] included and excluded in the selection process, including estimates of the proportion of the target population expected to be excluded from a chance of selection as a result of the excluded postcodes.

Table 4 Distribution of postcode areas[(a)] eligible for selection

	ACT	NSW	NT	QLD	SA	TAS	VIC	WA	Australia
Met	21	232	8	114	108	27	245	103	858
Exmet	0	265	3	175	80	38	224	68	853
Total	21	497	11	289	188	65	469	171	1711

[(a)] See section above 'Target population within postcodes' for a description of how 'postcode areas' were constructed.

Table 5 Distribution of postcodes eligible for selection

	ACT	NSW	NT	QLD	SA	TAS	VIC	WA	Australia
Met	26	261	17	130	125	35	258	119	971
Exmet	2	266	5	176	80	70	232	72	903
Total	28	527	22	306	205	105	490	191	1874

Table 6 Distribution of postcodes deleted from selection

	ACT	NSW	NT	QLD	SA	TAS	VIC	WA	Australia
Met	2	60	1	3	2	1	16	66	151
Exmet	0	89	15	124	126	9	199	161	723
Total	2	149	16	127	128	10	215	227	874

Table 7 Proportion of total population excluded from chance of selection

	ACT	NSW	NT	QLD	SA	TAS	VIC	WA	Australia
Met	0.05%	0.29%	0.74%	0.04%	0.03%	0.11%	0.10%	0.90%	0.24%
Exmet	0.00%	2.55%	65.87%	7.86%	16.87%	2.71%	5.59%	22.57%	8.29%
Total	0.05%	1.11%	34.96%	4.23%	4.91%	1.56%	1.63%	7.17%	3.18%

[5] A postcode area can either be a single residential postcode, a combination of adjacent residential postcodes, or a combination of residential and post office box postcodes.

Cluster size

It was felt to be important that an appropriate balance was found between the number of postcodes included in the study, and the number of children selected within a postcode (cluster size). Sufficient postcodes needed to be chosen across Australia to help ensure the representativeness of the sample and to help minimise sampling error, but there needed to be sufficient children selected within each postcode for operational efficiency and to allow for analysis of community level effects.

Having considered the approach taken in other studies, and weighing up the statistical, analytical and operation implications of different cluster sizes, it was decided that, in general, a (final) cluster size of around 20 children per cohort per postcode was desirable.

For example, based on simulations that were performed on data from the Western Australia Child Health Survey, it appeared that there would be little difference in the design effect, and hence the sampling error, of a cluster size between 10 and 20. Larger cluster sizes had the advantage of being more likely to allow community level analysis, and also resulted in some reduction in data collection costs. Cluster sizes of around 20 meant that about 1 in 9 postcodes containing children from the target populations would be included. This cluster size also balanced the number of postcodes available for selection with the number of children in each postcode. A larger cluster size would have resulted in more postcodes not having sufficient population to select a full cluster sample, while a smaller cluster size would have resulted in the selection of a high proportion of postcodes, negating the cost savings of cluster sampling and compromising the ability to analyse community level influences on child outcomes.

After the sample selection process was simulated several times to examine the sorts of selected postcode distributions that could be expected, it was decided to halve the cluster size for Adelaide, Tasmania and the Australian Capital Territory, as it was felt that the number of postcodes selected using a full cluster was not sufficient to ensure a representative sample.

A separate postcode selection process was used for the Northern Territory. The process for Darwin was the same as for the rest of Australia. For the rest of the Northern Territory, once the remote areas had been excluded (as outlined earlier), the remaining postcodes were combined into 3 areas, and selections were made from each. The initial sample selected from the Northern Territory was proportionally larger than for the rest of Australia. This was required because of the higher proportion of post office box addresses and out-of-date addresses. After the first 2 phases of fieldwork[6], it was apparent that one of the exmet areas would hardly yield any final sample. It was therefore decided to further increase the sample selections made in the other two areas for phases 3 and 4.

