THE LONGITUDINAL STUDY OF AUSTRALIAN CHILDREN: AN AUSTRALIAN GOVERNMENT INITIATIVE

Data user guide

December 2018
Abbreviations

ABS - Australian Bureau of Statistics
ACARA - Australian Curriculum, Assessment and Reporting Authority
ACASI - Audio Computer Assisted Self Interview
ACIR - Australian Childhood Immunisation Register
AEDC - Australian Early Development Census
AIFS - Australian Institute of Family Studies
ANU4 - Australian National University ranking of occupational prestige, 4th edition
ASCL - Australian Standard Classification of Languages
ANZSCO - Australian and New Zealand Standard Classification of Occupations
ASGC - Australian Standard Geographic Classification
ATSI - Aboriginal and Torres Strait Islander
BMI - Body Mass Index
BP - Study Child Blood Pressure
CA - Carer Allowance
CAI - Computer-Assisted Interview
CAPI - Computer-Assisted Personal Interview
CASI - Computer-Assisted Self-Interview
CATI - Computer-Assisted Telephone Interview
CBC - Centre-Based Carer
CCB - Childcare Benefit
CSR - Child Self-Report
DFRDB - Defence Forces Retirement and Death Benefits Scheme
DSP - Disability Support Pension
DSS - Department of Social Services
DVA - Australian Government Department of Veterans’ Affairs
EHC - Event History Calendar
EXEC CogState - Executive functioning
F2F - Parent 1 Face-to-Face Interview
FCF - Family Contact Form
FDC - Family Day Care
FDCQA - Family Day Care Quality Assurance
FTB - Family Tax Benefit
FTBA - Family Tax Benefit A
FTBB - Family Tax Benefit B
GJT/SLI - Rice Test of Grammaticality Judgement
GPS - Global Positioning System
HBC - Home-Based Carer
IOBS - Interviewer Observations
ISP - Income Support Payments
LDC - Long Day Care
LOTE - Language Other Than English
LSAC - Longitudinal Study of Australian Children
MBS - Medicare Benefit Scheme
MSN - Medicare Safety Net
MR - Matrix Reasoning test
NAPLAN - National Assessment Program—Literacy and Numeracy
NCAC - National Childcare Accreditation Council
OSHCQA - Outside School Hours Care Quality Accreditation
PID - Parent 1 During Interview Questionnaire
PIL - Parent 1 Leave-Behind Questionnaire
P2L - Parent 2 Self-Complete Questionnaire
PBS - Pharmaceutical Benefit Scheme
PLE - Parent Living Elsewhere
PM - Physical Measurements
PPVT - Peabody Picture Vocabulary Test
PPVT-III - Peabody Picture Vocabulary Test, 3rd Edition
QIAS - Quality Improvement and Accreditation System (for Long Day Care centres)
RAP - Study Child (SC) living away from parents, parents of the SC RAP known as Parent 1 RAP, Parent 2 RAP and PLE RAP
RPBS - Repatriation Schedule of Pharmaceutical Benefits
ROC - Receiver Operating Characteristic
RSE - Relative Standard Error
SACC - Standard Australian Classification of Countries
SEIFA - Socio-Economic Indexes for Areas
SLI - Specific Language Improvement
SRS - Simple Random Sample
TIS - Telephone Interpreter Service
TQ - Teacher Questionnaire
TUD - Time Use Diary
WAI - Who Am I?
WISC - Wechsler Intelligence Scale for Children
Acknowledgements and citation

The current version of the LSAC Data User Guide has been updated by AIFS.
The Wave 7 data files were prepared by the ABS and AIFS.

Readers wishing to refer to this document should cite the following:
This document is intended to provide the necessary information to be able to use the LSAC data, and includes the survey methodology, file structure and variable naming conventions. Particular issues are highlighted to ensure data analysts apply the LSAC data appropriately in their research.

The following documentation is also useful to data users and is available on the study website (growingupinaustralia.gov.au/data-and-documentation):

- questionnaires and interview specifications marked with variable names for Computer Assisted Interview (CAI), Computer Assisted Self Interview (CASI) and Computer Assisted Telephone Interviews (CATI) instruments
- a data dictionary
- technical papers on weighting, non-response and other issues
- data issues papers
- item rationale papers

Data users should read the ‘Important issues for data analysis’ section carefully. It outlines particular aspects of the sample design that have important implications for interpreting analyses from the study.

Further information on data usage is contained in the Data users information pages on the LSAC website (growingupinaustralia.gov.au/data-and-documentation).

Feedback about this data user guide is welcome. If there is something that you would find useful that is currently not included, or if you had difficulty understanding any of the guide’s content, please let us know by emailing us at: aifs-lsac@aifs.gov.au.

---

1 Feedback from data users suggests that marked questionnaires with interview specifications are often the best way to find sections relevant to proposed research topics, and to illustrate the breadth of information available in the study.
2 What is LSAC?

Growing Up in Australia: The Longitudinal Study of Australian Children (LSAC) continues to examine the impact of Australia’s unique social and cultural environment on the next generation.

The study tracks children’s development and life course trajectories in today’s economic, social and political environment. A major aim of the project is to identify policy opportunities for improving support for children and their families, and identifying the opportunities for early intervention.

The study investigates the effect of children’s social, economic and cultural environments on their wellbeing over the life course.

2.1 Objectives

LSAC has a broad multi-disciplinary base and examines policy-relevant questions about development and wellbeing. The research questions span parenting, family relationships, education, child care, employment and health.

The study’s longitudinal structure enables researchers to determine critical periods for providing services and welfare support, and to identify long-term consequences of policy innovations (for more details see LSAC Discussion Paper No.1, Introducing the Longitudinal Study of Australian Children).

The study is the first ever comprehensive, national Australian data collection on children as they grow up.

2.2 Who is involved?

LSAC is undertaken in partnership between the Department of Social Services (DSS), the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS), with advice provided by a consortium of leading researchers known as the LSAC Consortium Advisory Group (CAG).

The Wave 1 data collection was undertaken for AIFS by private social research companies i.e., Colmar-Brunton Social Research and I-view/NCS Pearson. Data collection for Waves 2–7 was undertaken by ABS.

2.3 Timelines

Development for the study commenced in March 2002 with a testing phase involving over 500 families that continued through 2003. Recruitment for the main study took place between March and November 2004, and over 10,000 children and their families agreed to participate. From 2004, participating families have been interviewed every two years, and between-wave mail-out questionnaires were sent to families in 2005 (Wave 1.5), 2007 (Wave 2.5) and 2009 (Wave 3.5). Additional between-wave questionnaires (Waves 4.5 and 5.5) were undertaken via online web forms from 2009 for the purposes of updating the contact details of study participants. In 2015-16, B cohort study children were invited to participate – along with one of their parents – in a comprehensive clinic appointment or shorter home visit for a comprehensive, one-off physical health and biomarker module, known as the Child Health CheckPoint (between Waves 6 and 7).

2.4 Sample design

The focus of the study is on the developmental pathways of two cohorts of Australian children, so the study child is the sampling unit of interest. A dual cohort cross-sequential design was adopted as shown in Figure 1.
Two cohorts of children were selected from children born within two 12-month periods:

- B cohort: children born March 2003–February 2004

Further information about the design of the sample is available in the ‘Survey methodology’ section of this guide, and in LSAC Technical Paper No. 1, Sample Design (available from the study website growingupinaustralia.gov.au/data-and-documentation).

### 2.5 Study informants

The study collects data from multiple informants:

- Study child is the cohort child.
- Study child RAP (SC RAP) is the respondent who is living away from the parental home (in Wave 7 only applicable to K cohort children).
- Parent 1 (P1) is defined as the parent who knows the study child best; in most cases this is the child’s biological mother.
- Parent 2 (P2) is Parent 1’s partner or another adult in the home with a parental relationship to the study child; in most cases this is the biological father, but step-fathers are also common.
- Parent living elsewhere (PLE) is a parent who does not live with the study child; most commonly the biological father after separating from the biological mother. This collection was started in Wave 2.
- Teachers and child care workers involved with the study child.

In addition, LSAC data are linked to the file from the National Childcare Accreditation Council, Medicare Australia, ABS Census, the National Assessment Program – Literacy and Numeracy (NAPLAN), Australian Early Development Census (AEDC) and Centrelink.

### 2.6 Mother/Father data

While P1 is usually the mother and P2 is usually the father, this is not always the case. However, many users prefer to analyse the data by parent gender (i.e. mother and father rather than P1 and P2). Therefore, all the variables collected for both P1 and P2 are also presented as mother and father variables. Note that P1 and P2 may be the guardians of the child and not the child’s biological parents. In this context, mother should be taken to mean ‘female parent/guardian’. Sometimes P1 (and/or P2) might change between waves. For instance, P1 may be reported as female across subsequent waves, although the parent may, in fact, be different people.

If there are two female parents, P1 is coded as Mother and P2 is coded as Father. This will be maintained if the parents swap between P1 and P2 in subsequent waves. This means that there are a small number of female fathers that analysts should be mindful of when working with these variables. In addition, data users can use the sex variable to identify these if needed.

The majority of study child respondents live with their families. However in Wave 7 for the first time there were cases where the study child respondent lived outside the parental home. In these cases the study child respondent is defined as the study child respondent away from parents (RAP). The parents of the study child RAP are known as P1 RAP, P2 RAP and PLE RAP and their information is presented in main wave data files.
3 Instruments

Table 1 summarises the data collection instruments used in each wave.

The following methods are used to collect study data.

- The face-to-face interview (F2F) is conducted with P1 and the study child (although, in Wave 1, P2 could complete some sections if this was more convenient). This component is undertaken with all participating families at each wave. Some interviews might be completed over the telephone in full; for example, with participating families in remote areas (see section 11.3.7).

- The P1 during interview questionnaire (PID) consisted of self-complete items for which it was considered important to achieve high response rates. In Wave 4 it became a computer-assisted self-interview (CASI).

- The P1 leave-behind questionnaire (P1L) consisted of lower priority self-complete items. Efforts are made to obtain this data from P1 while the interviewer is in the home. This form became part of the CASI.

- The P2 leave-behind questionnaire (P2L) consists of self-complete items. Efforts are made to obtain this data from Parent 2 while the interviewer is in the home. If this is not possible the questionnaire is left for completion at a later time.

- Child self-report interview (CSR) consists of items answered by the study child. For children younger than 10 years old it was administered by an interviewer. For children aged 10–15 years (K cohort, Waves 4, 5 & 6 and B cohort, Wave 6) it was administered via an audio computer-assisted self-interview (ACASI) and from 16 to 17 years (K cohort, Wave 7) by a computer-assisted self-interview (CASI). As part of the interview, physical measurements are taken and other assessments (such as measures of cognition or achievement) are administered to the study child.

- The study child completes an audio computer-assisted self-interview (ACASI) or a computer-assisted interview (CASI) by themselves. This method allows sensitive content to be answered by the child in total anonymity.

- The time use diary (TUD) documents a 24-hour period of the child’s life. In Waves 1, 2 and 3, the child’s family were asked to complete two TUDs, one for a week day and one for a weekend day. A different procedure was implemented in Wave 4. In Wave 4, the study child (K cohort only) was asked to complete one TUD. A TUD form with instructions on how and when to fill it in was sent to the study child prior to the interview. The study child was asked to fill in the TUD form on the day before the interview date. The next day, during the interview, the interviewer asked the child to describe ‘yesterday’ using the TUD form. The day the diary referred to could be any day of the week depending on when the interview was scheduled.

- The parent living elsewhere questionnaire (PLE) was first included in Wave 2 as a mail-back questionnaire. In Wave 3 it became a computer-assisted telephone interview (CATI).

- The RAP study child is the study child respondent living away from parents (from Wave 7 for K cohort). Study child (RAP) and P1 (RAP) both complete home interviews in their own separate homes. P2 (RAP) and Parent PLE (RAP) instruments are still administered in the same way for RAP study child’s parents as for other participants.

- The home-based carer questionnaire (HBC) is for children aged 0–1 and 2–3 years who receive child care in a home environment, most commonly from a grandparent.

- The centre-based carer questionnaire (CBC) is for children aged 0–1 and 2–3 years who receive child care from long day care programs in centres, schools, occasional care programs, multi-purpose centres and other arrangements.

- The teacher questionnaire (TQ) is for children aged 4–5 years and older who attend a school or, for some 4–5 year olds, a preschool or long day care centre.
# Table 1: Data collection modes by wave

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Mode</th>
<th>Completed by</th>
<th>Indicator variable</th>
<th>W1</th>
<th>W2</th>
<th>W3</th>
<th>W4</th>
<th>W5</th>
<th>W6</th>
<th>W7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face-to-face interview (F2F)</td>
<td>Paper</td>
<td>Parent 1</td>
<td>N/A</td>
<td>BK</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Face-to-face interview (F2F) Computer Assisted Interview – (CAI)</td>
<td>Computer</td>
<td>Parent 1</td>
<td>N/A</td>
<td></td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
</tr>
<tr>
<td>Parent 1 during interview (PID)</td>
<td>Paper</td>
<td>Parent 1</td>
<td>[*]p1dd</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
</tr>
<tr>
<td>Parent 1 during interview (CASI)</td>
<td>Computer</td>
<td>Parent 1</td>
<td>[*]p1dd</td>
<td>-</td>
<td>-</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
</tr>
<tr>
<td>Parent 1 leave behind (PIL)</td>
<td>Paper</td>
<td>Parent 1</td>
<td>[*]p1scd</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
</tr>
<tr>
<td>Parent 2 leave behind (P2L)</td>
<td>Paper</td>
<td>Parent 2</td>
<td>[*]p2scd</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
</tr>
<tr>
<td>Child self-report (CSR)</td>
<td>Computer</td>
<td>Study child</td>
<td>[*]csrd</td>
<td>K</td>
<td>K</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Audio computer-assisted self-interview (ACASI)</td>
<td>Computer</td>
<td>Study child</td>
<td>Need consent from: PI [<em>]id40e &amp; SC [</em>]id40f</td>
<td>-</td>
<td>-</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>B</td>
</tr>
<tr>
<td>Computer assisted self-interview (CASI)</td>
<td>Computer</td>
<td>Study child</td>
<td>Need consent from: PI [<em>]id40e &amp; SC [</em>]id40f</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>K</td>
</tr>
<tr>
<td>Time use diary (TUD)</td>
<td>Paper</td>
<td>Parent 1</td>
<td>N/A</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Time use diary (TUD)</td>
<td>Computer</td>
<td>Study child</td>
<td>Need consent from: PI [<em>]id40i &amp; SC [</em>]id40</td>
<td>-</td>
<td>-</td>
<td>K</td>
<td>K</td>
<td>BK</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Parent living elsewhere (PLE)</td>
<td>Paper – mailed out</td>
<td>PLE</td>
<td>[*]plescd</td>
<td>-</td>
<td>BK</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Parent living elsewhere (PLE CATI)</td>
<td>Computer/Telephone</td>
<td>PLE</td>
<td>[*]plescd</td>
<td>-</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
</tr>
<tr>
<td>Home-based carer (HBC)</td>
<td>Paper</td>
<td>Carer</td>
<td>[*]hbccbc</td>
<td>B</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Centre-based carer (CBC)</td>
<td>Paper</td>
<td>Carer</td>
<td>[*]hbccbc</td>
<td>B</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Teacher questionnaire (TQ)</td>
<td>Paper</td>
<td>Teacher</td>
<td>[*]tcd</td>
<td>K</td>
<td>K</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>B</td>
</tr>
<tr>
<td>Physical measurements (PM)</td>
<td>Computer</td>
<td>Study child</td>
<td>Need consent from: PI [<em>]id30d &amp; SC [</em>]id30e</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
</tr>
<tr>
<td>Who am I? (WAI)</td>
<td>Computer</td>
<td>Study child</td>
<td>cid44a1</td>
<td>K</td>
<td>-</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PPVT assessment (PPVT)</td>
<td>Computer</td>
<td>Study child</td>
<td>[*]ppvtd</td>
<td>K</td>
<td>K</td>
<td>BK</td>
<td>B</td>
<td>B</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Matrix reasoning (MR)</td>
<td>Computer</td>
<td>Study child</td>
<td>[*]id44a1</td>
<td>-</td>
<td>K</td>
<td>K</td>
<td>BK</td>
<td>B</td>
<td>B</td>
<td>-</td>
</tr>
<tr>
<td>Study child blood pressure (BP)</td>
<td>Computer</td>
<td>Study child</td>
<td>Need consent from: PI [<em>]id47a &amp; SC [</em>]id47b</td>
<td>-</td>
<td>-</td>
<td>K</td>
<td>K</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Interviewer observations (IOBS)</td>
<td>Computer</td>
<td>Interviewer</td>
<td></td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
<td>BK</td>
</tr>
<tr>
<td>Executive functioning (EXEC/CogSTATE)</td>
<td>Computer</td>
<td>Study Child</td>
<td>[*]id40m</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>K</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Event history calendar (EHC)</td>
<td>Computer</td>
<td>Study Child</td>
<td>EHCEK16 – employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes: The indicator variable can be used to see if data is present or not for a particular instrument (for more information see sections 8.6 &amp; 8.7). The [*] in the indicator variable should be replaced by the age indicator (a, c, d, e, f, g, h, i) as discussed below. In-between waves were administered using mail out surveys for Waves 1.5, 2.5 and 3.5. Waves 4.5 and 5.5 used online web forms to update contact details.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• Interviewers make observations (IOBS) with permission of the respondent about the interview, state of the house (where the interview was conducted) and the neighbourhood characteristics of where the respondent lives.

• In Wave 1 the Australian Early Development Census (AEDC) was included as a nested study, which involved the AEDC questionnaire being sent with the LSAC K cohort teacher questionnaire in Victoria, Queensland and Western Australia. The AEDC is a community-level measure of young children’s development based on a teacher-completed checklist. It consists of over 100 questions measuring five developmental domains: language and cognitive skills; emotional maturity; physical health and wellbeing; communication skills and general knowledge; and social competence. For more information see aedc.gov.au

• The family contact form (FCF) recorded information about any contact between the interviewer and the family of each of the selected children at the time of Wave 1, regardless of whether they agreed to participate in the study or not. The information was mainly used by the fieldwork agency, with the only information from the FCF available in the publicly released dataset being the information on the family’s home and neighbourhood. In subsequent waves, this information was included as part of the interviewer observations of the face-to-face interview.

• Between-wave questionnaires (Wave 1.5, Wave 2.5 and Wave 3.5) are brief questionnaires sent to respondents to complete and return in the year between main waves of data collection. Between-wave surveys help to maintain contact with study families and collect information about activities and development in the year between the main waves. For Waves 4.5 and 5.5, online web forms were used to update contact details of study participants.

3.1 Child assessments

3.1.1 Physical measurements

Weight
For the B cohort in Wave 1, the child’s weight was obtained by calculating the difference between the weight of Parent 1 (or another adult) with the child and the weight of the parent/other adult on their own. For the B cohort at all subsequent waves, and the K cohort at all waves, the child’s weight was measured directly.

In Wave 1 the scales used were Salter Australia glass bathroom scales (150 kg x 50 g). In Waves 2 and 3, these scales were used along with HoMedics digital BMI bathroom scales (180 kg x 100 g). In Waves 4, 5, 6 and 7, Tanita body fat scales were used.

Height
Height is measured for children aged two years and older. In Waves 1, 2 and 3, height was measured using an Invicta stadiometer, from Modern Teaching Aids. In Waves 4, 5, 6 and 7 a laser stadiometer was used. Two measurements were taken, and if the two measurements differed by 0.5 cm or more, a third measurement was taken. The average of the two closest measures was included on the data file.

Girth
This measurement is taken for children aged two years and older using a non-stretch dressmaker’s tape, positioning the tape horizontally over the navel. In all waves, two measurements were taken, and if these differed by 0.5 cm or more, a third measurement was taken. The average of the two closest measures was recorded on the data file.

Body fat
A body fat measurement was included in Waves 4, 5, 6 and 7, with the reading provided by the same scales used for weight (Tanita body fat scales). Issues with the body-fat measurement are outlined in the Data Issues Paper.

Head circumference
This measurement was only taken for the B cohort in Wave 1, using an Abbott head circumference tape. Two measurements were taken, and if these differed by 0.5 cm or more, a third measurement was taken. The average of the two closest measures was included on the data file.
Chapter 3: Instruments

Blood pressure

This measurement was taken for the K cohort in Waves 4 and 5 and for the B cohort in Waves 6 and 7 using the A&D Digital Blood Pressure Monitor - Model UA-767. The interviewer took two measurements, with a one-minute interval between the measurements. Both of the readings were included in the data file.

3.1.2 ‘Who am I?’ (WAI)2

The ‘Who am I?’ (WAI) assessment is a direct child assessment measure that requires children to copy shapes (a circle, triangle, cross, square and diamond) and write numbers, letters, words and sentences. For the LSAC testing, there was a change to WAI Item 11: ‘This is a picture of me’ was replaced with a sentence to be copied, ‘John is big.’ The WAI assessment was used for children aged 4–5 years (Wave 1 K cohorts and Wave 3 B cohorts) to assess the general cognitive abilities needed for beginning school.