[6] Sample selections and fieldwork were conducted in 4 phases nationally for both cohorts as described later in section 'Age range of children in the sample'.

Target population

Target population

Table 8 shows the Australian Bureau of Statistics June 2002 Estimated Resident Population estimates and the corresponding distribution of the potential samples for each cohort, if it was based on these estimates.

Table 8 ABS population estimates, June 2002 and corresponding regional distribution for each cohort

Population estimates, 30 June 2002						
	Children aged under 1 year			Children aged 4 years		
State/ territory	**Capital city**	**Rest of state**	**Total**	**Capital city**	**Rest of state**	**Total**
NSW	54,123	28,358	82,481	53,515	33,263	86,778
Vic	44,495	15,881	60,376	44,247	18,320	62,567
Qld	22,124	25,395	47,519	22,426	27,866	50,292
SA	12,555	4,902	17,457	13,150	5,431	18,581
WA	16,918	6,944	23,862	17,639	7,822	25,461
Tas	2,445	3,498	5,943	2,424	3,531	5,955
NT	1,710	1,967	3,677	1,769	1,707	3,476
ACT	4,018	5	4,023	4,273	4	4,277
Total	158,388	86,950	245,338	159,443	97,944	257,387
Corresponding distribution of sample						
	Children aged under 1 year			Children aged 4 years		
State/ territory	**Capital city**	**Rest of state**	**Total**	**Capital city**	**Rest of state**	**Total**
NSW	1,103	578	1,681	1,040	646	1,686
Vic	907	324	1,230	860	356	1,215
Qld	451	518	968	436	541	977
SA	256	100	356	255	106	361
WA	345	142	486	343	152	495
Tas	50	71	121	47	69	116
NT	35	40	75	34	33	68
ACT	82	0	82	83	0	83
Total	3,228	1,772	5,000	3,097	1,903	5,000

Note: these estimates have not been adjusted for the population that was not included in the LSAC target population due to the exclusion of particular remote postcodes. This adjustment was done during the sample selection and weighting processes.

Exclusions from sample selection

Earlier sections have dealt with the extent of coverage of the target population that was likely through the use of the Medicare enrolment database as the sampling frame, and the exclusion of children due to particular postcodes being excluded from the sample design.

Through almost all of the sample design development process, HIC had a requirement that only children with relatively recent[7] Medicare/ ACIR activity (ie children who had used a service for which a Medicare claim could be made, or who had had an immunisation registered with the ACIR) could be included. This process was required based on the rationale that it would minimise the chance of selecting a child who had died, but for whom an end-date had not yet been registered. However, this was likely to be an inaccurate method for excluding deaths and it meant that at least 10 per cent of 4 year old children would not have had a chance of selection, and most very young infants would also be excluded. The potential for bias through this requirement was quite large.

Just prior to the selection of the first phase sample, HIC replaced this requirement with one that involved a "fuzzy match" on name. This involved identifying from the Registrars' of Births, Deaths and Marriages 'fact-of-death' file, all children aged 5 years and under who had died in the last 5 years. Any child on the HIC database with the same recorded name (regardless of whether the age was the same) was then excluded from selection. An initial examination by HIC indicated that this would mean that 3 per cent of the selected sample would be withdrawn. Although this process excludes many more children than have actually died, it results in far less bias to the sample than the exclusion of children without Medicare activity.

The impact of including children without Medicare activity was not measured in the Dress Rehearsal[8] as only children with activity had been included. It is possible that there may be higher non-response from this population, due to address information being less up-to-date and also because proportionally more of these children are likely to be further away from main towns and may be harder for interviewers to contact.

In the short time available to determine whether the initial number of sample selections should be altered to account for the effects of fact-of-death matching, analysis of the available data indicated that the required final sample should still be achievable. Hence no adjustment to the initial sample was done at this stage.

Families with more than one target child

About 1.5 per cent of families with children in the target population have multiple births (i.e. twins, triplets etc) and about 5 per cent of families have both infants and children aged 4 years.

The scientific contribution resulting from examining the relatively small group of families containing siblings who were in the target populations was weighed against the respondent burden and operational complexities that would result from collecting data on multiple children within one family. Inclusion of two or more target children from one family would greatly increase respondent burden (and hence potential attrition) and time in the home. The operational complexities would also be considerable – survey instruments would have to be tailored for these situations and/or very clear instructions given to interviewers regarding what information needed to be collected for each child

[7] In the previous 6 months for the infants, and previous 12 months for the 4-5 year old children.

[8] The Dress Rehearsal for this study was conducted in August-October 2003 and as far as possible followed processes that were intended for the main wave of recruitment to the study. Further details are given in later sections of this report.

and what needed to only be asked once (i.e. family and household level data). In the case of multiple births, carers and teachers who are also being approached for information for this study would also have to complete forms for each child.

These processes would have required a reasonable amount of development (and processing) resources. In light of these concerns, it was decided to include only one child per family.

Age range of children in the sample and sample selection phases

Four phases of sample selection were undertaken. This process was guided by two principles regarding the desired ages and birth dates of the sample.

First, it was decided to include in the sample, children who are born in all months of the year. This was to ensure that the sample was representative of potential seasonal exposures that may impact on a child's development. For example, when a child is born may influence the child's susceptibility to certain health conditions (such as eczema or asthma), and the age that a child starts school (associated with their month of birth) may be an important variable in relation to a child's adjustment to school.

Second, it was also desirable to restrict the age range of children at interview. A restricted age range (particularly for the infant cohort), enabled the use of more age-targeted measures, and attempted to ensure that the majority of infants were aged at least 6 months old at the time of interview.

In addition, the phased approach ensured that the selected sample could be distributed into sensible workloads for interviewers and that the contact address information was as up-to-date as possible.

In order to meet the above requirements, four phases of selection were used, corresponding to four fieldwork periods. The details are indicated in Table 9. Due to the lead time required for the HIC mail-out/opt-out process and workload formation, the selection of children had to occur about two months prior to the start of fieldwork.

Table 9 Phases of sample selection in 2004

	Phase 1	Phase 2	Phase 3	Phase 4
Sample selection	16 January	27 February	2 April	21 May
Mail-out of letters	27 January	8 March	13 April	31 May
Data collection starts	15 March	26 April	31 May	19 July
Postcode areas	Group 1 (146)	Group 2 (147)	Group 1	Group 2
Dates of birth	March-August	March-August	September-February	September-February
Expected age range at interview	(4 years plus) 6-13 months	(4 years plus) 8-14 months	(4 years plus) 4-9 months	(4 years plus) 5-11 months

Note that Phase 1 and Phase 3 each covered the same group of postcode areas (146), and Phase 2 and Phase 4 covered the remaining postcode areas (147).

Also note that as a consequence of the sample selection process and the timing of the fieldwork, it was likely that about 700 (14 per cent) of the 4 year old cohort children (in New South Wales, Australian Capital Territory, Victoria and Western Australia) selected in the first 2 phases would have started school at the time of interview.

This design also involved selecting some children who were born just prior to the time of the phase 3 sample selection (that is, children born in February, and to a lesser extent, January 2004). It was recognised that a number of potential sample infants born in these months would not yet be enrolled in Medicare and would therefore be excluded from possible selection. However, it was decided to include children born in these months due to the importance of capturing potential season-of-birth effects, and the preference for the infant and 4 year cohorts to be exactly 4 years apart (as this facilitates inter-cohort comparison once the younger cohort reaches 4 years of age).

An advantage of the staged selection was that early data on recruitment to the study from the first phases was able to be used to make adjustments to the size of the sample selected in later phases. It was recognised that adjusting the sample in this manner leads to the need for different weights to be computed between the two sub-samples. However, the impact on the weights from this adjustment was judged to be minor compared to the consequences of a lower or higher than anticipated final sample size (insufficient sample might limit the analyses especially after the effects of attrition over several waves, and a larger sample may have added significantly to the cost). It was decided that it was preferable to make adjustments to the initial sample selections to ensure a final sample of close to 10,000 children rather than risk either insufficient or more than sufficient sample.