The study child was given his/her own answer booklet to draw and write in. What they wrote/drew was assessed by experienced researchers at the Australian Council for Educational Research (ACER). See Data Issues Waves 1 to 7 for details of the Rasch Modelling used to score the WAI (available from growingupinaustralia.gov.au/data-and-documentation/issues-papers).

3.1.3 Peabody Picture Vocabulary Test (PPVT)3

A short form of the Peabody Picture Vocabulary Test (PPVT-III), a test designed to measure a child’s knowledge of the meaning of spoken words and his or her receptive vocabulary for Standard American English, was developed for use in the study. This adaptation is based on work done in the USA for the Head Start Impact Study, with a number of changes made for use in Australia.

Different versions of the PPVT containing different, although overlapping, sets of items of appropriate difficulty were used for the children at ages 4–5, 6–7 and 8–9 years. A book with 40 plates of display pictures was used. The child points to (or says the number of) a picture that best represents the meaning of the word read out by the interviewer.

Scores are created via Rasch Modelling so that changes in scores represent real changes in functioning, rather than just changes in position relative to peers. See Data Issues Waves 1 to 7 for more details (available from growingupinaustralia.gov.au/data-and-documentation/issues-papers).

3.1.4 Matrix Reasoning4

Children completed the Matrix Reasoning (MR) test from the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) at ages 6–7, 8–9 and 10–11 years. This test of non-verbal intelligence presents the child with an incomplete set of diagrams (an item) and requires them to select the picture that completes the set from five different options. The data file includes raw scores (number of correct responses) and scaled scores based on age norms given in the WISC-IV manual. The instrument comprises 35 items of increasing complexity. Children start on the item corresponding to their age-appropriate start point. If a child does not answer correctly on the first or second start-point items, the examiner should ask two items prior to the age-appropriate start point (called ‘reverse administration’). Reverse administration was not implemented in the LSAC instrument. See the discussion of this issue in Data Issues Waves 1 to 7 (available from growingupinaustralia.gov.au/data-and-documentation/issues-papers).

3.1.5 Executive functioning (EXEC/CogState)5

The executive functioning of children in the K cohort was tested at Wave 6 using three Cogstate cognitive tests, including the Identification task (IDNT), One-back test (ONBT), and Groton Maze Learning Test (GML). In Wave 7, the same battery of tests was used to examine the executive functioning of the PI of K cohort children. The outcome variables are contained in the CogState dataset, where a series of cognitive testing batteries have been customised for use in LSAC. Each row of a CogState dataset represents one task in the CogState test battery for one study subject in one test session.

---

2 The ‘Who Am I?’ is copyrighted by the Australian Council for Educational Research, Melbourne, 1999.
5 Executive functioning was assessed via direct cognitive assessment using the Cogstate cognitive testing battery. The Cogstate program produces a variety of cognitive tests, which can be found at Cogstate.com/
The IDNT tests choice reaction time and overt attention. The subject must determine stimulus colour and then pick the appropriate button depending on the colour of the stimulus. The subject is shown a playing card on the screen and asked to respond as quickly as possible to the question: “is the card red?”. Rapid and accurate responding requires children to pay attention to the colour of the card, but not its suit or number. The ONBT is a task of working memory, where the subject is required to remember the image of the last item they saw on the screen and compare the memory of this image to the next stimulus.

The GML test contains five learning trials (i.e. the subject repeats the same task five times), where the subject is shown a 10 x 10 grid of tiles on a computer touchscreen. A 28-step pathway is hidden among these 100 possible locations. The child is instructed to move one step from the start location and then to continue, one tile at a time, toward the end. The subject repeats the task while trying to remember the pathway they have just completed and learns the 28-step pathway though the maze on the basis of trial and error feedback. The scores are interpreted by calculating the total number of errors made in attempting to learn the same hidden pathway. A lower score indicates better performance.


### 3.1.6 Rice Test of Grammaticality Judgement (GJT/SLI)

As children grow older, different methods are needed to assess the presence or absence of specific language impairment (SLI). That is, to identify whether children are meeting expected performance levels in achieving the adult standard of English grammar. Where LSAC children were identified in early waves to have poor language performance, it was not possible to distinguish the children with and without SLI. The Rice Grammaticality Judgement Task (GJ Task) was therefore introduced in Wave 6 for children of the K cohort.

The GJ Task is a short, automated (administered by ACASI) task that requires the study child to distinguish between grammatical and non-grammatical utterances known to be vulnerable to SLI in English-speaking children (Rice, Hoffman & Wexler, 2009). The study child listens through earphones as 20 pre-recorded items are spoken and enters their response by clicking the appropriate radio buttons (1 for ‘Right’, 5 for ‘Not so good’, and 9 for ‘Hear again’). Its sensitivity and specificity for SLI are .70 with a ROC of approximately 0.85.

### 3.2 Response rates

The number and percentages of survey instruments of each type that were completed at each wave are shown in Table 2. More detailed information on non-response can be found in the technical papers on weighting and non-response available at growingupinaustralia.gov.au/data-and-documentation/technical-papers

#### Table 2: Waves 1–7 instrument response

<table>
<thead>
<tr>
<th>Wave 1 instrument *</th>
<th>B cohort Eligibleb</th>
<th>B cohort Actual c</th>
<th>Percentage</th>
<th>K cohort Eligibleb</th>
<th>K cohort Actual c</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2F</td>
<td>5,107</td>
<td>5,107</td>
<td>100</td>
<td>4,983</td>
<td>4,983</td>
<td>100</td>
</tr>
<tr>
<td>PIL</td>
<td>5,107</td>
<td>4,341</td>
<td>85</td>
<td>4,983</td>
<td>4,229</td>
<td>85</td>
</tr>
<tr>
<td>P2L</td>
<td>4,630</td>
<td>3,696</td>
<td>80</td>
<td>4,286</td>
<td>3,388</td>
<td>79</td>
</tr>
<tr>
<td>TUD 1</td>
<td>5,107</td>
<td>4,031</td>
<td>79</td>
<td>4,983</td>
<td>3,867</td>
<td>78</td>
</tr>
<tr>
<td>TUD 2</td>
<td>5,107</td>
<td>3,751</td>
<td>73</td>
<td>4,983</td>
<td>3,582</td>
<td>72</td>
</tr>
<tr>
<td>WAI</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>4,983</td>
<td>4,880</td>
<td>98</td>
</tr>
<tr>
<td>PPVT</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>4,983</td>
<td>4,382</td>
<td>88</td>
</tr>
<tr>
<td>HBC</td>
<td>788</td>
<td>342</td>
<td>43</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>CBC</td>
<td>436</td>
<td>233</td>
<td>53</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>TQ</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>4,761</td>
<td>3,276</td>
<td>69</td>
</tr>
<tr>
<td>AEDC</td>
<td>1,366</td>
<td>720</td>
<td>53</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>WI.5</td>
<td>5,061</td>
<td>3,573</td>
<td>71</td>
<td>4,935</td>
<td>3,594</td>
<td>73</td>
</tr>
</tbody>
</table>

---

6 Executive functioning was assessed via direct cognitive assessment using the Cogstate cognitive testing battery. The Cogstate program produces a variety of cognitive tests, which can be found at Cogstate.com/.

### Chapter 3: Instruments

<table>
<thead>
<tr>
<th>Wave 2 Instrument</th>
<th>Eligible</th>
<th>B cohort</th>
<th>%</th>
<th>Eligible</th>
<th>K cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2F</td>
<td>5,107</td>
<td>4,606</td>
<td>90</td>
<td>4,983</td>
<td>4,464</td>
</tr>
<tr>
<td>PID</td>
<td>4,606</td>
<td>4,504</td>
<td>98</td>
<td>4,464</td>
<td>4,358</td>
</tr>
<tr>
<td>PIL</td>
<td>4,606</td>
<td>3,536</td>
<td>77</td>
<td>4,464</td>
<td>3,495</td>
</tr>
<tr>
<td>P2L</td>
<td>4,099</td>
<td>3,128</td>
<td>76</td>
<td>3,804</td>
<td>2,949</td>
</tr>
<tr>
<td>TUD 1</td>
<td>4,606</td>
<td>3,477</td>
<td>75</td>
<td>4,464</td>
<td>3,446</td>
</tr>
<tr>
<td>TUD 2</td>
<td>4,606</td>
<td>3,459</td>
<td>75</td>
<td>4,464</td>
<td>3,460</td>
</tr>
<tr>
<td>PPVT</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>4,464</td>
<td>4,409</td>
</tr>
<tr>
<td>MR</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>4,464</td>
<td>4,402</td>
</tr>
<tr>
<td>PLE mail-out</td>
<td>400</td>
<td>96</td>
<td>24</td>
<td>612</td>
<td>199</td>
</tr>
<tr>
<td>HBC</td>
<td>791</td>
<td>533</td>
<td>67</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>CBC</td>
<td>1,672</td>
<td>1,144</td>
<td>68</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>TQ</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>4,447</td>
<td>3,632</td>
</tr>
<tr>
<td>W2.5</td>
<td>5,107</td>
<td>3,246</td>
<td>64</td>
<td>4,983</td>
<td>3,252</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wave 3 Instrument</th>
<th>Eligible</th>
<th>B cohort</th>
<th>%</th>
<th>Eligible</th>
<th>K cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2F</td>
<td>5,107</td>
<td>4,386</td>
<td>86</td>
<td>4,983</td>
<td>4,331</td>
</tr>
<tr>
<td>PID</td>
<td>4,386</td>
<td>3,831</td>
<td>87</td>
<td>4,331</td>
<td>3,807</td>
</tr>
<tr>
<td>P2L</td>
<td>3,900</td>
<td>2,753</td>
<td>71</td>
<td>3,707</td>
<td>2,680</td>
</tr>
<tr>
<td>TUD 1</td>
<td>4,386</td>
<td>2,959</td>
<td>67</td>
<td>4,331</td>
<td>2,961</td>
</tr>
<tr>
<td>TUD 2</td>
<td>4,386</td>
<td>2,950</td>
<td>67</td>
<td>4,331</td>
<td>2,963</td>
</tr>
<tr>
<td>PPVT</td>
<td>4,386</td>
<td>4,266</td>
<td>97</td>
<td>4,331</td>
<td>4,273</td>
</tr>
<tr>
<td>WAI</td>
<td>4,386</td>
<td>4,197</td>
<td>96</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>MR</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>4,331</td>
<td>4,270</td>
</tr>
<tr>
<td>PLE CATI</td>
<td>346</td>
<td>272</td>
<td>77</td>
<td>510</td>
<td>403</td>
</tr>
<tr>
<td>TQ</td>
<td>4,114</td>
<td>3,395</td>
<td>83</td>
<td>4,275</td>
<td>3,643</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wave 4 Instrument</th>
<th>Eligible</th>
<th>B cohort</th>
<th>%</th>
<th>Eligible</th>
<th>K cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2F</td>
<td>5,107</td>
<td>4,242</td>
<td>82</td>
<td>4,983</td>
<td>4,164</td>
</tr>
<tr>
<td>CASI</td>
<td>4,242</td>
<td>4,210</td>
<td>99</td>
<td>4,164</td>
<td>4,116</td>
</tr>
<tr>
<td>P2L</td>
<td>3,706</td>
<td>2,677</td>
<td>72</td>
<td>3,512</td>
<td>2,645</td>
</tr>
<tr>
<td>CSR</td>
<td>4,242</td>
<td>4,181</td>
<td>99</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>ACASI</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>4,169*</td>
<td>4,094</td>
</tr>
<tr>
<td>TUD</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>4,169*</td>
<td>3,994</td>
</tr>
<tr>
<td>PPVT</td>
<td>4,242</td>
<td>4,185</td>
<td>99</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>MR</td>
<td>4,242</td>
<td>4,180</td>
<td>99</td>
<td>4,169*</td>
<td>4,103</td>
</tr>
<tr>
<td>PLE CATI</td>
<td>439</td>
<td>377</td>
<td>86</td>
<td>572</td>
<td>493</td>
</tr>
<tr>
<td>TQ</td>
<td>4,143</td>
<td>3,427</td>
<td>83</td>
<td>4,025</td>
<td>3,352</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wave 5 Instrument</th>
<th>Eligible</th>
<th>B cohort</th>
<th>%</th>
<th>Eligible</th>
<th>K cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2F</td>
<td>5,107</td>
<td>4,085</td>
<td>80</td>
<td>4,983</td>
<td>3,956</td>
</tr>
<tr>
<td>CASI</td>
<td>4,077</td>
<td>4,010</td>
<td>98</td>
<td>3,952</td>
<td>3,857</td>
</tr>
<tr>
<td>P2L</td>
<td>3,512</td>
<td>2,444</td>
<td>70</td>
<td>3,277</td>
<td>2,333</td>
</tr>
<tr>
<td>CSR</td>
<td>4,026*</td>
<td>4,014</td>
<td>100</td>
<td>3,872</td>
<td>3,850</td>
</tr>
<tr>
<td>ACASI</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>3,873*</td>
<td>3,844</td>
</tr>
</tbody>
</table>

*Table continued on next page*
<table>
<thead>
<tr>
<th>Wave 6 instrument</th>
<th>B cohort</th>
<th>K cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eligible</td>
<td>Actual</td>
<td>%</td>
</tr>
<tr>
<td>F2F</td>
<td>5,107</td>
<td>3,764</td>
</tr>
<tr>
<td>CASI</td>
<td>3,759</td>
<td>3,668</td>
</tr>
<tr>
<td>P2L</td>
<td>3,197</td>
<td>2,311</td>
</tr>
<tr>
<td>CSR</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>TUD</td>
<td>3,648*</td>
<td>3,597</td>
</tr>
<tr>
<td>EXEC</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>GJT</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>MR</td>
<td>3,648*</td>
<td>3,585</td>
</tr>
<tr>
<td>PLE CATI</td>
<td>559</td>
<td>398</td>
</tr>
<tr>
<td>TQ</td>
<td>3,678</td>
<td>3,102</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wave 7 instrument</th>
<th>B cohort</th>
<th>K cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eligible</td>
<td>Actual</td>
<td>%</td>
</tr>
<tr>
<td>F2F</td>
<td>5,107</td>
<td>3,381</td>
</tr>
<tr>
<td>P1 CASI</td>
<td>3,374</td>
<td>3,287</td>
</tr>
<tr>
<td>P2L</td>
<td>2,794</td>
<td>1,999</td>
</tr>
<tr>
<td>CSR/CAI</td>
<td>3,238</td>
<td>3,224*</td>
</tr>
<tr>
<td>SC ACASI/CASI</td>
<td>3,238</td>
<td>3,213</td>
</tr>
<tr>
<td>W 7.25 CATI</td>
<td>441</td>
<td>55</td>
</tr>
<tr>
<td>TUD</td>
<td>3,238</td>
<td>3,059</td>
</tr>
<tr>
<td>EXEC</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>PLE CATI</td>
<td>508</td>
<td>325</td>
</tr>
<tr>
<td>TQ or TCHB</td>
<td>3,160</td>
<td>2,567</td>
</tr>
<tr>
<td>EHC – Employment</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>EHC – Resident Living Away</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>EHC – Study</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Notes: SC ACASI = B cohort and SC CASI = K cohort. Wave 6 CSR instrument was used and in Wave 7 CAI was used.

* Questionnaire acronyms are detailed above in section 3, Table 1: Data collection modes by wave. ‘Eligible’ means the number of LSAC children for whom a questionnaire was applicable (e.g. children are eligible for a HBC questionnaire if the child’s main care is attended for 8 hours or more per week and this is home-based care). ‘Actual’ means the number of respondents for whom a form was returned. Response rates for waves 2 to 7 as proportion of Wave 1 families. * Represents instances where a child interview was completed and the main interview with the parents was not. Specifically, in Wave 4 there were five cases (K cohort). In Wave 5 there were eight cases for the K cohort and four cases for the B cohort. In Wave 6 there were 11 cases for K cohort and four cases for the B cohort. In Wave 7 there were seven cases for B cohort and 41 cases for K cohort. N/A = Not administered. Also in Wave 7 an ‘in-between’ wave activity was conducted to address the increase in refusals, hence W7.25 was developed. CSR/CAI = Child Self-report & Computer Assisted Interview (introduced first time in K cohort) - both are an interviewer administered survey with the study child.
3.2.1 Parent 1 questionnaires

In Wave 1, interviewers encouraged the parents to complete the P1L and P2L forms while the interviewer was in the home. Interviewers were also able to pick up forms in some cases, when forms were left behind. Forms not given to interviewers were mailed back. Two reminders were made for forms that were not returned.

In Wave 2, P1 had two forms to complete. Interviewers were instructed that the PID form must be completed when they were in the home (resulting in a high response rate). The PIL was generally left behind to be mailed back, as there was not enough time for these to be completed. Interviewers were generally not required to pick up the forms. Up to four reminders were made for forms that were not returned; however, the PIL forms showed lower response rates in Wave 2 compared with Wave 1. This may have been because P1 had already completed one form or because interviewers did not generally pick up forms.

For Wave 3, there was only one P1 self-complete form. Interviewers were instructed that this form must be completed while the interviewer was in the home. However, only two thirds of parents were able to do so. Three reminders were sent for forms not returned.

In Wave 4, P1 was asked to complete a CASI, which resulted in a response rate of 99% of eligible respondents. This was higher than the response rate of 88% of eligible respondents achieved in Wave 3 using the self-complete form.

In Wave 5, response rates were very similar to response rates obtained in Wave 4. This was due to no mode changes and attrition tapering off.

In Wave 6, response rates were similar to previous waves using the same mode. There was a slight decrease from the K cohort completion of the CASI from 98% in Wave 5 to 96% in Wave 6.

In Wave 7, response rates saw a very slight decrease in the B cohort completion of the CASI from 98% in Wave 6 to 97% in Wave 7. While there was a slight increase in the K cohort completion of the CASI from 96% in Wave 6 to 99% in Wave 7.

3.2.2 Parent 2, TUD and teacher forms

Response rates to the P2L and the TUD were broadly similar between waves (Waves 1, 2 and 3; between 67% and 79%), while the carer and teacher questionnaire response rates were much improved in Wave 2, with similar response rates at Wave 3. In Wave 4 the TUD response rate was 96%. The higher response rate could be contributed to changes in the procedure and in the informant. In Waves 4, 5 and 6 the interviewer collected the TUD information from the child instead of the parent. The data were collected as part of the interview rather than leaving a diary that previously required completion and return via mail by respondent families after the visit. In Wave 7 hard copy questionnaires were collected from P2 for both B and K cohorts. However, TUDs and teacher forms were collected from B cohort children only.

3.2.3 PLE response

The PLE questionnaire was introduced in Wave 2 and applies to children who see their ‘parent living elsewhere’ (PLE) at least once a year. There are three stages at which non-response can occur: (1) obtaining contact details from P1; (2) obtaining permission from P1; and (3) receiving a response from the PLE.

In Wave 2, contact details were given for 69% of cases for the B cohort and 70% of cases for the K cohort, and responses were received from 35% of PLEs sent a questionnaire for the B cohort and 47% for the K cohort.

Due to the relatively low response in Wave 2 to the mail-out questionnaire, a change in methodology was introduced in Wave 3. Where P1 had provided contact details, PLEs were telephoned and asked to respond to a computer-assisted telephone interview (CATI). The response from PLEs who were approached was very positive. Of the 856 PLEs that interviewers attempted to contact, interviews were achieved with 675 (79%) PLEs and only 53 (6%) PLEs refused an interview. Most of the remaining non-responses were due to not being able to contact the PLE.

In Wave 3, P1 was explicitly asked for their permission to contact the PLE. Therefore, it was easy for P1 to refuse to provide any information about the PLE or refuse the PLE’s participation. This meant that no information was obtained for 260 (18%) PLEs.

It is worth noting that from Wave 4 onwards, there was no direct question asking the P1 permission to contact the PLE. However, some P1 respondents refused the PLE’s participation by not providing contact details.

Table 3 summarises the PLE response rates from Waves 3 to 7.
### Table 3: Wave 3 to Wave 7: Information obtained with regard to PLE

<table>
<thead>
<tr>
<th></th>
<th>Wave 3</th>
<th>Wave 4</th>
<th>Wave 5</th>
<th>Wave 6</th>
<th>Wave 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>K</td>
<td>Total</td>
<td>B</td>
<td>K</td>
</tr>
<tr>
<td>PLE identified during P1 interview</td>
<td>578</td>
<td>837</td>
<td>1,415</td>
<td>674</td>
<td>878</td>
</tr>
<tr>
<td>Eligible PLE*</td>
<td>346</td>
<td>510</td>
<td>856</td>
<td>439</td>
<td>572</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>K</td>
<td>Total</td>
<td>B</td>
<td>K</td>
</tr>
<tr>
<td></td>
<td>773</td>
<td>911</td>
<td>1,684</td>
<td>778</td>
<td>817</td>
</tr>
<tr>
<td></td>
<td>732</td>
<td>756**</td>
<td>1,488</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *The PLE is considered eligible when: (1) the PLE satisfies the parental requirements; i.e. PLEs who see the study child at least once a year; (2) the PLE’s contact details are available; (3) P1 did not explicitly refuse permission to contact the PLE. ** There were 19 (RAP) PLEs identified during P1 interview and 9 (RAP) identified as Eligible PLE* in the K cohort.