Process for initial contact with selected families

The selection of postcodes was undertaken by AIFS in late November 2003 and advice was sought from I-view on which postcodes would be included in Phase 1 (and 3) and which in Phase 2 (and 4).

HIC was provided with the list of selected postcodes by phase and the number of children to be selected in each postcode for each cohort and, at the designated sample extraction times, selected children of the appropriate ages.

HIC then mailed-out an 'invitation to participate' letter naming the selected child, with a letter of support from the Institute and a brochure on *Growing Up in Australia*, to the Medicare cardholder. If families did not wish to be involved with the study, they could either ring a 1800 'opt-out' telephone line staffed by the Health Insurance Commission or return a reply-paid form to the Commission.

Families had four weeks to register their withdrawal from the study. At the end of this period, cardholder name and address details for families (along with whether the child was an infant or a 4 year old) who had not contacted the Health Insurance Commission to withdraw from the study were given to I-view, the data collection agency. This list also excluded any families for whom the initial letter was returned to the Health Insurance Commission, due to a change of address or other reasons (possibilities for locating a current address in such circumstances were being explored, but none were

viable in the timeframe). Note: As part of the Phase 4 mail-out, any return-to-senders or families who I-view had found had moved were re-matched against the HIC database and a letter resent if a new address was found.

Once I-view were given a list of contact names and addresses, I-view then sent a letter to everyone on the list, indicating when an interviewer would be in the area and providing a 1800 number for queries. Families with post office box addresses or other non-street addresses were asked to contact the 1800 number with residential address details. This letter was then followed by a personal visit to the families by the interviewer.

Postcode selection

Data from the March 2003 HIC statistical extract had been obtained from HIC for designing the sample selection procedures and for selecting the Dress Rehearsal sample. It was decided to use these data for the selection of postcodes for the main study. As the postcodes were to be selected in advance of the children, it was not possible to use the actual data for the target populations (children born March 1999 - February 2000 and March 2003 - February 2004) for the postcode selection, since some of the children would not be born at the time the postcodes needed to be selected.

If it had been possible to use the actual data for our target populations in the selection of postcodes, this would have ensured a self-weighting sample where the design weights for the selected children were the same (i.e. children have equal probability of selection). With approximately equal probabilities of selection, the sample would have been more robust, less likely to have problems with influential sample points or outliers, and would have maximised the contribution of each child to the overall analysis. As using data on the actual target population was not possible, analyses were undertaken to see how much effect there would be on the design weights by using data that was a year old, compared with obtaining more recent data (for example, obtaining an extract from HIC late in 2003 and using that).

Results from the analysis showed that there would be minimal impact on the variability of the weights and ultimately the sampling errors obtained. Therefore the decision was made to use the data already obtained from HIC.

Within strata, postcode areas were divided into two groups as shown in Table 10: Group A, those with sufficient children to ensure the (final) cluster size of 20 (or 10, for Adelaide, Tasmania and the ACT) per cohort, and Group B, those with less. The cut-off used during the sample selection process was 80 (or 40) 'active' infants and 4 year old children in the sample (the reason that 'active' children were used was because at that time it was required that only children with Medicare activity would be eligible for selection).

Table 10 Distribution of postcode areas eligible for selection

	Met	ACT	NSW	NT	QLD	SA	TAS	VIC	WA	Australia
Group A:	Met	21	232	8	114	108	27	202	82	794
80+ active(b) children	Exmet	0	168	3	117	34	38	95	32	487
	Total	21	400	10	231	142	65	297	114	1281
Group B:	Met	0	0	0	0	0	0	43	21	64
20-79 active children	Exmet	0	97	0	58	46	0	129	36	366
	Total	0	97	0	58	46	0	172	57	430
Total		21	497	11	289	188	65	469	171	1711

Due to the variation in the distribution of children across postcodes, and the intention to have (as far as possible) an equal probability of selection for all children, postcodes in Group A were selected on a 'probability proportional to size' (pps) basis and then a fixed number (cluster) of children were selected within the postcode. Postcodes in Group B were selected on an equal probability basis, and then all children within the

postcode were selected. The selection process ensured that children in Group A and Group B postcodes all had about the same probability of selection.

The size variable used for the pps selection was the sum of the number of children in each cohort in the postcode. Although in most cases the number of infants and 4 year olds was correlated, in some cases there were large discrepancies. However, an analysis of how this would affect the volatility of the weights was undertaken, and it was shown to have minimal impact. In addition, there was no other obvious size variable to use.

A number of simulations of the postcode selection process were undertaken. In these, we tested ordering postcodes by the main SEIFA (the Australian Bureau of Statistics' Socio-Economic Indices for Areas) index of disadvantage (within met/exmet strata) and also ordering the postcodes by number. Repeated tests found that the numerical ordering gave both a good distribution across SEIFA values and a good geographic distribution; whereas ordering by SEIFA produced a less adequate geographic distribution. Hence it was decided to order postcodes by number within stratum.

The selection process was undertaken using SAS PROC SURVEYSELECT, which can perform both systematic pps sampling and equal probability sampling. Postcodes with a very large number of children, which had a probability greater than 1 of selection, were split into two discrete units each with assigned a probability less than 1. This involved 7 postcodes.

Table 11 shows the distribution of postcode areas selected.

Table 11 Distribution of postcode areas selected

	Met	ACT	NSW	NT	QLD	SA	TAS	VIC	WA	Australia
Group A:	Met	8	54	3	22	26	5	43	17	178
80+ active(b) children	Exmet	0	28	3	24	3	7	13	5	83
	Total	8	82	6	46	29	12	56	22	260
Group B:	Met	0	0	0	0	0	0	2	2	4
20-79 active children	Exmet	0	6	0	5	4	0	8	5	28
	Total	0	6	0	5	4	0	10	7	32
Total		8	88	6	51	33	12	66	29	293

As indicated earlier, the number of children to be selected within each stratum was based on Australian Bureau of Statistics' (ABS) population estimates of the distribution of infants and children aged 4 years in each stratum, rather than the number of children on the Medicare frame. ABS figures were used for sample allocation, and HIC figures for actual probabilities of selection because the ABS figures were likely to be more accurate, so provided a better basis for determining the sample allocation between states, but the sample obviously had to be selected from the HIC frame. The use of a sum of the 0 and 4 year old cohorts as a measure of size at the first stage was an obvious consequence of wanting to select the same postcodes for both cohorts, which was done for both operational and analytical purposes.

Selection of children

Initial sample selection

The number and distribution of postcodes were determined based on the likely distribution of the infant plus 4 year old populations and the cluster sizes required for each postcode. The next decision was how many children to initially select in each of the postcodes to achieve the required cluster sizes.

Sample loss data from the Dress Rehearsal were analysed to determine the likely sample loss for the main wave. Tables 12 and 13 shows the results from the Dress Rehearsal.

Table 12 Dress rehearsal sample loss

	Victoria	Sydney	Rural/remote NSW/Qld	Total
Interviews	55.6	37.5	53.0	50.7
HIC opt-outs/ RTS	14.5	13.7	12.1	13.8
Other refusals	16.2	14.1	13.3	15.1
Non-contacts	17.3	33.9	21.1	22.1
Total sample	100.0	100.0	100.0	100.0

Table 13 Distribution of infant and 4 year old interviews

	Victoria	Sydney	Rural/remote NSW/Qld	Total
Infant interviews	152	61	67	280
Child interviews	151	33	63	247

For the Dress Rehearsal, only the postcodes in Victoria were selected at random. The other postcodes were chosen because they presented challenges for data collection (eg inner city, high proportion of English as a second language families, remote areas) and the recruitment rates to the study were expected to be lower. In addition, the number of selections with post office box addresses was boosted, in order to provide sufficient selections in the Dress Rehearsal to determine the likely response rate from these contact addresses.