### 3.2.4 Wave 7 RAP response

Delays in enumeration hindered the progress of identifying populations such as RAP children, RAP parents and RAP PLEs in Wave 7. This had flow-on effects in contacting these respondents, and the timing available for tracking or follow up. During Wave 7 enumeration, 24 RAP parent records were generated. Of these, 13 (54.2%) parents undertook an interview, while three parents (12.5%) refused, eight parents (33.3%) were not contactable.

Table 4 summarises the final RAP response rates for Wave 7.

### Table 4: Summary of RAP field response for Wave 7

<table>
<thead>
<tr>
<th></th>
<th>Study Child</th>
<th>Parent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Responding</td>
<td>27</td>
<td>35.5</td>
</tr>
<tr>
<td>Refusal*</td>
<td>4</td>
<td>5.3</td>
</tr>
<tr>
<td>Non-contact</td>
<td>45</td>
<td>59.2</td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>100</td>
</tr>
</tbody>
</table>

Note: * Includes avoidance

### 3.2.5 Wave 7.25 response

The fully responding rate for the K cohort was significantly lower than the B cohort as this required collecting the respondent engagement questions from both the P1 and the SC, as well as all of the CATI Wave 7 catch-up questions from the SC.

For both the B and K cohorts the non-contact rate was by far the largest with almost 50% of all records being unable to be contacted. Interviewers were advised to only make up to three call attempts before finalising selections (as is standard for follow-up refusal workloads). This would have had an impact on their ability to get hold of respondents.

Table 5 summaries the final response rates for Wave 7.25.

### Table 5: Response rates for Wave 7.25

<table>
<thead>
<tr>
<th>Field response</th>
<th>Cohort B</th>
<th>Cohort K</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Fully responding</td>
<td>55</td>
<td>12.5</td>
<td>13</td>
</tr>
<tr>
<td>Part responding*</td>
<td>131</td>
<td>29.7</td>
<td>154</td>
</tr>
<tr>
<td>Refusal**</td>
<td>84</td>
<td>19.1</td>
<td>42</td>
</tr>
<tr>
<td>Non-contact</td>
<td>171</td>
<td>38.8</td>
<td>242</td>
</tr>
<tr>
<td>Total</td>
<td>441</td>
<td>100.0</td>
<td>451</td>
</tr>
</tbody>
</table>

Notes: * Respondent engagement questions only (i.e. no CATI catch-up questions). ** For Ks, both the P1 and SC refused to take part or P1 refused for themselves and the SC.
4 The LSAC data release

LSAC data is available through Dataverse in partnership with the Australian data archive. The general release data and restricted release data are available at no cost. For more details about how to access the data, proceed to the Data User Guide for the Department of Social Services Longitudinal data sets on the National Centre for Longitudinal Data website (dss.gov.au/about-the-department/national-centre-for-longitudinal-data). This guide outlines the requirements for data users.

National Centre for Longitudinal Data (NCLD)

“Our mission is to see evidence of success as points of light all around us and join them up to create a universe of opportunity for our children. These results from the Footprints in Time data provide more points of light.”

Professor Mick Dodson AM
Director National Centre for Indigenous Studies, Australian National University

Our mission
To promote a longitudinal evidence base that informs policies and practices to improve the lifetime wellbeing of people and families in Australia.

To achieve our mission the NCLD will:
• develop a national framework (or architecture) for government investment in longitudinal data
• produce and manage quality longitudinal data sets and encourage their use
• foster collaboration between longitudinal survey developers, researchers and policy makers
• facilitate broader use of longitudinal data

About longitudinal data
Longitudinal data is data collected from the same individuals (or households, businesses or other entities) repeatedly over time. The data is able to show how actions and events can affect outcomes later in life.

The Longitudinal Studies and the National Centre for Longitudinal Data (NCLD) fact sheet explains the value of longitudinal data, and particularly how it can support decision makers to address critical questions.

The NCLD’s studies
The NCLD brings together longitudinal surveys managed within the Department of Social Services.
The four active longitudinal surveys managed in the NCLD are:
• Building a New Life in Australia (BNLA): The Longitudinal Study of Humanitarian Migrants
• Building a New Life in Australia (BNLA) fact sheet
• Footprints in Time: The Longitudinal Study of Indigenous Children (LSC)
• Footprints in Time (LSC) fact sheet
• Growing up in Australia: The Longitudinal Study of Australian Children (LSAC)
• Growing up in Australia: LSAC fact sheet
## 5 File structure

For the Wave 7 data general release, the following datasets are available.

### Table 6: Data release for waves and cohorts

<table>
<thead>
<tr>
<th>Number of datasets</th>
<th>Description of datasets</th>
<th>Main dataset for each wave and cohort</th>
<th>Data type</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Main datasets for each wave and cohort</td>
<td>lsacgrb0*, lsacgrb2, lsacgrb4, lsacgrb6, lsacgrb8, lsacgrb10, lsacgrb12, lsacgrk4*, lsacgrk6, lsacgrk8, lsacgrk10, lsacgrk12, lsacgrk14, lsacgrk16</td>
<td>Main</td>
</tr>
<tr>
<td>2</td>
<td>Study child household</td>
<td>hhgrb, hhgrk</td>
<td>Supplementary</td>
</tr>
<tr>
<td>1</td>
<td>P1 RAP household</td>
<td>p1raphgrk16</td>
<td>Supplementary</td>
</tr>
<tr>
<td>8</td>
<td>PLE household</td>
<td>plehgrb6, plehgrb8, plehgrb10, plehgrb12, plehgrk10, plehgrk12, plehgrk14, plehgrk16</td>
<td>Supplementary</td>
</tr>
<tr>
<td>3</td>
<td>Event history calendar</td>
<td>ehcegrk16, ehcgrk16, ehcgrk16</td>
<td>Supplementary</td>
</tr>
<tr>
<td>2</td>
<td>Executive functioning</td>
<td>execsc, execp1</td>
<td>Supplementary</td>
</tr>
<tr>
<td>23</td>
<td>Time use diary</td>
<td>tudb10, tudb12, tudk10, tudk12, tudk14</td>
<td>Supplementary</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• one cleaned data file with problematic cases deleted for each cohort for Waves 1, 2 and 3 (diaryb0, diaryb2, etc.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• one data file with the cases deleted from the above files after cleaning for each cohort for Waves 1, 2 and 3 (poortudsb0, poortudsb2, etc.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• one data file with all cases and no data cleaning performed on them for each cohort for Waves 1, 2 and 3 (ucdiaryb0, ucdiaryb2, etc.)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Wave 2.5</td>
<td>lsacgrb3, lsacgrk7</td>
<td>Supplementary</td>
</tr>
<tr>
<td>2</td>
<td>Wave 3.5</td>
<td>lsacgrb5, lsacgrk9</td>
<td>Supplementary</td>
</tr>
<tr>
<td>7</td>
<td>Medicare Australia</td>
<td>mbssc, pbssc, mbsp1, mbsp2, pbsp1, pbsp2, acir</td>
<td>Linked</td>
</tr>
<tr>
<td>1</td>
<td>NAPLAN</td>
<td>lsacnaplan</td>
<td>Linked</td>
</tr>
<tr>
<td>1</td>
<td>MySchool</td>
<td>lsacmyschool</td>
<td>Linked</td>
</tr>
<tr>
<td>1</td>
<td>AEDC</td>
<td>aedc^</td>
<td>Linked</td>
</tr>
<tr>
<td>3</td>
<td>Centrelink welfare</td>
<td>isp_summary, ftb_summary, concession_cards^</td>
<td>Linked</td>
</tr>
<tr>
<td>1</td>
<td>Child Health CheckPoint</td>
<td>lsacgrcp^</td>
<td>Substudy</td>
</tr>
</tbody>
</table>

**Notes:**  
* Wave 1.5 datasets have been added to the Wave 1 datasets. This was possible because all participants who responded at Wave 1.5 had to complete a Wave 1 interview. This is not the case with the other between-wave mailouts, as respondents may have completed any prior combination of interviews. This structure has been used to reduce the size of the main datasets and because some data are formatted using more than one record for each child.  
^ This is available with additional approval.
5.1 Main dataset

The main dataset consists of the data from all questionnaires except the time use diary, Wave 2.5, Wave 3.5, Wave 4.5, Wave 5.5, some household composition information and linked datasets. Data from the instruments are presented in the following order:

- FCF (Wave 1 files only)
- F2F
- P1 self-complete (except Wave 1 files)
- P2 self-complete
- PLE self-complete/interview (except Wave 1 files)
- Teacher/Carer questionnaire
- Wave 1.5 data (Wave 1 files only)

A number of derived variables are included in the output dataset alongside the raw responses used in their derivation. Additionally, the main datasets contain status variables (e.g. date of interview, whether each type of form was returned, etc.), ABS Population Census and NCAC data, and weights.

5.1.1 Australian Bureau of Statistics Census of Population and Housing data

Public data from the Australian Bureau of Statistics Census of Population and Housing have been added to the file to enhance the range of neighbourhood characteristics available for analysis with the LSAC data. Census data are available for the child’s residence from Waves 1 to 7.

The items currently included are:

- SEIFA – rounded off to the nearest 10 for on the general release file
- remoteness area classification
- percentage of persons aged under 5, 10 and 18 years
- percentage of persons born in Australia
- percentage of persons speaking English-only at home
- percentage of persons with Aboriginal and Torres Strait Islander (ATSI) origins
- percentage of persons who completed Year 12 schooling
- percentage of persons in above-median income category
- percentage of persons working
- percentage of households with internet capacity (in 2006 Census only)
- percentage of households with broadband (in 2006 Census only).

Census data is either linked at the Statistical Local Area (SLA), before 2011, or the Australian Statistical Geography Standard (ASGS) level, from 2011, or, where this wasn’t available, the child’s postcode. One estimate is provided for each time point representing a linear interpolation of the data at the censuses either side of the time period. For example, if a SLA had 4.2% of people with ATSI origins in 2001 and 6.5% with ATSI origins in 2006 then the estimate for the proportion in 2004 would be:

\[
\text{estimate} = 2001\text{Data} + (2006\text{Data} - 2001\text{Data}) \times \frac{\text{time}_{\text{since census}}}{\text{time}_{\text{between censuses}}}
\]

\[
\text{estimate} = 4.2\% + (6.5\% - 4.2\%) \times \frac{(2004 - 2001)}{(2006 - 2001)}
\]

\[
\text{estimate} = 4.2\% + 2.3\% \times 0.6
\]

\[
\text{estimate} = 5.6\%
\]

If data is only available for one of the censuses then no interpolation is performed. A ‘link type’ variable is included to tell data users whether the linkage was performed using statistical area level or postcode and whether the 2001 census, 2006 census, 2011 census or 2016 census or all were used.
5.1.2 National Childcare Accreditation Council data

A key research question in LSAC relates to the effect of child care on children’s developmental outcomes over time. While LSAC collected parent-reported information on children’s child care histories and carer reports on the child care environment, relatively little systematic information was collected on the quality of child care.

The National Childcare Accreditation Council Inc. (NCAC) as it was then had quality assurance data on every long day care (LDC) centre, some family day care (FDC) schemes and some before- and after-school care providers. The LSAC dataset includes linked NCAC data for most children using LDC or FDC at Wave 1, where contact details of this care were obtained and matched with NCAC data. The match rate obtained during the linkage process was 78% for Wave 1, 82% for Wave 2, 84% for Wave 3 and 92% for Wave 4.

One complication in using the NCAC data is due to the change of accreditation systems for both FDC and LDC. In Wave 1, all cases had FDC assessed under the guidelines laid out in second edition of the FDCQA Quality Practices Guide (NCAC, 2004), while from Wave 2 and onwards, all cases have been assessed under the third edition of this guide, introduced in July 2005. The revised guidelines contain the same quality areas (though some have been combined) but the number of principles used to assess these areas has been reduced from 35 to 30. The old scheme had 10 quality areas assessed by 35 principles, while the new scheme has seven quality areas assessed by 30 principles.

For LDC, all Wave 1 centres were assessed under the QIAS Validation Report, 2nd Edition (NCAC, 2003). From July 2006, accreditation decisions were made under the QIAS Quality Practices Guide, 1st Edition. As a consequence, some of the Wave 2 and 3 accreditations were made under the new scheme, while some were made under the old scheme.

Before- and after-school care arrangements were assessed by the guidelines laid out in the OSHCQA Quality Practices Guide, 1st Edition (NCAC, 2003). In Waves 2 and 3, some accreditations were made under the new scheme, while some were made under the old scheme.

The variables included are:

- date of accreditation
- date of validation
- accreditation status
- LDC v1 Quality area 1: Relationships with children
- LDC v1 Quality area 2: Respect for children
- LDC v1 Quality area 3: Partnerships with families
- LDC v1 Quality area 4: Staff interactions
- LDC v1 Quality area 5: Planning and evaluation
- LDC v1 Quality area 6: Learning and development
- LDC v1 Quality area 7: Protective care
- LDC v1 Quality area 8: Health
- LDC v1 Quality area 9: Safety
- LDC v1 Quality area 10: Managing to support quality
- LDC v2 Quality area 1: Staff relationships with children and peers
- LDC v2 Quality area 2: Partnerships with families
- LDC v2 Quality area 3: Programming and evaluation
- LDC v2 Quality area 4: Children’s experiences and learning
- LDC v2 Quality area 5: Protective care and safety
- LDC v2 Quality area 6: Health, nutrition and wellbeing
- LDC v2 Quality area 7: Managing to support quality
- FDC Quality area 1: Interactions
- FDC Quality area 2: Physical environment
- FDC Quality area 3: Children’s experiences, learning and development
- FDC Quality area 4: Health, hygiene, nutrition, safety and wellbeing
- FDC Quality area 5: Carers and coordination unit staff
Chapter 5: File structure

- FDC Quality area 6: Management and administration
- OHS Quality area 1: Respect for children
- OHS Quality area 2: Staff interactions and relationships with children
- OHS Quality area 3: Partnerships with families and community links
- OHS Quality area 4: Programming and evaluation
- OHS Quality area 5: Play and development
- OHS Quality area 6: Health, nutrition and wellbeing
- OHS Quality area 7: Protective care and safety
- OHS Quality area 8: Managing to support quality

Demographic data

The data used to develop the quality areas were collected from six sources:

- a self-study report prepared by centre management
- a validation survey completed by the director
- a validation survey completed by staff
- a validation survey completed by families
- a validation report completed by an independent peer
- a set of moderation ratings completed by independent moderators.

Data on 35 principles were collected. Each principle was related to one of the 10 quality areas. Response categories for each principle were: ‘unsatisfactory’, ‘satisfactory’, ‘good quality’ and ‘high quality’. Proportionally weighted factor-score regression coefficients for principle ratings were calculated to determine the extent to which each principle contributed to a quality area. For further information, see Rowe (2006).

As no data about the child was obtained, no consent was required from parents to collect this data (although parents did need to give details about their carers to assist in the linking).

5.2 Supplementary files

5.2.1 Household composition data

The study child household: At each wave of data collection, detailed information about every member of the household where the study child resides is collected. Information is collected about people currently residing in the study child’s household, as well as people who have come and gone between waves but lived with the study child for at least three months. This information is usually collected from Parent 1 only. However, from Wave 7 onwards, if a study child has moved out of the parental household, this information is collected directly from the study child. Parent 1 is still asked to provide information on their own household (P1 RAP).

The main household dataset for each cohort contains one record for each study child, detailing the composition of their household from their recruitment to the study to the most recent data collection. This dataset also includes ex-household members (with a variable indicating that they are no longer resident), such as parents living elsewhere who were resident at a previous wave. The details collected about the study child, P1 and P2 are included in each main dataset, along with a number of derived variables on household composition.

The study child’s household is always the household where the study child resides. When the study child resides with parents, the information is collected about the parental household and saved in the household file ‘hhgrb/k’. When the study child moves out of the parental household (SC RAP) to another household (independent living) the information is collected about all members of the household the study child moves to and is saved in the household file ‘hhgrb/k’.

PLE household: PLE household composition data is released from Wave 4 and contains detailed information about every member of the household in which the parent living elsewhere lives. The household data file is wave specific and released cross-sectionally at every wave, one record per study child.

P1 RAP household: Another household composition data file available in Wave 7 for the K cohort is the P1 RAP. This file contains detailed information about every member of the P1 RAP household and is saved in the file ‘piraphgrb/k’. The P1 RAP household is a parental household of study children who were living away from P1 during the Wave 7 interview.
5.2.2 Event history calendar

The event history calendar (EHC) was introduced in Wave 7 to collect retrospective reports of events and the timings of those events from the K cohort children. The primary focus of the EHC was to capture information on residential living arrangements, study and employment domains. Three data files are available with each corresponding to the specific domain (ehcrgrk16 – resident living away, ehcegk16 – employment and ehcsgrk16 – study). The files are structured as long format data, allowing multiple reports of events per child where possible. The EHC data file names are wave specific with the keyword ‘K16’ representing the 16 years of age of K cohort respondents. The EHC was able to capture all the changes that have occurred in these domains since the Wave 6 interview; or if the respondent was not interviewed in Wave 6, the two years preceding the date of the Wave 7 interview.

5.2.3 Executive functioning

Executive functioning data was collected from K cohort study children in Wave 6 and the parents (P1) of K cohort study children in Wave 7 interviews. This information is available in two separate data files (execksc and execkp1). The data file names are respondent-specific with keywords KSC and KPI representing study children and parents of the K cohort. However, the first letter of variable names in these data files represents wave-specific/child age indicator information. Further information about Cogstate data collection is available in LSAC Technical Paper No. 19, *Executive Functioning: Use of Cogstate Measures in the Longitudinal Study of Australian Children* (available from the study website: growingupinaustralia.gov.au/data-and-documentation/technical-papers).

5.2.4 Time use diary data

In Waves 1–3, responding families were given two time use diaries (TUDs) to complete at each wave. Each record in the TUD data relates to a single diary; that is, each child can have up to two records (one for each TUD).

This paper form TUD gathered information on children’s activities and the context of 96 15-minute periods in each 24-hour block. In addition to these variables, the TUD data includes the child’s unique identification number in order to allow linkage with the main dataset. It also includes the following general descriptors:

- date diary should be completed
- day of week diary should be completed
- diet of the study child on the day in question (Waves 2 and 3)
- relationship of the diary writer to the child
- over what duration the diary was completed
- actual day and date of completion
- hours of work done by respondent on day of completion (Waves 2 and 3)
- the kind of day described in the diary.

Due to scanning problems in Wave 1, and other data quality issues that are likely to apply equally across waves, a number of imputations and corrections have been applied to the TUD data (see *Data Issues: Waves 1 to 7* for details, available at growingupinaustralia.gov.au/data-and-documentation/issues-papers). So, researchers can determine the effect of these imputations/corrections to the data on any analysis. An uncorrected version of the TUD data is also provided, as well as files containing imputations/corrected versions of cases that were considered unsuitable for data analysis even after correction.

LSAC Technical Papers 4 and 13 include detailed discussions of issues that should be considered when using the time use data. The technical papers are available from growingupinaustralia.gov.au/data-and-documentation/technical-papers.

In Wave 4 a new methodological approach was undertaken due to a shift from the parent being the informant to the study child being the informant. In Waves 4–6 only the K cohort completed the TUD, which was substantially different from the TUDs that the parents had completed in earlier waves. With the child being the informant, the interviewer was directly involved in working with the child to transfer information from the diary into a computer instrument. In Wave 7 the TUD was collected only for B cohort. Waves 4–7 had the form of an ‘ABS Activity Episode’ diary. This data is stored as a long file, as opposed to the wide files the previous diaries were stored as.
Example analysis

SAS

The following code gives the proportion of children eating or drinking while watching a TV, video, DVD or movie at any time of day for the B cohort at Wave 1. Statements 1 and 2 tell SAS to create a new dataset beginning with the data in the mtud.diary2 file (you will need to use your own library name). The third statement tells SAS to treat the time use data as a multidimensional array (x) containing 96 rows of 40 columns each. The next statement tells SAS to set up a new array of 96 variables (TVeat) into which the data for eating in front of the TV will be derived.

Statements 5–8 contain a do loop, which runs across all 96 time periods. Statement 5 tells SAS to create a variable ‘i’ to keep track of which time period is being worked on, and to give it the values 1-96 in turn. Statement 6 tells SAS to allocate the value 100 at the position in the ‘TVeat’ array for the current time period if the child was eating or drinking (column 4 in the array ‘x’) and was watching a TV, etc. (column 12 in ‘x’). Statement 7 says the value of 0 will be assigned if the child either wasn’t eating or drinking or wasn’t watching TV, etc., and the diarist was sure of the child’s activities for the time period. This means that cases where the diarist wasn’t sure, or didn’t fill any information in for activities in this time period, will have missing data. Statement 8 finishes the do loop, and statement 9 finishes the data step so SAS runs the above statements.