Detailed analysis of the data was undertaken, comparing socio-demographic data of the Dress Rehearsal samples with Census data of children of the same age, at postcode level. This analysis revealed it was likely that there would be differential sample loss between strata. However, it was decided that there was not enough information to predict these differences in advance, so it was decided to not vary the sampling fractions across strata for the main study.

However, it was clear from the Dress Rehearsal data that the sample loss for the 4 year old cohort was likely to be higher than for the infant cohort for all strata. For the main study, it was therefore decided to select slightly more 4 year olds than infants.

At the time that decisions were being made about the number of children to select, it was expected that the recruitment rates for the main study would be higher than those obtained for the dress Rehearsal, for a number of reasons (for example, more publicity about the study, smaller workloads for interviewers, lower proportion of post office box addresses). The expectation was that an overall recruitment rate of at least 55 per cent of children selected by HIC would be achieved.

With a projected 55% recruitment rate, for Group A postcodes, it was estimated that 18 infants and 19 children aged 4 years would be selected for the postcodes with a final cluster size of 10 per phase, and 9 infants and 10 children aged 4 years would be selected for those with cluster size of 5 per phase. For Group B postcodes (those postcodes likely to have less than a full size cluster of target children), all appropriate children would be selected.

HIC selection of children

Within each selected postcode, the children were listed in date of birth order, with any children from the same multiple birth grouped together in a random order. The required number of children were selected by taking a random start and then applying a skip interval through the list. This ensured that children across a range of birth months were selected.

The sample selection was undertaken first for the infants and only one infant on any given Medicare card retained (so that only one child was selected from any family). It was decided that 'de-selecting' children this way was the best approach overall to selecting only one child if there was more than one eligible child on the same card.

An alternative approach that was considered involved randomly selecting only one child per Medicare card for inclusion on the sampling frame BEFORE selection of individual children. This in effect would mean that families are being selected, rather than children. It would mean that the sample obtained was representative of families of target children, rather than of all target children, which would result in unequal selection probabilities, and hence sampling weights, for children. This would impact on the overall level of accuracy achieved by the sample.

Given that the child was defined by FaCS as the sampling unit, it was decided that the number of children in the sample from multiple births should reflect the number of children in the population from multiple births. This was felt to be a more important issue than the very slight sample bias introduced by the preferred selection process described above.

Once the infants were selected, all 4 year old children on the same Medicare card as the selected infants were excluded from the 4 year old population. The selection process was then repeated for the 4 year old cohort. This process does introduce a slight selection bias but the alternatives were not felt to be preferable.

For example, we considered using a conditional selection methodology that would provide a more methodologically rigorous approach for selecting the 4 year olds. Conditional selection methodologies adjust the selection probabilities to account for previous selection mechanisms, which is considered to be more methodologically sound than just removing children from the sample, and which would still ensure that a 4-year-old child could only be selected if no other child in the family (or Medicare Card) has been selected. Methods such as Karmel Selection allow for this by adjusting the probability of selection based on the probability of selection of the first cohort. The disadvantages of using a conditional selection methodology includes the complexities it would have introduced both to the sample design and the programming, and it would also have resulted in unequal probabilities of selection for children, which may have impacted on the overall efficiency of the sample. Weighing the costs of this approach (complexity and a potential decrease in accuracy of estimates from the study) against the benefits (a possible slight reduction in bias), it was judged that the simple method of excluding 4 year olds on the same Medicare card as infants already selected, was the better approach.

Results of sample selection

Information on the numbers of children selected and recruited into the study are included at the end of this section of this report and full details are given in the report on weighting and non-response (Technical paper no. 2).

Response after 2-3 weeks of Phase 1 fieldwork indicated that the target number of children may not be recruited. This was due to the loss of initial sampling pool due to the fact-of-death matching combined with a higher proportion of out-of-date addresses than expected, and a higher refusal rate to the HIC enquiry line than was found in the Dress Rehearsal.

There was only a small window of opportunity to adjust the Phase 3 number of selections, based on this incomplete Phase 1 field data. Analysis of these data, combined with analysis of flow data from the Dress Rehearsal (eg looking at what proportion of interviews/ appointments were made in the same time frame), plus insights from the data collection agency, indicated that an increase in sample selections was likely to be required. In effect, this was principally to adjust for the sample selections that had been excluded through fact-of-death matching.

In addition, it appeared that there was higher than expected sample loss from the child cohort, whereas the infant recruitment was on track. However, given the preliminary nature of the information available at this time, it was decided to adjust both cohorts in the same way. The action taken was to increase the number of selected children in each cohort by 1 in all Phase 3 postcodes where this was possible[9].

It should be noted that while the fact-of-death matching excluded more selections than had been expected, methodologically, it was a significant improvement over the inactivity exclusion previously proposed by HIC. Although the decision to change the sample selection requirements happened very late in the day and meant small changes to

[9] For Group B postcodes, the number of selected children could not be increased, as all appropriate children were already being selected.

the intended sample sizes across phases, the impact on the representativeness of the final sample should be less than otherwise would have been the case.

Calculation of the required adjustment to the phase 4 sample were based on data from the almost complete Phase 1 plus 2-3 weeks of Phase 2. These data indicated that the infant sample was being recruited at the originally expected rate, but the 4 year old cohort recruitment was significantly lower than expected. It was decided that no increase in sample selections was needed for the infant cohort in Phase 4, but that substantial additional sample selections were required for the 4 year old cohort.

It was still difficult to estimate the number of additional sample that was needed to help ensure that 5,000 children from the 4 year old would be recruited. Various options were considered. The final decision was to increase the number of 4 year old children selected in Phase 4 postcodes by 4, and to select an additional 3 children in the Phase 3 capital city postcodes in the larger states (the restriction to capital city postcodes was so that the travel to these areas would not incur additional costs), and by 2 and 1 respectively in the smaller states. A larger increase was applied to the Northern Territory postcodes.

It was decided that increasing the sample at the later phases would be done by selecting more children from the population being used for phases 3 and 4, rather than also increasing the sample from the population that was used for phases 1 and 2 (which would have been preferable in terms of maintaining consistent weights). This was decided because only small numbers of additional children needed to be selected per postcode, and the complexity of the sample selection process would have been increased substantially for HIC in a very tight timeframe if the additional sample was to be split between the 2 sub-populations. In addition, most of the children from the older population would have turned one or 5 years old by the time of phase 3 (and 4) interviewing (and some would be almost 18 months or 5 ½ years).

Table 14 Final Wave 1 response

	Infant		Child	
	No.	Per cent	No.	Per cent
Mail-out sample	8921	100.0%	9893	100.0%
Recruited sample	**5107**	**57.2%**	**4983**	**50.4%**
Sample loss:				
Refusals:	2909	32.6%	3710	37.5%
HIC opt-out	*1398*	*15.7%*	*1587*	*16.0%*
I-view 1800 line	*335*	*3.8%*	*398*	*4.0%*
Interviewer	*1111*	*12.5%*	*1423*	*14.4%*
Non-contacts	970	10.9%	1502	15.2%

The official response rate for this study, as shown in the above table, is calculated as the proportion of families who were recruited to the study out of those who were sent a letter by HIC. Another way of calculating the response rate is to exclude the non-contacts from the calculation, as these families did not have an opportunity to actively say whether or not they wanted to be in the study. This response rate is 64.2% for the infants and 59.4% for the 4 year olds.

Summary

The following table summarises the key features of the sample design.