Statements 10–13 produce the means of the variables in the ‘TVeat’ array (which SAS gives the names TVeat1 to TVeat96 by default). The mean here will be the percentage of children from whom an activity was known that ate or drank in front of the TV, etc., at each time period. Line 12 uses the day weight variable ‘bweightd’ to ensure the proportion is representative of the population and represents each day of the week equally.

```
data diary2;
  set mtud.diary2;
  array x [96,40] b2da0101--b2de0196;
  array TVeat [96];
  do i=1 to 96;
    if x[i,4]=1 and x[i,12]=1 then TVeat[i]=100;
    else if (x[i,4]=0 or x[i,12]=0) and x[i,1] ^=1 then TVeat[i]=0;
  end;
run;
proc means data=diary2;
  var TVeat1-TVeat96;
  weight bweightd;
run;
```

This data can be used to produce a graph known as a tempogram.

Figure 2 shows the data produced by the example program along with the equivalent data for the K cohort at Waves 1 and 2. It shows that children did more of this as they got older, and that this activity was most common in the early mornings.
Figure 2: Tempogram of children watching TV, video, DVD or movie while eating or drinking by wave and cohort.

SPSS

The equivalent code to derive the TVeat variable in SPSS would appear as:

```plaintext
do repeat
eat b2da0401 b2da0402 ... b2da0496/
tv b2da1201 b2da1201 ... b2da1296/
dk b2da0101 b2da0101 ... b2da0196/
tve tveat1 to tveat96.
  if (eat=1 or tv=1) tve=1.
  if ((eat=0 or tv=0) and dk=0) tve=0.
end repeat.
```

STATA

The equivalent code to derive the TVeat variable in STATA would look like:

```plaintext
foreach n of numlist 1/9 {
  gen tveat`n'=1 if (b2da040`n'==1 & b2da120`n'=1)
  replace tveat`n'=0 if ((b2da040`n'==0 | b2da120`n'==0) & b2da010`n'=0)
}
foreach n of numlist 10/96 {
  gen tveat`n'=1 if (b2da04`n'==1 & b2da12`n'=1)
  replace tveat`n'=0 if ((b2da04`n'==0 | b2da12`n'==0) & b2da01`n'=0)
}
```
5.2.5 Wave 2.5 data

The data from the Wave 2.5 mailout is included in two separate datasets. Unlike Wave 1.5 in relation to Wave 1, families that responded to Wave 2.5 did not necessarily respond to Wave 2. Merging these with the Wave 2 datasets would have resulted in a number of largely blank cases on the data file.

The data in the Wave 2.5 file consists of questionnaire items, a small number of derived items and linked census data based on the postcodes of responding families at the time of Wave 2.5. Unfortunately, formatting of the questionnaires resulted in some respondents skipping items that they should have answered. Imputation has been performed on some items where it was possible to infer the data for these questions based on responses to other questions. See Data Issues: Waves 1 to 7 (available from growingupinaustralia.gov.au/data-and-documentation/issues-papers).

5.2.6 Wave 3.5 data

The data from the Wave 3.5 mailout is included in a separate dataset, in the same way that data from Wave 2.5 was included.

The data in the Wave 3.5 file consists of questionnaire items, a small number of derived items and linked census data based on the postcodes of responding families at the time of Wave 3.5. Imputation has been performed on some items where it was possible to infer the data for these questions based on responses to other questions. See Data Issues: Waves 1 to 7 for further information (available from growingupinaustralia.gov.au/data-and-documentation/issues-papers).

5.2.7 Distance to coast data

Distance to coast has been generated for every residential address in Waves 1–7 by geocoding latitude and longitude information. The distance to the coast data for each cohort (B and K) are stored in a separate data file. The dataset contains one record per study child with multiple distance-related variables representing different waves of data collection as denoted by the first letter of the variable name. See Distance to coast data information (available from growingupinaustralia.gov.au/sites/default/files/distance_to_coast_data_information.pdf), providing information on distance calculation and confidentialisation strategy. Distance to coast data are only available with restricted release data files.

5.3 Linked data

5.3.1 Medicare Australia data

In Wave 1, 97% of parents of study children gave consent for their children’s data to be linked with Medicare Australia data on an ongoing basis. This includes data from the Medicare Benefit Scheme (MBS), the Pharmaceutical Benefit Scheme (PBS) and the Australian Childhood Immunisation Records (ACIR). Data from these sources provide details of usage history of MBS, PBS and ACIR services. Linkage was successful for 93% of children (incomplete consent forms resulted in data not being released for about 400 children).

Since the child’s use of medical services is ongoing, the Medicare Australia data are not broken into waves but are provided as three separate files:

ACIR: Each record in the file represents an immunisation that the child has had.

MBS: Each record on this file represents a benefit claim.

PBS: Each record represents a benefit claim.

In Wave 7, Parent 1 and Parent 2 themselves consented to their data linkage for MBS, PBS and RPBS.

ACIR file

Records are currently available for payments received from birth to early 2013. The following variables are included on the file:

- child identification number
- vaccination code
- vaccination name
- scrambled provider ID
Some of the vaccination codes contain dose numbers, indicating a vaccine that has been received in a series of doses. The sequence of doses for these has been included in the dataset (i.e. 1st, 2nd, etc.). If a dose is missing, it means that it was either not reported to ACIR or it was missed.

**MBS file**

Records are currently available for services between January 2002 (or birth for the B cohort) and early 2015. The following variables are included on this file:

- child identification number
- item number
- item name
- amount of benefit paid
- hospital indicator
- scrambled provider ID
- date of payment
- date of service.

Some cases have very small or negative benefit amounts. In relation to negative benefits, this indicates that an adjustment has been made to the Medicare benefit records. There are several reasons why this may happen:

- It is a correction of a data entry made against the wrong individual reference number on a Medicare card (i.e. service is initially incorrectly recorded against someone else on the same card).
- The provider has issued an amended account.
- A new cheque has been issued to replace lost/stolen/unpresented cheques.

In relation to small benefits:

- There are a number of item numbers that have small benefits; for example, many pathology-related claims.
- There are also small amounts for things such as bulk bill incentives (generally around $5–6).
- The claimant had reached the Medicare Safety Net (MSN) threshold. Once the threshold has been reached, the family’s out-of-pocket expenses are tallied and a payment is calculated for a percentage of the substantiated amounts. In effect, there can be two payments made for the same doctor’s visit – one to the doctor for the service and one to the claimant for MSN purposes.

**PBS file**

The final of these datasets contains the PBS data. Again, each record represents a benefit claim. Records are available for medications supplied between May 2002 (or birth for the B cohort) and early 2015. The following information is included for each record:

- child identification number
- item code
- item name
- quantity
- benefit paid
- prescription type (original, repeat or unknown)
- payment category
- payment status
- date of payment
- date of supply.

**Example derivations**

There are simple techniques in SAS, SPSS and STATA to summarise across multiple records to create derived items from the Medicare datasets. The following code samples create a variable (ben07) for the amount of PBS benefits paid for a child in 2007. Note that this variable will initially be missing for cases that had no PBS claims.
in 2007 as well as those for which data linkage was unsuccessful. The 'match' file can be used to distinguish between these cases and set ben07 to 0 for those with no claims. This file contains a variable called 'medicare', which is 1 if linkage is successful for a case and 0 otherwise.

**SAS**

```sas
proc means data=m.pbs nway sum;
   class hicid;
   var benefit;
   where datesupp>=mdy(1,1,2007) and datesupp<=mdy(1,1,2008);
   output out=temp sum=ben07;
run;

data temp;
   merge temp m3.match;
   by hicid;
   if medicare=1 and ben07=. then ben07=0;
run;
```

**SPSS**

```spss
temp.
select if (datesupp >= date.dmy(1,1,2007) & datesupp <= date.dmy(31,12,2007)).
aggregate
   /outfile='/temp.sav'
   /break=hicid
   /ben07=sum(benefit).
get
   file='/temp.sav'.
match files /file=*
   /file='/match.sav'
   /by hicid.
if (medicare=1 & missing(ben07)) ben07=0.
execute.
```

**STATA**

Note that the collapse command will delete all other data than hicid and ben07. Ensure it is saved to a new file.

```stata
collapse (sum) ben07=benefit if (datesupp>=mdy(1,1,2007) & datesupp<=mdy(1,1,2008)), by(hicid)
merge hicid using match
replace ben07=0 if (medicare==1 & ben07==.)
keep if ben07!=.
sort hicid
save temp, replace
```
5.3.2 AEDC data

Every three years since 2009, the Australian Government has undertaken a census of all children in their first year of full-time schooling: see aedc.gov.au/parents

The Australian Early Development Census (AEDC) data for B cohort children were obtained from the Department of Education. The Department of Education is responsible for the AEDC. The Social Research Centre manages the data. The data contains no variable labels or value labels but these can be found in the Data Dictionary provided on the AEDC website: aedc.gov.au/data-users/resources-for-data-users/data-dictionary

5.3.3 NAPLAN data

In Wave 3, 81% of parents of K cohort children gave consent for their child’s data to be linked with NAPLAN data for the duration of the study. Linkage was successful for 96% of these children. For 4% of children, the NAPLAN data were not found, either because these children had not sat NAPLAN tests yet or they sat the NAPLAN tests in 2008 or 2009 but a match was not found. Families who did not give consent or who did not participate at Wave 3 were asked again at Wave 4. Out of 964 families who were followed up in Wave 4, 847 gave consent to link NAPLAN results.


LSAC Technical Paper B Using National Assessment Program - Literacy and Numeracy (NAPLAN) data in the Longitudinal Study of Australian Children (LSAC) includes a detailed discussion of the NAPLAN data linkage process and data issues, and should be considered when using the LSAC NAPLAN data. The report is available from growingupinaustralia.gov.au/data-and-documentation/technical-papers

5.3.4 ACARA MySchool data

Data has been obtained from ACARA. ACARA is responsible for collating NAPLAN data received from Australian schools, collecting school characteristics and managing the MySchool website. Some of the data ACARA collects and collates on Australian schools is publicly available on the MySchool website. School data about the schools LSAC participants attend has been linked onto the LSAC survey datasets and is available to data users. See growingupinaustralia.gov.au/data-and-documentation/technical-papers

5.3.5 Centrelink welfare data

During Wave 7 enumeration consent was collected from the K cohort study child’s parents (P1 and P2) to link their Centrelink welfare benefits back to 1 January 1999 and from the K cohort study child to link back to their 16th birthday. The data includes information on income support payments, Family Tax Benefit, Carer Allowance and concession cards. The data released with Wave 7 are extracted up until the end of the 2016/17 financial year (30 June 2017), apart from the Family Tax Benefit data, which is only extracted up until 30 June 2015 as it is based on entitlement calculated after reconciliation with tax data.

The linked Centrelink data is provided in separate datasets from the main LSAC data files and there are both general release and restricted release versions. These files are not supplied automatically with the LSAC data files and have to be explicitly requested. The general release version of the Centrelink data can be applied for by data users applying for either the general release or the restricted version of the main LSAC files at no additional cost. The restricted version of the Centrelink data can also be applied for users of the general release LSAC file.

Applicants for the restricted Centrelink files will need to present a project rationale for access to the restricted data making it clear why this data is essential for their research. This will entail either specifying why particular data items are required or why the research questions require access to episodic income support data. See below for a description of the data available in the two versions of the Centrelink files.

General release Centrelink files

ISP_Summary: Contains data for ISP receipt aggregated at financial-year level. For each participant who has received an income support payment in a particular year there will be a single observation. The following information is included in the summary file:

- benefit type received by the participant for the greatest duration during the year
- number of days that the participant received an income support payment and duration they received the primary benefit type
- duration in receipt of rent assistance, home ownership status and rent type
number of days the participant received other income while in receipt of an income support payment
number of days the participant was partnered
indicators for receipt of carer allowance payment and low income card

FTB_Summary: Contains data for Family Tax Benefit (FTB) summarised aggregated at financial-year level based on a participant’s reconciled eligibility and entitlement determined after receipt of their taxable income provided by the ATO. Information is only provided up to two years prior to the extraction date at which point the data is considered ‘mature’; that is, the vast majority have tax data against which their entitlement can be reconciled. The information provided includes:
- no. of days the participant was eligible for FTB (in total), FTB-A and FTB-B
- no. of days the participant was eligible for an ISP while eligible for FTB
- no. of days customer was partnered with a primary partner while eligible for FTB
- no. of days the participant was partnered with ex-partners while eligible for FTB
- count of children assessed as FTB children
- total validated adjusted taxable income (customer + primary partner + ex-partners).

Concession_cards: Contains episodes of concession cards data for participants where a participant held a concession card. As a participant can have multiple concession cards during the same time duration, in some cases this file has overlapping episodes of concession cards for a participant. Information includes the benefit type that qualified them for a concession, the concession card type and the number of dependent children.

Restricted release Centrelink files

ISP_Episodic: Holds the information for each episode of ISP receipt. In addition to the variables in the summary file the following information is provided:
- entitlement rate
- activity requirements
- reason for end of payment
- earnings amount and work hours
- educational details – student status, course level and type, highest educational level before episode
- rent amount
- homelessness
- medical conditions (currently a binary indicator pending confidentialisation) and impairment rating
- vulnerability indicator.

FTB_Customer_Reconciled: Has the same structure as the ISP_Summary file. Additional information provided includes:
- age, citizenship, Indigenous indicator, overseas indicator, preferred written language, remoteness area
- no. of days eligible for FTB-A (by rate type)
- no. of days eligible for of FTB-B
- FTB-A and FTB-B pre-reconciliation eligibility amounts (paid and notional)
- FTB-A and FTB-B post-reconciliation entitlement amounts
- maintenance income and amount of FTB-A not paid due to MI test
- no. of days overseas
- count of FTB shared care children
- no. of days also eligible for an ISP
- adjusted taxable income broken down by components.

FTB_Child_Reconciled: This file holds the reconciled data for the FTB children for which a participant received FTB payments in an entitlement year. The data contains one observation for each FTB customer – FTB child combination for each entitlement year during which the participant/customer received FTB payment for the corresponding child. Details for children aged 16 or over are not included due to privacy considerations. Information includes:
- age, gender, overseas indicator and duration
- post-reconciliation durations for FTB-A and FTB-B
- regular and shared care durations
- FTB-A supplement amount.
5.4 Child Health CheckPoint data

A comprehensive, one-off physical health and biomarker module, known as the Child Health CheckPoint, was added for the B cohort between LSAC Waves 6 and 7. B cohort families who took part in a LSAC Wave 6 home interview were eligible for the Child Health CheckPoint module. In 2015–16, the B cohort child and one of their parents participated in a comprehensive clinic appointment or shorter home visit. A second parent was also invited to provide a genetic sample. The study child was aged 11–12 years at the time of assessment. The aim of this additional phase was to learn more about the health of young Australians between childhood and adolescence. Contents of the general release, restricted release and supplementary files is available in the Child Health Checkpoint Data User Guide. Further information about Child Health CheckPoint is available from the study website: mcri.edu.au/research/projects/longitudinal-study-australian-childrens-child-health-checkpoint

Ideally, a physical health and biomarker module would have been offered to both B and K cohorts. However, because the CheckPoint was funded by a national competitive grant scheme, there were only sufficient funds to assess one of the two LSAC cohorts. The B cohort was chosen over the K cohort because the younger cohort has early-life data collected prospectively; were commencing puberty, which was important to many CheckPoint measures; and were at an age where the study children were less likely to become disengaged or too busy to participate.

During the LSAC Wave 6 home visit, the interviewer briefly introduced the Child Health CheckPoint and collected written consent to pass their contact details to the CheckPoint team solely for purposes of recruitment to the CheckPoint module. The majority of the Wave 6 interviews took place in March–September 2014. Permission for contact was received from 3,513 families (93% of Wave 6 families and 69% of the original cohort). See mcri.edu.au/research/projects/longitudinal-study-australian-childrens-child-health-checkpoint
6 Variable naming conventions

The variable naming convention was developed so that variables have predictable names across waves and informants, and so that thematically linked variables have similar names wherever possible. A two-page ‘help sheet’ is provided in this Data User Guide (see Appendix A) to help users learn these conventions.

6.1 Questionnaire variables

Variable names follow the standard format in most cases. Exceptions to this naming convention (derived items and household composition variables) are explained in the sections that follow.

Standard format: A tt xxxxx

Where:

A = child age indicator
tt = topic indicator
xxxxx = specific question identifier.

6.1.1 Child age indicator (alpha)

The child age indicator is an alpha symbol that indicates the child’s age, allowing for comparisons between the cohorts where data have been collected for both cohorts at that age. For instance:

- a indicates the child is aged 0–1 years (which is the B cohort in Wave 1)
- b indicates the child is aged 2–3 years (which is the B cohort in Wave 2)
- c indicates the child is aged 4–5 years (which is the B cohort in Wave 3, and the K cohort in Wave 1)
- d indicates the child is aged 6–7 years (which is the B cohort in Wave 4, and the K cohort in Wave 2)
- e indicates the child is aged 8–9 years (which is the B cohort in Wave 5, and the K cohort in Wave 3)
- f indicates the child is aged 10–11 years (which is the B cohort in Wave 6, and the K cohort in Wave 4)
- g indicates the child is aged 12–13 years (which is the B cohort in Wave 7 and the K cohort in Wave 5)
- h indicates the child is aged 14–15 years (which is the K cohort in Wave 6)
- i indicates the child is aged 16–17 years (which is the K cohort in Wave 7)

This is an example of how the child age indicator is used for the item ‘Parent 1 rating of parenting self-efficacy’:

- Wave 1 B cohort: apa01a
- Wave 2 B cohort: bpa01a
- Wave 3 B cohort: cpa01a
- Wave 4 B cohort: dpa01a
- Wave 5 B cohort: epa01a
- Wave 6 B cohort: fpa01a
- Wave 7 B cohort: gpa01a
- Wave 1 K cohort: cpa01a
- Wave 2 K cohort: dpa01a
- Wave 3 K cohort: epa01a
Wave 4 K cohort: fpa01a
Wave 5 K cohort: gpa01a
Wave 6 K cohort: hpa01a
Wave 7 K cohort: ipa01a

Those items of information that do not change (e.g. details of birth, age child began or stopped something, etc.) are given the age indicator \( z \) so that they have a consistent variable name across cohorts regardless of the age of the child when the information was obtained. For example, \( zhs03a \) indicates ‘birth weight of the study child’ regardless of whether the information was collected when the child was aged 0–1 years, as for the B cohort, or aged 4–5 years, as for the K cohort.

### 6.1.2 Topic indicator (alpha)

The topic indicator is taken from the topic field of the data dictionary. Efforts were made to make the abbreviations used meaningful (e.g. family demographics is \( \text{fd} \)).

A list of topics and their abbreviations is provided in Table 7.

**Table 7: Topics used in LSAC datasets**

<table>
<thead>
<tr>
<th>Abbrev.</th>
<th>Topic</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>ce</td>
<td>Centrelink data</td>
<td>Statistical information about payments and services</td>
</tr>
<tr>
<td>fad</td>
<td>Family demographics</td>
<td>Demographic information relating to the family such as education, ethnicity and religion</td>
</tr>
<tr>
<td>fn</td>
<td>Finances</td>
<td>Financial information such as income and use of government benefits</td>
</tr>
<tr>
<td>ed</td>
<td>Education</td>
<td>Scales that measure the effect of study on parenting</td>
</tr>
<tr>
<td>gd</td>
<td>General development</td>
<td>Scales that contain items from multiple domains of child development</td>
</tr>
<tr>
<td>hb</td>
<td>Health behaviour and risk factors</td>
<td>Behaviours and other risk factors that potentially impinge upon the health of the study child or his/her family. These include behaviours such as parental smoking and drinking as well as risk factors such as a parent experiencing diabetes during pregnancy.</td>
</tr>
<tr>
<td>he</td>
<td>Home education environment</td>
<td>Information on factors likely to impinge on the child’s learning while at home such as parental support for education, number of books in the home and TV use. Also contains information on parent interaction with teachers such as parent teacher interviews including from the teacher’s perspective</td>
</tr>
<tr>
<td>ho</td>
<td>Housing</td>
<td>Information on housing such as number of bedrooms, tenure type and payments</td>
</tr>
<tr>
<td>hs</td>
<td>Health status</td>
<td>Information about the physical and mental health status of the study child or his/her family such as body mass index, diagnosis of conditions and number of hospital stays</td>
</tr>
<tr>
<td>id</td>
<td>Identifiers</td>
<td>Questionnaire process variables such as sequence guides, consents and details of proxy respondents</td>
</tr>
<tr>
<td>lc</td>
<td>Learning and cognition outcomes</td>
<td>Information on the child’s development in the areas of learning and cognition including language, literacy and numeracy</td>
</tr>
<tr>
<td>pa</td>
<td>Parenting</td>
<td>Information on parenting styles and other information affecting parenting such as self-efficacy</td>
</tr>
<tr>
<td>pc</td>
<td>Program characteristics</td>
<td>Characteristics of the educational or child care program such as type of program, number of days or hours the child attends and staff satisfaction</td>
</tr>
<tr>
<td>pe</td>
<td>Parent living elsewhere</td>
<td>Details of the child’s PLE such as the relationship to study child, interactions with resident parents and child support</td>
</tr>
<tr>
<td>pl</td>
<td>Parental leave in Australia</td>
<td>Data from the Parental Leave in Australia Survey – a nested study</td>
</tr>
<tr>
<td>pw</td>
<td>Paid work</td>
<td>Information on work status such as employment, occupation and work/family interactions</td>
</tr>
<tr>
<td>re</td>
<td>Relationships</td>
<td>Information on the quality of relationships primarily focused on the relationship between Parent 1 and Parent 2, but also on broader family harmony</td>
</tr>
<tr>
<td>sc</td>
<td>Social capital</td>
<td>Information on social capital such as attitudes to neighbours and the neighbourhood and use of services</td>
</tr>
<tr>
<td>se</td>
<td>Social and emotional outcomes</td>
<td>Information relevant to the social and emotional development of the child such as temperament, behaviour and emotional states</td>
</tr>
<tr>
<td>tp</td>
<td>Teaching practices</td>
<td>Practices employed by teachers and child care workers in their work such as time use, use of resources and general philosophies</td>
</tr>
</tbody>
</table>
For example:
apa01a (PI rating of self-efficacy) has ‘pa’ as the second and third letters as its topic is ‘Parenting’; and
zhso3a (Birth weight of study child) has ‘hs’ as the second and third letter as its topic is ‘Health status’.

### 6.1.3 Specific question identifier (alphanumeric)

The last five digits of a variable name make up the specific question identifier (if required). These digits contain whatever information is necessary to uniquely identify each item. Each has an arbitrary two-digit question number, not related to the questionnaire positioning. Items of related content are grouped together as much as possible.

For example:
bhs12a is whether P1 is concerned about the child’s weight
bhs12b is whether P1 considers the child to be ‘underweight’, ‘normal weight’, ‘somewhat overweight’ or ‘very overweight’.

The sixth digit of the variable name can also be an informant or subject indicator where a question is asked of or about more than one person. The indicators used are:

- **a** Parent 1
- **b** Parent 2
- **c** Study child
- **m** Mother
- **f** Father (or family home for census data)
- **t** Teacher/Carer
- **i** In-between waves respondent
- **p** Parent living elsewhere
- **y** Study Child Offspring (ya-1st offspring, yb-2nd offspring and yc-3rd offspring)
- **x** Other biological parent of the Study Child Offspring (xa – Other biological parent of 1st Child, xb – Other biological parent of 2nd Child, xc – Other biological parent of 3rd Child)

For example:
bhs13a is Parent 1’s rating of their own overall health status.
bhs13b is Parent 2’s rating of their own overall health status.
bhs13c is Parent 1’s rating of the study child’s overall health status.
bhs13p is the PLE’s rating of their own overall health status.
bhs13m is the mother’s rating of their own overall health status.
bhs13f is the father’s rating of their own overall health status.

An exception to the above rule is in the area of child care and education (variables with topic indicators pc and tp). Here the prefixes a, b, c, d and e are used to mean different things at each wave depending on the options available to the child at that age (see Table 8).

All items that form a scale have a single question number. Where applicable, the name of the item also indicates the relevant subscale or sub-subscale (please note that this is done only where it is possible to do so, due to the eight-character limit for the name of an item).

An example of how this is applied is shown with the Conduct Problems and Peer Problems subscales of the Strengths and Difficulties Questionnaire (see Table 9). These are subscales that both P1 and the teacher filled out in Waves 1 and 2 for the K cohort.

As shown:
- The 6th character in the variable name in this case represents an informant indicator: ‘a’ is for Parent 1, ‘t’ is for teacher.
- The 7th character indicates the subscale: 4 for Conduct, 5 for Peer. (Note: the subscales 1 for Prosocial, 2 for Hyperactivity and 3 for Emotional are also available as part of the SDQ.)
- The final character uniquely identifies each item. (Note: different items were used for the Conduct subscale in waves 1 and 2 due to the change in the child’s age).
Table 8: Subject indicators for education and childcare variables

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Age 0-1</th>
<th>Age 2-3</th>
<th>Age 4-5</th>
<th>Age 6-7</th>
<th>Age 8-9</th>
<th>Age 10-11</th>
<th>Age 12-13</th>
<th>Age 14-15</th>
<th>Age 16-17</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1st child care</td>
<td>1st child care</td>
<td>Main educational program</td>
<td>Main educational program</td>
<td>Main educational program</td>
<td>Main educational program</td>
<td>Main educational program</td>
<td>Main educational program</td>
<td>Main educational program</td>
</tr>
<tr>
<td>c</td>
<td>3rd child care</td>
<td>3rd child care</td>
<td>2nd child care</td>
<td>After-school care</td>
<td>After-school care</td>
<td>After-school care</td>
<td>After-school care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>Other child care</td>
<td>3rd child care</td>
<td>Child care at other times</td>
<td>Other child care</td>
<td>Other child care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>Program child would attend if attending school</td>
<td>Program child would attend if attending school</td>
<td>Program child would attend if attending school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o</td>
<td>Any extra care</td>
<td>Any extra care</td>
<td>Any extra care</td>
<td>Any extra care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 6: Variable naming conventions

### Table 9: Variable names of SDQ² conduct and peer problems subscales

<table>
<thead>
<tr>
<th>Conduct problems</th>
<th>Wave 1 Parent 1 K cohort name</th>
<th>Wave 1 Teacher K cohort name</th>
<th>Wave 2 Parent 1 K cohort name</th>
<th>Wave 2 Teacher K cohort name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Often loses temper</td>
<td>cse03a4a</td>
<td>cse03t4a</td>
<td>dse03a4a</td>
<td>dse03t4a</td>
</tr>
<tr>
<td>Generally, well behaved, usually does what adults request</td>
<td>cse03a4b</td>
<td>cse03t4b</td>
<td>dse03a4b</td>
<td>dse03t4b</td>
</tr>
<tr>
<td>Often fights with other children or bullies them</td>
<td>cse03a4c</td>
<td>cse03t4c</td>
<td>dse03a4c</td>
<td>dse03t4c</td>
</tr>
<tr>
<td>Often argumentative with adults</td>
<td>cse03a4d</td>
<td>cse03t4d</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Can be spiteful to others</td>
<td>cse03a4e</td>
<td>cse03t4e</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Often lies or cheats</td>
<td>N/A</td>
<td>N/A</td>
<td>dse03a4f</td>
<td>dse03t4f</td>
</tr>
<tr>
<td>Steals from home, school or elsewhere</td>
<td>N/A</td>
<td>N/A</td>
<td>dse03a4g</td>
<td>dse03t4g</td>
</tr>
</tbody>
</table>

#### Peer problems

<table>
<thead>
<tr>
<th>Peer problems</th>
<th>Wave 1 Parent 1 K cohort name</th>
<th>Wave 1 Teacher K cohort name</th>
<th>Wave 2 Parent 1 K cohort name</th>
<th>Wave 2 Teacher K cohort name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rather solitary, tends to play alone</td>
<td>cse03a5a</td>
<td>cse03t5a</td>
<td>dse03a5a</td>
<td>dse03t5a</td>
</tr>
<tr>
<td>Has at least one good friend</td>
<td>cse03a5b</td>
<td>cse03t5b</td>
<td>dse03a5b</td>
<td>dse03t5b</td>
</tr>
<tr>
<td>Generally liked by other children</td>
<td>cse03a5c</td>
<td>cse03t5c</td>
<td>dse03a5c</td>
<td>dse03t5c</td>
</tr>
<tr>
<td>Picked on or bullied by other children</td>
<td>cse03a5d</td>
<td>cse03t5d</td>
<td>dse03a5d</td>
<td>dse03t5d</td>
</tr>
<tr>
<td>Gets on better with adults than with other children</td>
<td>cse03a5e</td>
<td>cse03t5e</td>
<td>dse03a5e</td>
<td>dse03t5e</td>
</tr>
</tbody>
</table>

Note: * The SDQ is copyrighted by Robert Goodman, UK, 1999.

### 6.2 Derived variables

The derived items start with an age indicator, as outlined in section 6.1.1, followed by an informant or subject indicator and then a mnemonic that relates to the subject matter of the derived item. For example, the Peer subscale of the SDQ for the K cohort teacher in Wave 2 is dtpeer, where d = child aged 6–7 years, t = teacher and peer = Peer subscale of SDQ.

### 6.3 Study child household composition variables

In order to keep the variable names under eight characters, it was necessary to have a slightly different convention in the Wave 2 data release. Household composition variables have the following structure:

\[ A\ f\ ##xmmm \]

Where:

- \( A = \) Child age indicator
- \( f = f \) (for ‘family’)
- \( ## = \) Question number (numeric)
- \( x = \) Sub-question indicator (optional)
- \( mmm = \) person identifier

Note:

- The age indicator above is as described in section 6.1.1.
- ‘f’ is a constant to indicate that it is the household composition that is being described.
- The question number and sub-question indicator indicate the question being responded to.

The person identifier indicates the member number, or other identification information. For every household, the study child is member 1, the Wave 1 P1 is member 2, and the Wave 1 P2 is member 3 (or will be missing if there is no P2 at Wave 1). Any additional people in the household at the time of Wave 1 are given member numbers 4 through to whatever is required. Each household member retains the same member number throughout the study, even if they leave and re-enter the study child’s home.
Due to the requirements of the CAI instrument, some families have ‘gaps’ in member numbering; for example, where someone is Member 5 but Member 4 has never been assigned.

Member 1 is denoted by ‘m1’ in the above convention, member 2 as ‘m2’ and so on as required.

As families change from Wave 2 on, the new P1, P2, mother or father could have any member number apart from 1. For this reason, an extra set of variables has been derived to give the details for the P1, P2, mother and father at any age. This subscript is an age indicator and then either ‘p1’, ‘p2’, ‘m’, or ‘f’.

A set of indicator variables tracks the household member number of P1, P2, mother and father at each wave. For example, bp2mn tells you the household member number of P2 when the child is aged 2–3, while cmmn gives the member number of the mother when the child is aged 4–5.

Some examples:

- zf02m1 is the gender of the study child (z = unchanging characteristic, f = ‘Family’, 02 = gender, m1 = study child)
- bf01m2 is whether the Wave 1 P1 is present in the household when the child is aged 2–3 (b = child aged 2–3, f = ‘family’, 01 = present for wave, m2 = Wave 1 P1)
- cf03m3 is whether the Wave 1 P2 is present when the child was aged 4–5 (or whether there was a P2 at all in Wave 1 for the K cohort) (c = child aged 4–5, f = ‘family’, 01 = present for wave, m3 = Wave 1 P2)
- af08am is the relationship of the mother to the study child when the child was aged 0–1 (a = ages 0–1, f = ‘family’, 08 = relationship to study child, am = mother of child at age 0–1)
- df01cp1 is whether the P1 of the child when aged 4–5 is present in the household when the child is aged 6–7. (d = child aged 6–7, f = ‘family’, 01 = present for wave, cp1 = child’s P1 when child is aged 4–5)
- cf13dp2 is whether the P2 of the child when aged 6–7 had a medical condition or disability at the time the child was 4–5 (c = child aged 4–5, f = ‘family’, 13 = whether person has a disability, dp2 = P2 when child is aged 6–7).

Table 10 shows the information that is available for each person.

### Table 10: Question numbers used in variable names for household member characteristics

<table>
<thead>
<tr>
<th>Question number</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Present for wave</td>
</tr>
<tr>
<td>02</td>
<td>Gender</td>
</tr>
<tr>
<td>03</td>
<td>Age</td>
</tr>
<tr>
<td>04</td>
<td>Date of birth</td>
</tr>
<tr>
<td>05</td>
<td>Temporarily away from home (as per Wave 1 question)</td>
</tr>
<tr>
<td>06</td>
<td>Relationship to Parent 1</td>
</tr>
<tr>
<td>07</td>
<td>Relationship to Parent 2</td>
</tr>
<tr>
<td>08</td>
<td>Relationship to study child</td>
</tr>
<tr>
<td>08z</td>
<td>Relationship to study child partner</td>
</tr>
<tr>
<td>09</td>
<td>Country of birth</td>
</tr>
<tr>
<td>10</td>
<td>Year of first arrival in Australia</td>
</tr>
<tr>
<td>11</td>
<td>Language other than English spoken at home</td>
</tr>
<tr>
<td>12</td>
<td>ATSI status</td>
</tr>
<tr>
<td>13</td>
<td>Has a condition or disability for six months or more (as per Wave 1 question)</td>
</tr>
<tr>
<td>13a</td>
<td>1st specific condition</td>
</tr>
<tr>
<td>13b</td>
<td>2nd specific condition</td>
</tr>
<tr>
<td>14</td>
<td>Date stopped living with study child</td>
</tr>
<tr>
<td>15</td>
<td>Reason stopped living with study child</td>
</tr>
<tr>
<td>16</td>
<td>Temporarily away from home (as per Wave 2 question)</td>
</tr>
<tr>
<td>16o</td>
<td>Temporarily away from home (other) (as per Wave 2 question)</td>
</tr>
<tr>
<td>17</td>
<td>Has a condition or disability for six months or more (as per Wave 2 question)</td>
</tr>
</tbody>
</table>

Table continued on next page
<table>
<thead>
<tr>
<th>Question number</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>17a</td>
<td>Has sight problems (as per Wave 2 question)</td>
</tr>
<tr>
<td>17b</td>
<td>Has hearing problems (as per Wave 2 question)</td>
</tr>
<tr>
<td>17c</td>
<td>Has speech problems (as per Wave 2 question)</td>
</tr>
<tr>
<td>17d</td>
<td>Has blackouts, etc. (as per Wave 2 question)</td>
</tr>
<tr>
<td>17e</td>
<td>Has difficulty learning (as per Wave 2 question)</td>
</tr>
<tr>
<td>17f</td>
<td>Limited use of arms or fingers (as per Wave 2 question)</td>
</tr>
<tr>
<td>17g</td>
<td>Difficulty gripping (as per Wave 2 question)</td>
</tr>
<tr>
<td>17h</td>
<td>Limited use of legs and feet (as per Wave 2 question)</td>
</tr>
<tr>
<td>17i</td>
<td>Other physical condition (as per Wave 2 question)</td>
</tr>
<tr>
<td>17j</td>
<td>Other disfigurement (as per Wave 2 question)</td>
</tr>
<tr>
<td>17k</td>
<td>None of the above conditions (as per Wave 2 question)</td>
</tr>
<tr>
<td>18</td>
<td>Restricted in everyday activities</td>
</tr>
<tr>
<td>18a</td>
<td>Has difficulty breathing (as per Wave 2 question)</td>
</tr>
<tr>
<td>18b</td>
<td>Has chronic pain (as per Wave 2 question)</td>
</tr>
<tr>
<td>18c</td>
<td>Has nervous condition requiring treatment (as per Wave 2 question)</td>
</tr>
<tr>
<td>18d</td>
<td>Has mental illness requiring supervision (as per Wave 2 question)</td>
</tr>
<tr>
<td>18e</td>
<td>Has head injury (as per Wave 2 question)</td>
</tr>
<tr>
<td>18f</td>
<td>Has other long-term condition (as per Wave 2 question)</td>
</tr>
<tr>
<td>18g</td>
<td>Has other condition requiring treatment (as per Wave 2 question)</td>
</tr>
<tr>
<td>18h</td>
<td>None of the above restrictions (as per Wave 2 question)</td>
</tr>
<tr>
<td>19</td>
<td>Date began living with the study child</td>
</tr>
<tr>
<td>20</td>
<td>Household member was in the household for at least three months but moved in and left between current and previous waves</td>
</tr>
<tr>
<td>21</td>
<td>Person type</td>
</tr>
<tr>
<td>22</td>
<td>Young carer activities</td>
</tr>
<tr>
<td>23</td>
<td>Migration status</td>
</tr>
</tbody>
</table>

6.4 PLE household composition variables

From Wave 4, the household information for the child’s parent living elsewhere (PLE) has been collected. PLE household composition variables have a similar structure to that of the study child household composition variables:

A f ##xple#

Where:

- A = Child age indicator
- f = f (for ‘family’)
- ## = Question number (numeric)
- x = Sub-question indicator (optional)
- ple## = person identifier within PLE household with ple (for Parent Living Elsewhere) and # member number

Note:

The age indicator is as described in section 6.1.1.

‘f’ is a constant to indicate that it is the household composition that is being described.

The question number and sub-question indicator indicate the question being responded to.

The person identifier comprises the constant ‘ple’ to indicate that it is the PLE household and the member number. For every PLE household, the study child is member 1 (ple1) and PLE is member 2 (ple2). For example, variable f02ple2 refers to a PLE gender when a study child is 10–11 years old. Any additional member in the household is assigned a PLE member number that remains the same throughout the study, even if they leave and re-enter the PLE’s home.
Table 11 shows the information that is available for each PLE.

**Table 11: Question numbers used in variable names for PLE household member characteristics**

<table>
<thead>
<tr>
<th>Question number</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Present for wave</td>
</tr>
<tr>
<td>02</td>
<td>Gender</td>
</tr>
<tr>
<td>03</td>
<td>Age</td>
</tr>
<tr>
<td>04</td>
<td>Date of birth</td>
</tr>
<tr>
<td>05</td>
<td>Temporarily away from home (as per Wave 1 question)</td>
</tr>
<tr>
<td>06a</td>
<td>Relationship to PLE</td>
</tr>
<tr>
<td>08</td>
<td>Relationship to study child</td>
</tr>
<tr>
<td>09</td>
<td>Country of birth</td>
</tr>
<tr>
<td>10</td>
<td>Year of first arrival in Australia</td>
</tr>
<tr>
<td>11</td>
<td>Main language spoken at home</td>
</tr>
<tr>
<td>12</td>
<td>ATSI status</td>
</tr>
</tbody>
</table>

A PLE household file also includes the following variables (the asterisk refers to the child age indicator):

- *datplec – date of PLE CATI interview*
- *plepar – whether PLE has a partner*
- *pleparmn – PLE partner member number in PLE household*
- *dfd02p3 – date of recent PLE marriage*
- *dfd02p4 – date of PLE cohabitation.*

### 6.5 Age invariant indicator variables

There are five variables at the start of each of the main data files that contain no age indicator. These are:

- hicid – unique identifier assigned when child was selected by Medicare Australia
- cohort
- wave
- stratum – stratum at the time of selection
- pcodes – postcode at the time of selection

Users wishing to create long datasets should note the presence of these variables when removing age indicators.

### 6.5.1 Study child unique identifier

Each study child has a single, unique identification variable to ensure matching and merging across instruments, files and waves. This number was allocated at the time of selection by Medicare Australia.

The first digit indicates which cohort the child is in (1–4 = Infant; 5–8 = Child) and what fieldwork phase (see ‘Methodology’ section for more detail) the child was selected to be part of in Wave 1 (phase 1 = 1 and 5, phase 2 = 2 and 6, etc.).

The second is the state the child was selected from (1 = NSW, 2 = Vic., etc.).

The third indicates the part of the state the child was selected from (1=2 = capital city; 3–4 = rest of state).

The remaining five digits are a random number allocated by Medicare Australia.

Note that the stratum for selection may differ from the location of the child at interview and that the fieldwork phase may change from wave to wave.
6.6 Indicator variables

There are indicator variables in the main data files that indicate which parts of an interview were incomplete. These variables were created to flag to data users (through yes/no values) that no data, or only partial data, exists for an instrument (e.g. the CASI) or an informant (e.g. Parent 1). The data may be incomplete due to a number of different reasons. There may be no data if a self-complete form was not returned; the parent/child did not provide consent to obtain/provide the data; one of the informants refused to participate; or when the interview was only partially completed.

For example, on the day of the interview the parent may consent to the child participating but refuse to participate themselves. In this example, there would be data for the sections where the study child is the informant; however, there would be no data for the sections where P1 is the informant. To identify these cases a data user can use the following indicator variable *nopar (* refers to the age indicator). Another example is a teacher’s responses. To identify cases where a teacher form was not returned, a data user can examine the variable *tcd. A data user can also examine the following indicator variables: *partresp to identify cases that were incomplete due to an interview stopping half way as opposed to just certain sections being refused or *hhresp to identify cases where the household interview was completed.

There are a large number of indicator variables. Data users are encouraged to investigate the reasons for data being incomplete through these variables. Note that the indicator variables do not follow the general variable naming conventions described above. Some indicator variables are listed in Table 1. Indicator variables can be found in the data dictionary under the topic ‘Identifiers’, along with other variables that fall under that topic. For more information refer to the data dictionary.