Reference population - scope	Cohort 1 Infant: Australian children born between March 2003 and February 2004. Cohort 2 Child: Australian children born between March 1999 and February 2000.
Sample size	5,000 for each cohort intended; final sample recruited was 5,107 infant and 4,983 child cohort.
Sampling frame	Medicare enrolment database
Coverage	Slight over-coverage of child cohort according to comparison with ABS population estimates (101.5%); under-coverage of infant cohort (88.5%) compared with ABS, due to late registration of births with Medicare.
Exclusions	The sample is broadly representative of all Australian children, with the exception of children living in remote areas. About 40 per cent of children in remote areas were not given a chance of selection, and these children will not be estimated for in the population estimates produced. Other children who were not given a chance of selection, or who were excluded after selection, are not considered to impact on the representativeness of the sample, and therefore are included in the population estimates. These include: • Children living in postcodes with less than 6 of the target population, where the postcode was not amalgamated with an adjacent postcode (about 1-2 per cent of children in each cohort); • Children in families where another child in the family was selected for the main sample (less than 1 per cent); • Children with the same (or similar) name to a child (born after 1 March 1999) listed on the National Death Index (approximately 3 per cent in each cohort).
Primary sample units	Postcodes, or groups of postcodes – stratified by state/territory and by capital city statistical division/rest of state ('met'/ 'exmet') for the purposes of production of estimates. Postcodes were further stratified into 2 groups by size – postcodes that had at least the minimum number of children that needed to be selected (about 75 per cent of postcodes), and postcodes that had less than that amount.
Selection process	A random selection of a number of postcodes then a random selection of a number of in-scope children within each selected postcode. For postcodes with at least the minimum number of children, probability proportional to size selection was used; for the other postcodes, equal probability selection was used.

	Postcodes were selected by AIFS and then divided into 2 groups by I-view for data collection purposes. Children were selected by HIC in 4 phases, corresponding to the 4 data collection phases. Children born March – August were selected for the first 2 phases, and children born September – February were selected for the second 2 phases. For postcodes with at least the minimum sample size, the children were listed in date of birth order within a postcode, with any children from the same multiple birth grouped together. The required number of children was selected by taking a random start and then applying a skip interval through the list. For postcodes with less than the minimum number of children, all children were selected. The sample selection for the infants was undertaken first.
Cluster size	Over the 4 phases of sample selection, for postcodes with more than the minimum number of children, in NSW, Vic, Qld, WA and exmet SA 73 infants and 86 4-5 year olds were initially selected, and in met SA, Tas and ACT 37 infants and 44 4-5 year olds were selected.
Families with more than one target child	Only one child per family was included in the sample selection process. This includes cases of multiple birth, as well as families with both an infant and a 4-5 year old.
Children on multiple cards	Children were selected on the basis of the card where there was the most recent Medicare activity. If no recent activity, the card with an adult female was selected. If no adult female on either card, then the primary card was selected.

References

Hanna, B. (2002). 'An alternative way to estimate population in local areas: The second phase of the Medicare Project', paper presented to the APA conference, 2-4 October, 2002

Hull, B., Lawrence, G., MacIntyre, C.R. & MacIntyre, P. (2001). 'Immunisation Coverage: Australia 2001' National Centre for Immunisation Research and Surveillance of Vaccine Preventable Diseases, University of Sydney.

National Longitudinal Survey of Children and Youth (NLSCY) (1999). *Overview of survey instruments for 1998-99 data collection cycle 3.* (Catalogue no. 89FOO78XPE, no. 3). Canada: Statistics Canada.

Nicholson, J.M., Sanson, A., and the LSAC Research Consortium (2003). A new longitudinal study of the health and wellbeing of Australian children: How will it help? (commentary) *Medical Journal of Australia, 178 (6),* 282-284.

Sanson, A., Nicholson, J., Ungerer, J., Zubrick, S., Wilson, K., and the LSAC Consortium Advisory Group (2002). 'Introducing the Longitudinal Study of Australian Children', LSAC Discussion Paper No 1. Australian Institute of Family Studies, Melbourne.

Soloff, C., Misson, S., Lawrence, D., Johnstone, R. and Slater, J. 'Wave 1 weighting and non-response', LSAC Technical paper No. 2. . Australian Institute of Family Studies, Melbourne.