6.7 Variable labelling convention

The labels used for the variable dataset take the following general form:

(Age) - (Informant/subject) - (Questionnaire position) - (Construct label)

Age is a label for the age indicator from the variable name, so:

- a = 0/1
- b = 2/3
- c = 4/5
- d = 6/7
- e = 8/9
- f = 10/11
- g = 12/13
- h = 14/15
- i = 16/17

If no age indicator is present in the variable name, or the age indicator is z, then this part of the variable label will not be included.

For example:

- label zf04m1 = ‘SC - DOB’, here no age is associated with the variable because it doesn't change with time, hence no age indicator is included.
- label df03m1 = ‘6/7 - SC - Age’, this variable is a variable that changes over time so the age indicator is required in order to establish when the question was answered.

Informant/subject gives the informant or subject of the question as contained in the variable name. For household composition variables involving P1, P2, mother or father, the age of the study child at which the person’s status as parent is determined will also be indicated (e.g. M@0/1 is the mother when the child is aged 0–1 years old). If the information only exists for one subject or informant in the study this part of the variable label will not be included.

Questionnaire position indicates the location of the question the data was obtained from within the LSAC questionnaires (e.g. F2F H2 is question H2 of the face-to-face interview). This part of the variable label is left blank for derived items such as scales and other non-input items but included for mother/father variables where the location of both the P1 and the P2 variables are given.
Construct label provides a description of what information is actually contained in the variable (e.g. ‘Sex’, ‘Birthweight’, etc.). This part of the variable name will be consistent for each variable representing the same construct for a different subject/informant or wave.

For example:

The Parent 1’s rating of their own health quality at Wave 1 for the B cohort (ahs13a) has the variable label ‘0/1 - P1 - PIL D1 - Global Health Measure’: (0/1 is the age indicator, P1 is the informant/subject indicator, PIL D1 indicates the variable comes from the first question of section D of the P1 leave-behind questionnaire, ‘Global Health Measures’ is the construct label).

Total score for the P1 parental warmth scale for the K cohort at Wave 2 (dbwarm) id ‘6/7 - P2 - warm parenting’ (6/7 is the age indicator, P2 is the informant indicator, there is no questionnaire position as the variable is calculated from multiple questions, ‘warm parenting’ is the construct label).

### 6.8 Missing value conventions

Missing data are coded as follows:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>Not applicable (when explicitly available as an option in the questionnaire)</td>
</tr>
<tr>
<td>-2</td>
<td>Don’t know</td>
</tr>
<tr>
<td>-3</td>
<td>Refused or not answered</td>
</tr>
<tr>
<td>-4</td>
<td>Section refused</td>
</tr>
<tr>
<td>-9</td>
<td>Not asked due to one of the following reasons:</td>
</tr>
<tr>
<td></td>
<td>(a) A question was skipped due to the answer to a preceding question (e.g. if a child never repeated a grade, the following question regarding what grade the child repeated was not asked/skipped).</td>
</tr>
<tr>
<td></td>
<td>(b) A form was not returned or consent to participate was not given (e.g. if a teacher form was not returned then the teacher’s responses for this hicid are set to -9. To identify cases for which a form was not returned/or consent was not provided a data user can use an indicator variable (see Table 1 for details)).</td>
</tr>
<tr>
<td></td>
<td>(c) One of the informants refused to participate (e.g. if a parent refused to participate but not a child then the parent’s responses are set to -9. To identify cases when the parent refused to participate, a data user can use the *nopar indicator variable).</td>
</tr>
<tr>
<td></td>
<td>(d) A form was partially completed (e.g. P1 completed the interview over the phone (P1 CATI) but the face-to-face component did not occur. To identify these cases, a data user can use the *partresp indicator variable). (See section 6.6 for more detail.)</td>
</tr>
<tr>
<td>-99</td>
<td>Missing data – data not collected where it might be expected (e.g. the respondent skipped a question they should have answered in a self-complete form), or made missing due to an unreliable value (e.g. weight of P1 recorded as 800 kg).</td>
</tr>
</tbody>
</table>

Other specific examples of (-99)

(a) Negative income (loss)

(b) Before baby’s birth

(c) No set amount

For further details about how missing LSAC income data is treated see Technical Paper No. 14 *Imputing income in the Longitudinal Study of Australian Children* available at growingupinaustralia.gov.au/data-and-documentation/technical-papers
A number of tools can be used to navigate the LSAC dataset:

- marked-up (or labelled) instruments
- frequencies
- online LSAC data dictionary
- rationale document
- Excel spreadsheets of the data dictionary (good for creating hard copies).

Users should also consider which documents they want to print out and which they want to look at electronically. We have found that the marked-up questionnaires and interview specifications are best printed and provide the easiest method of browsing to familiarise yourself with the data available. The data dictionary is best used for searching for specific items and mapping items from wave to wave.

These tools are described in more detail below.

7.1 Marked-up instruments

The associated variable name has been added beside each question in the questionnaires and interview specifications. Derived variables are also included. See Figure 3 for an example.

**Figure 3:** Examples of the marked-up questionnaires

A mock questionnaire (interview specifications) has also been generated for the CASI and CAI instruments used in Waves 2–6. Figure 4 is a sample of this.
7.2 Frequencies

The frequencies are a listing of the response categories for each question and the number of cases in each category. Table 12 provides an example of the listing.

Table 12: Example of the weighted frequencies

<table>
<thead>
<tr>
<th>hhs5S5c</th>
<th>Frequency</th>
<th>Percentage (%)</th>
<th>Cumulative frequency</th>
<th>Cumulative percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-9</td>
<td>202.5867</td>
<td>5.37</td>
<td>202.5867</td>
<td>5.73</td>
</tr>
<tr>
<td>-3</td>
<td>23.94586</td>
<td>0.68</td>
<td>226.5325</td>
<td>6.40</td>
</tr>
<tr>
<td>No</td>
<td>1,379.145</td>
<td>38.99</td>
<td>1,605.678</td>
<td>45.40</td>
</tr>
<tr>
<td>Yes</td>
<td>1,931.322</td>
<td>54.60</td>
<td>3,537</td>
<td>100.00</td>
</tr>
</tbody>
</table>

The frequencies are useful for simple queries related to particular questions (e.g. how many of the births had a normal delivery, or what are the codes used for Wave 1 question A15). Variables for which there were a wide variety of responses, meaning unaltered frequencies would run for several pages (e.g. study child weight), have been rounded off to enable the grouping of responses.

7.3 Data dictionary

This is available as both an online version and in Excel. Both versions of the data dictionary are searchable and can be sorted. Each record describes a single variable and has the following fields:

- variable name
- variable name without age (useful for sorting)
- topic number (allows derived items to be sorted in with the input variables they come from)
- question id (i.e. variable name without age or subject/informant, useful for sorting)
- file (each of the main datasets are allocated a file name that denotes the cohort and age of the study child at each wave (i.e. Wave 1 = files B0 and K4, Wave 2 = files B2 and K6, Wave 3 = files B4 and K8, etc.))
- position in file order (the order of the variables in the files)
- wave
- cohort
- position of question in questionnaires
- person label
- child’s age
- variable label briefly describing each data item
- topic
- construct
Chapter 7: Documentation

- measure
- question as found in the survey instruments
- response categories
- population with data
- SAS format
- notes field indicating other information about the data item users should know.

7.3.1 Excel data dictionary

The Excel data dictionary contains two spreadsheets, one with the complete detailed listing of variable attributes, another with a shorter listing in a print-ready format. The print-ready format contains the variable name, question, responses and population fields; however, it is not a difficult task for users to make their own printable versions if they prefer other fields.

The Excel version can be easily filtered using the drop-down menus in the first row of the spreadsheet. For example, to find all of the items on teacher practices in the lsacgr6 file (K cohort at Wave 2) first click on the drop-down menu in the 'File' field as shown in Figure 5 and select 'B2'. Next, repeat the process for the 'Topic' field, selecting 'Teaching practices'.

After the search is finished all variables can be displayed by either clicking the 'show all' option in each of the fields that have been filtered (see Figure 5) or by selecting 'Data > Filter > Show All' from the menus.

More advanced searches can be performed using the 'Custom Filter' option, which produces a dialogue box to assist with your searching. For example, to find all the questions that contain the word 'internet', go to the 'question' column and open up the filter menu and click on 'Custom filter', in the dialogue box change 'equals' to 'contains' and type 'internet' next to this.

Figure 5: Example of filtering in Excel
7.3.2 Using wildcards for filtering

A good understanding of the variable naming convention is valuable for using the data dictionary. Both the online and Excel versions of the data dictionary can be searched and filtered using wildcards, which can be used to return thematically linked sets of variables. Two wildcard characters are used by both these programs:

* represents any combination of letters and characters

? represents any single character

Some examples of the use of these characters are as follows:

- `apw23a*` returns a range of variables `apw23a1a` through to `apw23a4b`.
- `apw23a4?` returns two variables `apw23a4a` and `apw23a4b`.
- `?pw23a4a` shows if this variable exists over different waves.
- `apw23?4a` shows if this variable exists for different people in the same wave.
- `?pw23?4a` shows if this variable exists for different people in different waves.

7.3.3 Some useful tips for navigating the data dictionary

- Only items currently on the main datasets are included in the data dictionary. The User Guide provides information on the composition of other datasets.
- Items on the data dictionary are in the same order as on the data files but can easily be sorted into other orders, for example, grouping topics.
- Searching the online data dictionary finds whole words (e.g., searching for ‘child’ won’t find ‘children’ as well). However, an asterisk will represent any combination of characters. So, searching for ‘child*’ will find ‘child’, ‘children’, ‘childcare’, etc.
- The introduction page for the data dictionary contains a list of topics and constructs that can be used for finding the information you want.
- The ‘Question ID’ field gives the variable name without any wave or person indicators. Filtering by this field is the best way to tell which questions were asked of or about which people at which wave.
- The ‘Topic ID’ field gives the topic and associated two-digit question number for each item where this is appropriate. It can be used to link derived items with their associated input items.

Please contact the LSAC Data Management team if you need any help with using the data dictionaries.

---

7 The data dictionary reflects the variables that are included in the main datasets (i.e., lsacgrb0, lsacgrb2, lsacgrb4, lsacgrb6, lsacgrb8, lsacgrbl0, lsacgrbl2, lsacgrk4, lsacgrk6, lsacgrk8, lsacgrkl0, lsacgrkl2, lsacgrkl4, lsacgrkl6) Items from the study child household and the PLE household modules, the NAPLAN items and the Medicare items are not in the data dictionary.
The data from many of the responses to questions have been transformed to assist data users.

8.1 Transformations to ensure consistency

LSAC contains a number of items that have been asked slightly differently across waves. Where this is logically supportable, items are recoded to match the variables produced from other waves. These recoded versions are provided in addition to the original item response. Some examples of this:

- Income is generally collected as a continuous variable; however, for the PLE in Wave 2 it was collected using five categories. To assist users in comparing the responses of different informants, an additional variable containing the continuous income information recoded into these five categories is added wherever income has been collected continuously.

- In Wave 1, respondents were asked if the child received any regular child care from a grandparent. In Wave 2, respondents were given the option of this being a maternal or paternal grandparent. In addition to the two variables giving this information separately for maternal and paternal grandparents, an extra variable has been added for whether the child is being cared for by a grandparent.

8.2 Transformations to update information

From Wave 2 onwards, there are a number of places in the questionnaire where respondents are asked about what has happened with something since the last interview (or in the last two years if the study child is living in a new household). For example, in Wave 1, P1 was asked how many homes the study child had lived in since birth, while in subsequent waves P1 was asked how many homes the study child had lived in since the last interview. The datasets for the subsequent waves contain variables on the number of homes since the last interview and a tally of all the homes the study child has ever lived in.

8.3 Summary measures for scales

The appropriate summary measure for each scale is included, based on advice from the Consortium Advisory Group. Where it is possible to logically implement either a mean or a sum score for a psychological scale or subscale, the preference of the Consortium Advisory Group was to provide the calculation of means, except in cases where convention would dictate another scoring system. This enabled the calculation of scale level derivations where data measuring a construct has multiple contributing data items and where some contributing items are missing. Using a sum calculation for these scales would have led to the exclusion of cases with any missing data. All contributing data items to these scales are included on the datasets.

For scales where there are different sets of items for children at different ages or for different informants, multiple versions of the same scale are calculated based on just those items shared between two versions of the scale. For example, the parenting hostility scale began as a five-item measure for 0–1 year olds but had one item dropped for children aged 4–7 years, and a further item dropped for children aged 8–9 years. On the file for 0–1 year olds, three different versions of the scale are calculated: one using all five items, another using just the four items included for children aged 4–7 years, and another using just those three items used for children aged 8–9 years. As a general rule, data users should select the variable containing the greatest number of contributing items that is appropriate for their purpose. So, for analyses using just the hostility scale at age 0–1 years, or for those comparing the hostility scale at ages 0–1 and 2–3 years, analysts should use the five-item version. Analysts
comparing hostility between the ages of 0 and 7 years should use the four-item version, and analysts comparing hostility between the ages of 0 and 9 years should use the three-item version.

### 8.4 Outcome Index measures

A unique component of the derivation and analysis work was the development and derivation of the LSAC Outcome Index; a composite measure that indicates how children are developing. LSAC tracks the development of children across multiple domains, and the Outcome Index provides a means of summarising this complex information for policy makers, the media and the general public, as well as data users.

In contrast to some other indices, which focus on problems or negative outcomes, the LSAC Outcome Index, wherever possible, incorporates both positive and negative outcomes, reflecting the fact that most children have good developmental outcomes. Thus, the Outcome Index has the ability to distinguish groups of children developing poorly from those developing satisfactorily.

The rationale and methodology used to develop the Outcome Index are described in the LSAC Technical Paper No. 2, *Summarising children’s wellbeing: the LSAC Outcome Index*. Papers on the derivation of the Waves 2 and 3 Outcome Index are forthcoming. Any users planning to use the Outcome Index are strongly advised to read the technical papers (available from growingupinaustralia.gov.au/data-and-documentation/technical-papers) as they contain important information about the correct use of the variable. From Wave 4 the Outcome Index is not calculated.

When undertaking longitudinal analysis involving the Outcome Index, analysts should be cautious about using outcome indices from different waves in a pooled data file, as different measures may have been used at different waves to create the sub-domains.
Two types of data are available to data users:
- restricted release data
- general release data.

9.1 Restricted release data

The only information not included is name, address and other contact details for the child, family, child care agency and teacher or carer. Access to the in-confidence datasets may be granted where data users are able to demonstrate a genuine need for the additional data and they meet the necessary additional security requirements. It should be noted that from Wave 7 in-confidence data will be known as restricted release data.

9.2 General release data

In addition to the information removed for the restricted release data, some other items have also been removed, and some items have either been transformed, had response categories collapsed or have been top-coded (i.e. recoding outlying values to a less extreme value). These items have been flagged in the “Confidentialisation” column of the data dictionary. It is important for data users to be aware that these items are eligible for confidentiality if required but not all items may require confidentiality in a given wave.

Confidentialisation of general release data are detailed below.

The following items are removed:
- qualitative data provided by respondents
- census and postcode data for the location of carers and schools.

The following items are transformed:
- postcode – postcodes are given an indicator so that all children selected in the same postcode can be identified
- date left hospital after birth – number of days between birth and departure.

The following items have response categories collapsed (i.e. response categories combined to form an aggregate category):
- parents’ occupation – output at two-digit Australian and New Zealand Standard Classification of Occupations (ANZSCO) level, or rounded off to the nearest five if ANU four ratings of occupational prestige
- occupation in previous job – output at two-digit ANZSCO level
- Socio-Economic Index for Areas (SEIFA) variables – rounded to the nearest 10
- country of birth (coded as 0 if fewer than five contributors)
- religion (coded as 0 if fewer than five contributors)
- language other than English (LOTE) (coded as 0 if fewer than five respondents).

The following data items are top-coded:
- income
- housing costs
- child support paid by Parent 2
- children and parents’ current height, weight and waist circumference
- number of hours spent in child care.
The following topics are suppressed in relation to Study Child pregnancy and offspring information:

LSAC assessed disclosure risk assessment of Study Child offspring information available in K cohort (less than 5 cases). Topics that were considered as highly vulnerable to exposure to privacy risk were family demographics, health behaviour and risk factors, health status, home education environment, offspring program characteristics, paid work, parenting, parent living elsewhere, relationships and social capital. This information is available in restricted data file whereas the information has been suppressed and is presented as missing in general release data file.
Limited imputation of data is undertaken in LSAC. In general, imputation occurs only when there is a clear contradiction between data items and there is a good reason to believe one item over the other. Some basic principles are applied for this task.

10.1 Virtual roll-forward

‘Roll-forward’ is the term in CAI design (all forms of computer-assisted interviewing, including CASI, ACASI and CATI) that refers to the use of data from a previous wave of data collection to determine the questions that need to be asked in a subsequent wave. For Wave 2 a limited set of data was rolled forward, largely to assist with the household composition module. Time and resource implications meant that roll-forward could not be used in some other parts of the questionnaire where it may have reduced respondent burden.

For example, in Wave 2, respondents were again asked about the age the child stopped being breastfed, in order to obtain the information from those cases where this had not yet happened at the time of Wave 1. In re-asking this question, some respondents gave different answers to their Wave 1 responses. Given that recollection of respondents is likely to be more accurate closer to the event (i.e. the cessation of breastfeeding), it was decided that in cases where Wave 1 data exists, the Wave 1 value is taken as correct and the Wave 2 value is ignored (i.e. as if the Wave 1 data had been rolled forward and the question never asked in Wave 2). This means a single variable is produced that represents the best estimate from the two waves of data. (Users are able to tell at which wave the timing data was collected by referring to the question from each wave asking if the child is still being breastfed.)

Note: From Wave 3 onwards there is a greater use of roll-forward, which reduced the number of situations where such conflicts could occur.

10.2 Longitudinal contradictions

Another possible contradiction in the data may occur where respondents report at a subsequent wave that an event took place at a time before a previous wave, when the previous wave’s data indicated that this event hadn’t happened yet.

In these cases, the time of the previous wave is treated as the time of the event. For example, if a parent reported at Wave 2 that the child stopped being breastfed after two months; however, at Wave 1 the child was three months old and was reported as still being breastfed, the age of breastfeeding cessation would be set to three months.

This strategy for fixing the time of an event is also used for the:
- date when new members joined the household
- length of attendance at a particular child care facility
- date left the household for Wave 1 members and temporary members (bf14m1, bf14m2, etc.)
- age stopped breastfeeding (zf05c)
- age first had non-breast milk (zhb07)
- age first had solid food (zhb10)
- age entered child care arrangements (bpc11a, bpc11b, etc.)
- age last lived with two biological parents (bpe23c).
10.3 Other imputations

On inspection of the data, problems were revealed in a small number of items. These problems were solved using imputation and are listed below.

- Employment status: Some assumptions are made to assist in coding the parent to employed, unemployed or not in the labour force where missing values were present.

- Type of educational program (K cohort, Wave 1): There appeared to be some confusion with parents and interviewers as to whether the child was in pre-school or pre-Year 1 at school. The type of education program variable was amended based on the teacher data and other information provided in the questionnaire.

- Parental income: Outlying values, particularly those with responses to other questions (e.g. categorical income, sources of income) that make the income value appear incorrect, were adjusted. For further information about imputations related to parental income, see LSAC Technical Paper No. 14 *Imputing income in the Longitudinal Study of Australian Children* (available from growingupinaustralia.gov.au/data-and-documentation/technical-papers).

- Parental height: It was found that there were some changes in height between waves for some parents of study children. While most were minor (most likely due to estimation error) some were more substantial, and called into question the reliability of differences in body mass index recordings between waves.

- Time use diary data: In earlier Waves of LSAC, Responses were recorded by marking an oval to indicate whether an activity/situation occurred in each 15-minute time period. A number of ‘false positives’ were discovered in the Wave 1 TUD data. Imputation was used to reduce the number of false positives. A number of imputations were also performed to improve data quality in all three waves. This is no longer an issue in most recent Waves where TUD has been collected using the CAI/CSR type of data collection method.

Further details of these imputations are given in *Data Issues: Waves 1 to 7* (available from growingupinaustralia.gov.au/data-and-documentation/issues-papers).
Chapter 11: Survey methodology

11 Survey methodology

LSAC employs a cross-sequential design that follows two cohorts of children, initially aged 0–1 years (B cohort) and 4–5 years (K cohort) in 2004. Families are visited by interviewers every two years to collect data for the main waves of the study. In the ‘between’ years, a mailout survey was conducted at Waves 1.5, 2.5 and 3.5 to help maintain contact with families and obtain some additional information. At Waves 4.5 and 5.5, a web form was used primarily to update contact details.

The key features of the initial sample design and methodology for each wave are included in this section. A full description of the sample design is given in LSAC Technical Paper No. 1 Sample design, and details of the weighting and non-response analysis are given in subsequent technical papers. For technical papers see: growingupinaustralia.gov.au/data-and-documentation/technical-papers

11.1 Sample design

A two-stage clustered sample design was employed, first selecting postcodes then children, with the clustered design allowing analysis of children within communities and producing cost savings for interviews.

Stratification was used to ensure proportional geographic representation for states/territories and capital city statistical division/rest of state areas. The sample was stratified by state, capital city statistical division/balance of state and two strata based on the size of the target population in the postcode.

Postcodes were selected with probability proportional to size selection where possible, and with equal probability for small population postcodes. Children from both cohorts were selected from the same 311 postcodes. Some remote postcodes were excluded from the design, and the population estimates were adjusted accordingly.

Children were selected with approximately equal chance of selection for each child (about one in 25).

Apart from some remote areas, the sample was selected to be representative of all Australian children (citizens and permanent residents) in each of two selected age cohorts:

- children born March 2003–February 2004 (B cohort)

11.1.1 Sample selection and recruitment

The sample was selected from Medicare Australia’s enrolment database. Within the selected postcodes, the population was ordered by date of birth and then a random start and skip applied to select the children. The actual number of children selected depended on which stratum the postcode was in but for most postcodes the aim was to recruit about 20 children per cohort per postcode.

The selection of children and corresponding Wave 1 fieldwork occurred in four phases, partly to reduce the age range of children at interview and partly because some of the target population had not been born at the time of the first phase selection.

Families of 18,800 selected children received letters of invitation to take part in the study sent by Medicare Australia. Families could ‘opt-out’ of the study by phoning an 1800 number or returning a reply-paid slip. Medicare Australia 1800 staff were given training about the study and were able to answer queries and make notes of other information (e.g. telephone numbers).
After a 4-week opt-out period, Medicare Australia gave the contact names and addresses of remaining families to I-view, the Wave 1 data collection agency. I-view then sent another letter to families saying when an interviewer would be in their area.

I-view maintained a 1800 number for families selected in the study, which was transferred to the ABS who took responsibility for the data collection from Wave 2 onwards.

### 11.2 Development and testing of survey instruments

#### 11.2.1 Pre-testing

Pre-testing of new material and processes is undertaken at each wave of the study, comprising small-scale pre-tests and cognitive interviews. In Waves 1 and 2, more formal piloting was also undertaken. Small-scale testing is also undertaken for the between-wave surveys.

**Wave 1**
- Development began in March 2002.
- Small-scale pre-testing occurred in September–October 2002.
- A pilot test with about 50 families from each cohort was conducted in March–April 2003.

**Wave 2**
- Small-scale pre-testing occurred in September–October 2004.
- A pilot test with 86 families was conducted in April 2005.

**Wave 3**
- Development began in March 2006.
- Pre-testing occurred in a number of stages from mid 2006 to March 2007.
- No pilot test was required.

**Wave 4**
- Development began in February 2008.
- Pre-testing occurred in a number of stages from mid-August 2008 to June 2009.
- No pilot test was required.

**Wave 5**
- Development began in February 2010.
- Pre-testing occurred in a number of stages from mid-June 2009 to March 2010.
- No pilot test was required.

**Wave 6**
- Development began in May 2012.
- Pre-testing occurred in a number of stages from August 2012 to September 2013.
- No pilot test was required.

**Wave 7**
- Development began in May 2014.
- Pre-testing occurred in a number of stages from August 2014 to September 2014.
- No pilot test was required.
11.2.2 Dress rehearsal

In Wave 1, a dress rehearsal (DR) sample of 526 families was recruited to test the content and processes intended for the main waves of the study. Over 1,000 children were initially selected from 25 postcodes in Victoria, Sydney and rural/remote New South Wales and Queensland. Postcodes in Victoria were selected at random but the other postcodes were selected as areas that may provide challenges to the data collection process. Other dress rehearsals have also been completed.

- Wave 1 DR – August–November 2003 (526 families interviewed)
- Wave 2 DR – September–November 2005 (423 families interviewed)
- Wave 3 DR – July–October 2007 (420 families interviewed)
- Wave 4 DR – July–October 2009 (387 families interviewed)
- Wave 5 DR – July–August 2011 (451 families interviewed)
- Wave 6 DR – June–August 2013 (351 families interviewed)
- Wave 7 DR – June–September 2015 (309 families interviewed)

After each dress rehearsal, both processes and content have been refined to increase efficiency and reduce the time in the home.

11.3 Data collection

11.3.1 Interview length

In Wave 1, an average of 126 minutes was allowed for time in the home by the interviewer. In-home data collection with the B cohort averaged about 1.5 hours, while interviews for the K cohort averaged about 2.5 hours.

In Wave 2, although an average of 90 minutes had been allowed for the time in the home, the actual time was shorter, averaging 66 minutes for the B cohort and 85 minutes for the K cohort.

In Wave 3, an average of 100 minutes was allowed for time in the home; the actual time was 91 minutes for the B cohort and 98 minutes for the K cohort.

In Wave 4, an average of 110 minutes was allowed for time in the home; the actual time was 102 minutes for the B cohort and 108 minutes for the K cohort.

In Wave 5, an average of 110 minutes was allowed for time in the home; the actual time was 98 minutes for both cohorts.

In Wave 6, an average of 110 minutes was allowed for time in the home; the actual time was 108 minutes for the B cohort and 116 minutes for the K cohort.

In Wave 7, an average of 110 minutes was allowed for time in the home; the actual time was 114 minutes for the B cohort and 115 minutes for the K cohort.

11.3.2 Interviewers

As part of a standard ABS interviewer induction, ABS interviewers receive two weeks of intensive training across a range of standard procedures and practices. All interviewers received eight hours of home learning (this included a computer-based learning module, home study exercises and the reading of interviewer instructions).

In Wave 1, 150 interviewers and field supervisors from I-view were trained during a series of four-day sequential training courses conducted in Melbourne, Brisbane, Perth and Sydney during February and early March 2004. The principal trainers were the same for all courses, ensuring consistency in training.

Psychologists conducted the training for ‘Who am I?’, the PPVT and the interviewer observations. A large part of the training involved practice interviews, with one day devoted to interviews with parents and children.

For Wave 2, 147 interviewers from ABS were trained in a series of three-day training courses in Sydney, Melbourne, Brisbane and Perth during March and April 2006. Two training teams were used, comprising staff from both AIFS and ABS. This time, AIFS staff undertook the direct assessment training, after receiving training from a child psychologist (the use of computer-assisted interviewing for the direct assessments helped ensure the consistent administration of these assessments).
For Wave 3, 176 interviewers from ABS were trained in a series of two-day training courses in Brisbane, Melbourne, Sydney and Perth during March and April 2008. Interviewers who had not worked on LSAC previously were given background training in LSAC before the two-day course commenced. Two training teams were used, comprising staff from ABS, AIFS and DSS. Again, AIFS staff undertook the direct assessment training.

For Wave 4, 181 interviewers from ABS were trained in a series of three-day training courses in Brisbane, Melbourne, Sydney and Perth. Two training teams were used, comprising staff from the ABS, AIFS and DSS. As in previous waves, AIFS staff undertook the direct assessment training.

For Wave 5, 198 interviewers from ABS were trained in a series of three-day training courses in Brisbane, Melbourne, Sydney, Adelaide and Perth. New-to-LSAC interviewers (defined as anyone who did not participate in Main Wave 4) attended the first day of classroom training where topics such as ‘Background to the study’, ‘Physical measurements’, ‘Direct assessments’ and ‘Notebook security’ were covered. All interviewers attended Days 2 and 3 when the P1 interviews and the K and B child interviews were covered in detail (apart from what was done on Day 1). New interviewers were teamed with an experienced interviewer, allowing for mentoring throughout the training course and for the new interviewers to be the interviewer during practice sessions.

For Wave 6, 200 interviewers from ABS were trained in a series of four-day training courses in Brisbane, Melbourne, Sydney, Adelaide and Perth. All interviewers attended the full four-day training program due to the large amount of new content and procedures. During the practice sessions, interviewers were split into groups of three (rather than pairs as in previous waves). This allowed for a more realistic practice with each interviewer taking the role of the parent, child and interviewer. Where possible in the training sessions and in the practice sessions, new LSAC interviewers were paired with experienced LSAC interviewers. ABS staff conducted all of the training.

For Wave 7, 200 interviewers were in the initial training sessions (March–April), and then another 20 in a top-up training held in July 2016. All interviewers attended the full four-day training program due to the large amount of new content and procedures. During the practice sessions, interviewers were split into groups of three (rather than pairs as in Waves 1–5). This allowed for a more realistic practice with each interviewer taking the role of the parent, child and interviewer. Where possible in the training sessions and in the practice sessions, new LSAC interviewers were paired with experienced LSAC interviewers. ABS staff conducted all the training.

11.3.3 Fieldwork periods

**Wave 1**

Selected postcodes were divided into two groups for maximum field efficiency. The target population was also divided into two groups: children born March–August (older) in one group and children born September–February (younger) in the other.

The fieldwork was divided into four phases:

- Phase 1 started in mid-March 2004 for the older children in the first group of postcodes.
- Phase 2 started at the end of April for the older children in the second group of postcodes.
- Phase 3 started in June for the younger children in the first group of postcodes.
- Phase 4 started in late July for the younger children in the second group of postcodes.

Follow-up continued throughout 2004. The blue line in Figure 6 shows the distribution of interviews over time for Wave 1 fieldwork.

**Wave 2**

Again, there were broadly four fieldwork periods, although the dates for these varied from state to state. Regional offices of the ABS were able to organise the work to suit the availability of interviewers and other work. As far as possible, ABS tried to interview the children born in March–August in the first two periods, and children born in September–February in the later fieldwork periods. Eighty-four per cent of the interviews were conducted prior to September 2006.

Figure 6 shows the distribution of interviews over time for the Wave 2 fieldwork. Fieldwork started later than in Wave 1 due to the additional work required to prepare the CAI instrument.
Wave 3
Fieldwork was organised as per Wave 2. The green line in Figure 6 shows the distribution of interviews over time for the Wave 3 fieldwork.

Wave 4
As the children get older, the age differences within a cohort are less significant. To assist the efficiency of work allocations to interviewers, the focus in Wave 4 was more on the location of the sample and interviewers with less emphasis given to following interviews within the set phases. The dark blue line in Figure 6 shows the distribution of interviews over time for the Wave 4 fieldwork.

Wave 5
Fieldwork was organised based on the location of the sample and interviewers. Figure 6 shows that the distribution of interviews for the Wave 5 fieldwork was more spread out across the months than for previous waves.

Wave 6
Fieldwork was organised based on the location of the sample and interviewers. Figure 6 shows that the distribution of interviews for the Wave 6 fieldwork was distributed across the months, similar to Wave 5.

Wave 7
Fieldwork was organised based on the location of the sample and interviewers. Figure 6 shows that the distribution of interviews for Wave 7 decreased greatly in September, which can be mostly attributed to the ABS Census Post Enumeration Survey priorities during this time. Enumeration was extended from the originally planned end in December 2016 to May 2017.

Figure 6: Month of interview for study families in Waves 1 to 7

11.3.4 Contact process
Wave 1
For most families, the interviewer only had the name and address of the Medicare cardholder and which cohort the child was in. In a small number of cases, families who were keen to participate had contacted the 1800 numbers and supplied phone numbers and/or best times to call.
Interviewers were required to make up to six visits to the address, at different times of the day and on different days of the week. A major challenge was that 7% of addresses were post office box addresses, and although families with these addresses were specifically requested to make contact with the 1800 number to supply a residential address, only a small proportion did so. In addition, many of the residential addresses held by Medicare were found to be out of date by the time the interviewers visited. Interviewers made significant attempts to locate families for whom they did not have a current residential address, by referencing the White Pages and electoral rolls and speaking with neighbours and other local contacts.

**Between waves**

Contact is maintained with study families between waves by sending birthday cards, annual calendars and newsletters and via the between-wave mailout and online questionnaires. These processes have resulted in some families contacting the ABS to update their contact information, which helps when trying to arrange appointments for the main waves of interviewing.

**Subsequent waves**

Pre-interview letters plus a brochure outlining the processes for that wave were sent to all families who had not opted out of the study since the previous wave, unless it was confirmed that the address was out-of-date. Interviewers then followed up with a telephone call to make an appointment for an interview. If the contact information was out-of-date, the interviewers tried to contact secondary contacts of P1 (these details were given by P1 in Wave 1 and are updated each wave) to locate the family. One visit to the address was also made. If the family could not be located, the interviewer referred this back to the office for tracking.

After an appointment for interview was made, the interviewer confirmed the appointment the day before the appointment.

**11.3.5 Foreign language interviews**

**Wave 1**

As part of the Medicare Australia mailout, a brochure was included with information about the study in nine languages. Medicare Australia staff made use of the Telephone Interpreter Service (TIS) to assist with calls where required.

Apart from this brochure, no other study material was (or has been) translated into other languages, and instead interpreters were used. An interpreter was required in 3% of interviews, with over 50 languages involved. In most cases ($n = 138$), a member of the family or friend was preferred as the interpreter. In 76 cases, an I-view employee was able to act as interpreter and, in 96 cases, an interpreter was employed.

**Wave 2**

A total of 110 interviews (1%) were conducted in a language other than English, in 23 different languages. Family or friends assisted in 58 cases, ABS interpreters helped in 37 cases, and a TIS interviewer was used for 15 families. An interpreter was arranged whenever requested or judged necessary by the interviewer. The reduction in use of interpreters between waves is presumably due to the increased confidence in English that had been gained by respondents in this time.

**Wave 3**

A total of 97 interviews needed an interpreter, in 24 languages. Family or friends assisted in 58 cases, ABS interpreters helped in 31 cases, and a TIS interviewer was used for eight families.

**Wave 4**

A total of 93 interviews needed an interpreter, in 26 languages. Family or friends assisted in 50 cases, ABS interpreters helped in 29 cases, and a TIS interviewer was used for 14 families.

**Wave 5**

A total of 81 interviews needed an interpreter, in 18 languages. Family or friends assisted in 47 cases, ABS interpreters helped in 24 cases, and a TIS interviewer was used for 10 families.
Wave 6
A total of 64 interviews needed an interpreter, in 17 languages. Family or friends assisted in 42 cases, ABS interpreters helped in 18 cases, and a TIS interviewer was used for four families.

Wave 7
A total of 55 interviews needed an interpreter, in 19 languages. Family or friends assisted in 31 cases, ABS interpreters helped in 21 cases, and a TIS interviewer was used for three families.

11.3.6 Indigenous communities
Although the sample selection process excluded 40% of areas classified as remote by the ABS (areas that typically have a high Indigenous population) there were still a number of postcodes selected that contained some remote Indigenous communities, hence strategies have been put in place to enumerate these communities.

Where feasible, communities were visited or telephoned, and personal contact made with a number of community organisations from whom assistance was gained to identify whether families were in residence and willing to be interviewed. Travel to remote communities was only undertaken if there was an appointment for an interview.

Aboriginal and Torres Strait Islander families are included in representative numbers in non-remote centres. However, there has been a higher rate of attrition from the study among these families. See the weighting and non-response technical papers for more details (available from growingupinaustralia.gov.au/data-and-documentation/technical-papers).

11.3.7 Remote areas
In the initial sample, there were 12 postcodes selected in areas classified as ‘remote’ by the ABS Australian Standard Geographic Classification (ASGC) Remoteness Classification. Interviewers were either recruited from these areas or travelled to these areas when the field agency did not have a suitable interviewer in the locality.

Where visits were not possible, telephone interviews were conducted: 12 (0.12%) in Wave 1, 42 (0.46%) in Wave 2, 87 (0.10%) in Wave 3, 83 (0.99%) in Wave 4, 73 (0.91%) in Wave 5, 59 (0.81%) in Wave 6 and, in Wave 7, 49 (0.76%) of call interviews were conducted when visits were not possible.

11.4 Fieldwork response

11.4.1 Wave 1 recruitment
The final response to the recruitment of children was 54% of those families who were sent the initial letter by Medicare Australia. The response rate was higher for the B cohort, with 57% of families (5,107) agreeing to take part, compared with 50% of K cohort families (4,983).

About 35% of families who were sent the initial letter refused to take part in the study. The main reasons given to interviewers for not participating in the study were: not interested/too busy (57%), not capable/moving/overseas (9%), husband refused (9%), and illness/death (8%). The remaining 13% of families could not be contacted, despite intensive efforts from interviewers.

Non-response analysis was undertaken to determine how representative the sample is of all Australian children in the scope of this study, and adjustments have been made to the survey weights to allow for this. For further information on the weighting and non-response, see LSAC Technical Paper No. 3, Wave 1 weighting and non-response analysis (available from growingupinaustralia.gov.au/data-and-documentation/technical-papers).

11.4.2 Response in later waves
Table 13 summarises the response from families in later waves, using the Wave 1 sample and ‘available’ sample as the bases for comparisons.
Table 13: Sample size and response rate for each wave and cohort of LSAC

<table>
<thead>
<tr>
<th>Wave</th>
<th>B cohort</th>
<th>K cohort</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Resp. rate of Wave 1 (%)</td>
<td>Resp. rate of available sample (%)</td>
</tr>
<tr>
<td>Main waves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 1</td>
<td>5,107</td>
<td>100</td>
<td>4,983</td>
</tr>
<tr>
<td></td>
<td>5,047</td>
<td>98.8</td>
<td>4,913</td>
</tr>
<tr>
<td>Wave 2</td>
<td>4,606</td>
<td>90.2</td>
<td>4,464</td>
</tr>
<tr>
<td>available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4,971</td>
<td>97.3</td>
<td>4,829</td>
</tr>
<tr>
<td>responding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 3</td>
<td>4,386</td>
<td>85.9</td>
<td>4,331</td>
</tr>
<tr>
<td>available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4,929</td>
<td>96.5</td>
<td>4,774</td>
</tr>
<tr>
<td>responding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 4</td>
<td>4,242</td>
<td>83.0</td>
<td>4,169</td>
</tr>
<tr>
<td>available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4,884</td>
<td>96.6</td>
<td>4,735</td>
</tr>
<tr>
<td>responding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 5</td>
<td>4,085</td>
<td>80.0</td>
<td>3,956</td>
</tr>
<tr>
<td>available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4,483</td>
<td>87.8</td>
<td>4,395</td>
</tr>
<tr>
<td>responding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 6</td>
<td>3,764</td>
<td>73.7</td>
<td>3,537</td>
</tr>
<tr>
<td>available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4,318</td>
<td>84.6</td>
<td>4,176</td>
</tr>
<tr>
<td>responding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 7</td>
<td>3,381</td>
<td>66.2</td>
<td>3,089</td>
</tr>
<tr>
<td>available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between waves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 1.5</td>
<td>5,061</td>
<td>99.1</td>
<td>4,935</td>
</tr>
<tr>
<td>sent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 1.5</td>
<td>3,573</td>
<td>70.0</td>
<td>3,584</td>
</tr>
<tr>
<td>returned</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 2.5</td>
<td>4,859</td>
<td>95.1</td>
<td>4,712</td>
</tr>
<tr>
<td>sent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 2.5</td>
<td>3,268</td>
<td>63.5</td>
<td>3,287</td>
</tr>
<tr>
<td>returned</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 3.5</td>
<td>4,772</td>
<td>93.4</td>
<td>4,641</td>
</tr>
<tr>
<td>sent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 3.5</td>
<td>3,012</td>
<td>59.0</td>
<td>2,972</td>
</tr>
<tr>
<td>returned</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Excludes in-between Waves 4.5 and 5.5 where the data is not relevant for users of the LSAC datasets. They were used only to update contact details. * Available sample excludes those who opted out of the study between waves. Some additional families also opted out permanently during the fieldwork process. b Those who had a home visit.

Table 14 details the reasons why interviews were not obtained in Waves 2–7.
### Table 14: Response status and reasons for non-response by wave

<table>
<thead>
<tr>
<th>Response status</th>
<th>Wave 2</th>
<th>Wave 3</th>
<th>Wave 4</th>
<th>Wave 5</th>
<th>Wave 6</th>
<th>Wave 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responding</td>
<td>9,070</td>
<td>8,717</td>
<td>8,411</td>
<td>8,041</td>
<td>7,301</td>
<td>6,470</td>
</tr>
<tr>
<td>Refusal</td>
<td>284</td>
<td>436</td>
<td>637</td>
<td>774</td>
<td>938</td>
<td>1,118</td>
</tr>
<tr>
<td>Non-contact</td>
<td>540</td>
<td>552</td>
<td>526</td>
<td>715</td>
<td>555</td>
<td>803</td>
</tr>
<tr>
<td>Away entire enumeration period</td>
<td>61</td>
<td>93</td>
<td>135</td>
<td>88</td>
<td>39</td>
<td>34</td>
</tr>
<tr>
<td>Death of study child</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>43</td>
</tr>
<tr>
<td>Total starting sample</td>
<td>9,960</td>
<td>9,799</td>
<td>9,709</td>
<td>9,619</td>
<td>8,879</td>
<td>8,494</td>
</tr>
</tbody>
</table>

Note: N/A indicates not applicable.
12 Important issues for data analysis

A data issues paper for Waves 1 to 7 is available and details issues that have been identified over the course of the study. Data users should be aware of these issues when using LSAC data for their analysis (see growingupinaustralia.gov.au/data-and-documentation/technical-papers). Other important issues are addressed below.

12.1 Weighting and external validity

The LSAC study design, based on a complex probability sample, is specifically designed to produce valid estimates at the population level. Unlike clinically based or convenience samples, the LSAC sample is population based by design. By properly accounting for the survey design when analysing the data, it is possible not only to make inferences about the children and families participating in the study but to make valid inferences about the entire population of children in the relevant age groups.

The LSAC sampling strategy has three important elements that distinguish it from a simple random sample (SRS):

- **stratification** to ensure proportional representation of all states and both capital city and ex-metropolitan areas
- **clustering** by postcode to both reduce field enumeration costs and allow the study of community-level effects on children’s development and wellbeing
- **weighting** to adjust for potential non-response bias and to provide population estimates.

It is the responsibility of data users to determine when and how each of these needs to be accounted for when developing their analyses.

12.1.1 Stratification

Stratification, by state and part of state, was employed to ensure that all geographic areas within Australia are represented in the sample in proportion to their population. This produces a more even distribution of the sample across geographic areas than could be expected from a simple random sample.

The use of stratification can be expected to reduce standard errors compared with a simple random sample with no control over the geographic spread of the sample. As such, when trying to extrapolate to the population, the stratification should be incorporated in the analysis of results from the survey in order to correctly calculate standard errors and confidence intervals.

12.1.2 Clustering

The use of clustering in the sample design has important consequences for the analysis of data from the study. Clustering is useful in reducing the field costs associated with the survey enumeration. Clustering also has the added benefit of making possible the analysis of community-level effects, by ensuring that a sufficient sample is selected from each postcode included in the survey.

However, the use of clustering violates the standard assumption of independence of the observations that is fundamental to many statistical routines in major statistical packages. When children or carers have more similar characteristics within a given postcode than children or carers selected purely at random, the responses within postcodes will be correlated. This correlation will lead to an increase in the standard errors and size of the confidence intervals. The extent of this increase is measured by the *design effect*, which is the ratio of the
variance of an estimate from the survey to the variance that would have been achieved by a simple random sample of the same size.

Failure to account for clustering in the analysis can lead to under-estimating the size of standard errors and confidence intervals. In some circumstances, this can result in misleading conclusions of statistical significance.

12.1.3 Weighting

The Wave 1 weights provided in the LSAC data files take into account both the probability of selecting each child in the study and an adjustment for non-response. An analysis of possible differences in the characteristics of respondents and non-respondents was undertaken and identified two factors associated with the probability of participating in the survey – whether the mother speaks a language other than English at home, and whether the mother has completed Year 12. Both of these factors were incorporated into the Wave 1 survey weighting so that, to the best extent possible, the use of the sample weights offset the bias that may be introduced into the data due to differential non-response patterns.

At each subsequent wave of data collection, weights have been adjusted to account for the differential probability of response as estimated by regression. The weights are then calibrated back to the stratum benchmarks and a small number of cases have their weights top or bottom coded to prevent any case having too great or small an effect on the data.

From Wave 3 onwards, it was required to produce longitudinal as well as cross-sectional weights for the first time. Cross-sectional weights adjust the sample attained at current wave to be representative of the population at the time of selection (i.e. when first interviewed), while longitudinal weights do the same for the sample that has responded to all waves of the survey.

More detailed information on the weighting variables can be found in LSAC Technical Papers No. 3, 5, 6, 9, 10, 15 and 16 (available at growingupinaustralia.gov.au/data-and-documentation/technical-papers).

Three types of weight are included in the LSAC datasets:

- **Child population weights** – these weights are used to produce population estimates based on the LSAC data (e.g. based on LSAC data there are 22,464 children born in March 2003–February 2004 in Australia that were never breastfed).

  The sum of the responding B cohort child population weights is 243,026 and the sum of the K cohort child population weights is 253,202, which are the ABS-estimated resident population counts of children aged 0 years 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 (e.g. 5,107 B cohort and 4,983 K cohort in Wave 1).

  This weight is 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.

- **Time use data day weight** (for Waves 1, 2 and 3 only) – this is the sample weight adjusted so that each day of the week receives equal weight in analyses of time use data.

Data files for Wave 1 and Wave 2 each have one population weight and one sample weight. Given that there are no cases that responded to Wave 2 and didn’t respond to Wave 1, these weights can be used for both longitudinal and cross-sectional analyses.

At Wave 3, two sample weights and two population weights are necessary as this is the first time that respondents could return to the study after missing a wave. The first of these weights the full Wave 3 sample and should be used for cross-sectional analyses. The second weights the sample that has responded to all waves, and should be used for longitudinal analyses.

A complete list of LSAC weighting variables is given in Tables 15 (B cohort) and 16 (K cohort).
<table>
<thead>
<tr>
<th>Variable name</th>
<th>Cohort</th>
<th>Waves cases responded to</th>
<th>Used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>aweight</td>
<td>B</td>
<td>Population 1</td>
<td>Wave 1 cross-sectional analyses</td>
</tr>
<tr>
<td>aweights</td>
<td>B</td>
<td>Sample 1</td>
<td>Wave 1 cross-sectional analyses</td>
</tr>
<tr>
<td>aweightd</td>
<td>B</td>
<td>Day 1</td>
<td>Wave 1 cross-sectional analyses</td>
</tr>
<tr>
<td>bweight</td>
<td>B</td>
<td>Population 1 &amp; 2</td>
<td>Wave 2 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 2</td>
</tr>
<tr>
<td>bweights</td>
<td>B</td>
<td>Sample 1 &amp; 2</td>
<td>Wave 2 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 2</td>
</tr>
<tr>
<td>bweightd</td>
<td>B</td>
<td>Day 1 &amp; 2</td>
<td>Wave 2 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 2</td>
</tr>
<tr>
<td>cweight</td>
<td>B</td>
<td>Population 1 &amp; 3</td>
<td>Wave 3 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 3</td>
</tr>
<tr>
<td>cweights</td>
<td>B</td>
<td>Sample 1 &amp; 3</td>
<td>Wave 3 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 3</td>
</tr>
<tr>
<td>cweightd</td>
<td>B</td>
<td>Day 1 &amp; 3</td>
<td>Wave 3 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 3</td>
</tr>
<tr>
<td>bcwt</td>
<td>B</td>
<td>Population 1, 2 &amp; 3</td>
<td>Longitudinal analyses involving all waves up to Wave 3</td>
</tr>
<tr>
<td>bcwts</td>
<td>B</td>
<td>Sample 1, 2 &amp; 3</td>
<td>Longitudinal analyses involving all waves up to Wave 3</td>
</tr>
<tr>
<td>bcwt</td>
<td>B</td>
<td>Day 1, 2 &amp; 3</td>
<td>Longitudinal analyses involving all waves up to Wave 3</td>
</tr>
<tr>
<td>dweight</td>
<td>B</td>
<td>Population 1 &amp; 4</td>
<td>Wave 4 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 4</td>
</tr>
<tr>
<td>dweights</td>
<td>B</td>
<td>Sample 1 &amp; 4</td>
<td>Wave 4 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 4</td>
</tr>
<tr>
<td>eweight</td>
<td>B</td>
<td>Population 1 &amp; 5</td>
<td>Wave 5 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 5</td>
</tr>
<tr>
<td>eweightd</td>
<td>B</td>
<td>Sample 1 &amp; 5</td>
<td>Wave 5 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 5</td>
</tr>
<tr>
<td>bdwt</td>
<td>B</td>
<td>Population 1, 2 &amp; 4</td>
<td>Longitudinal analyses involving Waves 2 &amp; 4, or Waves 1, 2 &amp; 4</td>
</tr>
<tr>
<td>bdwts</td>
<td>B</td>
<td>Sample 1, 2 &amp; 4</td>
<td>Longitudinal analyses involving Waves 2 &amp; 4, or Waves 1, 2 &amp; 4</td>
</tr>
<tr>
<td>cdwt</td>
<td>B</td>
<td>Population 1, 3 &amp; 4</td>
<td>Longitudinal analyses involving Waves 3 &amp; 4, or Waves 1, 3 &amp; 4</td>
</tr>
<tr>
<td>cdwts</td>
<td>B</td>
<td>Sample 1, 3 &amp; 4</td>
<td>Longitudinal analyses involving Waves 3 &amp; 4, or Waves 1, 3 &amp; 4</td>
</tr>
<tr>
<td>bcdwt</td>
<td>B</td>
<td>Population 1, 2, 3 &amp; 4</td>
<td>Longitudinal analyses involving all Waves up to Wave 4</td>
</tr>
<tr>
<td>bcdwts</td>
<td>B</td>
<td>Sample 1, 2, 3 &amp; 4</td>
<td>Longitudinal analyses involving all Waves up to Wave 4</td>
</tr>
<tr>
<td>fweight</td>
<td>B</td>
<td>Population 1 &amp; 6</td>
<td>Wave 6 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 6</td>
</tr>
<tr>
<td>fweights</td>
<td>B</td>
<td>Sample 1 &amp; 6</td>
<td>Wave 6 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 6</td>
</tr>
<tr>
<td>bcdefwt</td>
<td>B</td>
<td>Population 1, 2, 3, 4, 5 &amp; 6</td>
<td>Longitudinal analyses involving all waves up to Wave 6</td>
</tr>
<tr>
<td>bcdefwts</td>
<td>B</td>
<td>Sample 1, 2, 3, 4, 5 &amp; 6</td>
<td>Longitudinal analyses involving all waves up to Wave 6</td>
</tr>
<tr>
<td>gweight</td>
<td>B</td>
<td>Population 1 &amp; 7</td>
<td>Wave 7 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 7</td>
</tr>
<tr>
<td>gweights</td>
<td>B</td>
<td>Sample 1 &amp; 7</td>
<td>Wave 7 cross-sectional analyses and longitudinal analyses involving Waves 1 &amp; 7</td>
</tr>
<tr>
<td>bcdefgwt</td>
<td>B</td>
<td>Population 1, 2, 3, 4, 5, 6 &amp; 7</td>
<td>Longitudinal analyses involving all waves up to Wave 7</td>
</tr>
<tr>
<td>bcdefgwts</td>
<td>B</td>
<td>Sample 1, 2, 3, 4, 5, 6 &amp; 7</td>
<td>Longitudinal analyses involving all waves up to Wave 7</td>
</tr>
<tr>
<td>Variable name</td>
<td>Cohort</td>
<td>Type</td>
<td>Waves cases responded to</td>
</tr>
<tr>
<td>---------------</td>
<td>--------</td>
<td>------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>cweight</td>
<td>K</td>
<td>Population</td>
<td>1</td>
</tr>
<tr>
<td>cweights</td>
<td>K</td>
<td>Sample</td>
<td>1</td>
</tr>
<tr>
<td>cweightd</td>
<td>K</td>
<td>Day</td>
<td>1</td>
</tr>
<tr>
<td>dweight</td>
<td>K</td>
<td>Population</td>
<td>1 &amp; 2</td>
</tr>
<tr>
<td>dweights</td>
<td>K</td>
<td>Sample</td>
<td>1 &amp; 2</td>
</tr>
<tr>
<td>dweightd</td>
<td>K</td>
<td>Day</td>
<td>1 &amp; 2</td>
</tr>
<tr>
<td>eweight</td>
<td>K</td>
<td>Population</td>
<td>1 &amp; 3</td>
</tr>
<tr>
<td>eweights</td>
<td>K</td>
<td>Sample</td>
<td>1 &amp; 3</td>
</tr>
<tr>
<td>eweightd</td>
<td>K</td>
<td>Day</td>
<td>1 &amp; 3</td>
</tr>
<tr>
<td>dweight</td>
<td>K</td>
<td>Population</td>
<td>1, 2 &amp; 3</td>
</tr>
<tr>
<td>dweights</td>
<td>K</td>
<td>Sample</td>
<td>1, 2 &amp; 3</td>
</tr>
<tr>
<td>dweightd</td>
<td>K</td>
<td>Day</td>
<td>1, 2 &amp; 3</td>
</tr>
<tr>
<td>fweight</td>
<td>K</td>
<td>Population</td>
<td>1 &amp; 4</td>
</tr>
<tr>
<td>fweights</td>
<td>K</td>
<td>Sample</td>
<td>1 &amp; 4</td>
</tr>
<tr>
<td>dfwt</td>
<td>K</td>
<td>Population</td>
<td>1, 2 &amp; 4</td>
</tr>
<tr>
<td>dfwts</td>
<td>K</td>
<td>Sample</td>
<td>1, 2 &amp; 4</td>
</tr>
<tr>
<td>efwt</td>
<td>K</td>
<td>Population</td>
<td>1, 3 &amp; 4</td>
</tr>
<tr>
<td>efwts</td>
<td>K</td>
<td>Sample</td>
<td>1, 3 &amp; 4</td>
</tr>
<tr>
<td>defwt</td>
<td>K</td>
<td>Population</td>
<td>1, 2, 3 &amp; 4</td>
</tr>
<tr>
<td>defwts</td>
<td>K</td>
<td>Sample</td>
<td>1, 2, 3 &amp; 4</td>
</tr>
<tr>
<td>gweight</td>
<td>K</td>
<td>Population</td>
<td>1 &amp; 5</td>
</tr>
<tr>
<td>gweights</td>
<td>K</td>
<td>Sample</td>
<td>1 &amp; 5</td>
</tr>
<tr>
<td>defgwt</td>
<td>K</td>
<td>Population</td>
<td>12, 3, 4 &amp; 5</td>
</tr>
<tr>
<td>defgwts</td>
<td>K</td>
<td>Sample</td>
<td>12, 3, 4 &amp; 5</td>
</tr>
<tr>
<td>hweight</td>
<td>K</td>
<td>Population</td>
<td>1 &amp; 6</td>
</tr>
<tr>
<td>hweights</td>
<td>K</td>
<td>Sample</td>
<td>1 &amp; 6</td>
</tr>
<tr>
<td>defghwt</td>
<td>K</td>
<td>Population</td>
<td>1, 2, 3, 4, 5 &amp; 6</td>
</tr>
<tr>
<td>defghwts</td>
<td>K</td>
<td>Sample</td>
<td>1, 2, 3, 4, 5 &amp; 6</td>
</tr>
<tr>
<td>iweight</td>
<td>K</td>
<td>Population</td>
<td>1 &amp; 7</td>
</tr>
<tr>
<td>iweights</td>
<td>K</td>
<td>Sample</td>
<td>1 &amp; 7</td>
</tr>
<tr>
<td>defghiwts</td>
<td>K</td>
<td>Population</td>
<td>1, 2, 3, 4, 5, 6 &amp; 7</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>Sample</td>
<td>1, 2, 3, 4, 5, 6 &amp; 7</td>
</tr>
</tbody>
</table>
12.1.4 Survey estimation and analysis techniques

Survey estimation and analysis techniques are available that can take all three key features of the study design into account, and many of these techniques are now included in commercially available software. Incorporating the study design features into analyses of the study can produce externally valid results at the full population level. Estimates of means, proportions and totals incorporating the study design provide the best estimate of the true means, proportions and totals within the total population.

Analytic techniques, particularly modelling, aim at exploring relationships within the data and are able to estimate the best fitting model for the underlying population, not just the best fitting model for the sample, when properly applied to account for the study design.

12.1.5 Useful references

An overview of population survey methods is given by Levy and Lemeshow (1999). They discuss the use of stratification, weighting and clustering in survey design and the impact it has on the analysis of sample survey data.

For a thorough discussion of the mathematical techniques used to analyse data from complex surveys, see Chambers and Skinner (2003).

12.1.6 Software

There is now a range of software available from a number of vendors that supports the analysis of data from complex survey designs incorporating stratification, clustering and weighting. These include SAS (using the SURVEYMEANS and SURVEYREG procedures), STATA (using the svy commands), and SPSS (through the SPSS Complex Samples add-on module), as well as software packages specifically designed for the analysis of sample survey data such as WesVar and SUDAAN.

Use of the appropriate analytic techniques from one or more of these packages is recommended for researchers analysing the LSAC data. Results that properly account for the sample design features will have the greatest external validity and should be appropriate for drawing inferences about the total population of children from which the sample was taken.

The following shows a template for using the SURVEYREG and SURVEYMEANS procedures in SAS.

```
proc surveyreg data=<filename> total=<stratumfile>;
  stratum stratum;
  cluster pcodes;
  model <standard SAS model details>;
  weight weights;
run;
```

```
proc surveymeans data=<filename> total=<stratumfile>;
  stratum stratum;
  cluster pcodes;
  var <variable names>;
  weight weights;
run;
```

Where:
- `stratum`: is a variable you can calculate for lsac0 using the formula:
  ```
  stratum=int(mod(hicid,10000000)/100000);
  ```
- `pcodes`: is the postcode of selection (already on the data file)
- `weights`: is the sample weight (preferred to the population weight for this analysis)
<stratumfile> is a file that contains the number of Primary Sampling Units (in this case postcode clusters) in each stratum. It is included on the data CD or can be set up using the following code.

```sas
data stratum;
  input stratum _total_; datalines;
  11 295
  13 168
  14 160
  21 202
  22 58
  23 95
  24 316
  31 116
  33 121
  34 108
  41 110
  43 34
  44 131
  51 82
  52 86
  53 32
  54 103
  61 28
  63 38
  71 9
  73 3
  74 1
  81 23
; run;
```

12.2 Unit of analysis

The child is the unit of selection in LSAC and estimates produced from this survey are of children, not of parents or families. It is important that this point is understood when producing population estimates from this survey.

Using the estimates to count families/parents will produce an over-count of the number of families/parents, due to the multiple (or over) counting of children from multiple births. Although this will not make a huge 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, it is not considered a worthwhile use of resources given the small number of analyses this could possibly meaningfully affect.

12.3 Age at interview

Different ages of children should be accounted for in any analyses focused on age-dependent measures such as cognitive and motor development. Figures 7 and 8 show the age distribution of the two cohorts at each wave. The figures show the age of the child as a base figure (i.e. 0, 2, 4, 6, 8, 10, 12, 14 and 16 years) plus a number of
months. For example, a B cohort child aged three years and one month at time of interview in Wave 2 is shown against ‘13’ on the x-axis (see the red line).

**Figure 7:** Age distribution of B cohort sample at each wave

**Figure 8:** Age distribution of K cohort sample at each wave
12.4 Time between interviews

Effort is made to ensure that the time between interviews is close to two years; however, in some cases this is not possible. Figures 9 and 10 show the distribution of the intervals between waves.

**Figure 9:** Distribution of time between interviews, B cohort, Waves 1–7

**Figure 10:** Distribution of time between interviews, K cohort, Waves 1–7
12.5 Cross-cohort comparisons

It should be noted that the two cohorts of LSAC were selected and weighted to represent similar but different populations. For the B cohort, the reference population is ‘0-year-old children in Australia in 2004 excluding those from certain remote postcodes’, while for the K cohort the reference population is ‘4-year-old children in Australia in 2004 excluding those from certain remote postcodes’. One implication of this is that the K cohort will have a greater number of children born overseas as there was more time for families to immigrate to Australia between the birth of their child and selection into the study. The 2001 census contained 4.4% of four year olds born overseas compared with 0.8% of 0 year olds. In comparison, the weighted percentages for these figures in LSAC at Wave 1 were 4.2% vs 0.4%.

However, there are other demographic differences between the populations that are reflected in the benchmarks used to weight the two cohorts. Figure 11 shows the population percentages in each state by part of state and by gender stratum for the B and K cohorts. The B and K cohort figures generally match closely; however, the population from which the K cohort was selected was a little more likely to live in capital cities (66.5% vs 63.6%). Figure 12 shows the population proportions for mothers having completed Year 12 by state and part of state for each cohort. The B cohort population was more likely to have completed Year 12 in every part of the country, with the ABS census figures nationally being 56.6% for the B cohort against 48.3% for the K cohort. Figure 13 shows the population proportions for mothers speaking a language other than English at home by state and part of state for each cohort. These proportions were more closely matched between the B and K cohorts.

The implications of this are that even though the two cohorts have been weighted using similar variables, it does not mean that the variables that they have been weighted on are not responsible for the differences observed between the two. For example, because the two cohorts have had non-response due to maternal education adjusted for, it does not mean they will have equal proportions of mothers who had completed Year 12 when the weights are applied. Therefore, different levels of maternal education could explain differences observed between the two samples in the educational attainment of children.

**Figure 11:** Cohort benchmarks by state, part of state and gender

![Chart showing population percentages by state, part of state, and gender for B and K cohorts.](chart.png)

**Note:** There are no respondents from non-metropolitan ACT.
Figure 12: Proportion of mothers who completed Year 12, cohort benchmarks by state and part of state

- NSW: Met Xmet
- VIC: Met Xmet
- QLD: Met Xmet
- SA: Met Xmet
- WA: Met Xmet
- TAS: Met Xmet
- NT: Met Xmet
- ACT: Met

Note: There are no respondents from non-metropolitan ACT.

Figure 13: Proportion of mothers who speak a language other than English at home, cohort benchmarks by state and part of state

- NSW: Met Xmet
- VIC: Met Xmet
- QLD: Met Xmet
- SA: Met Xmet
- WA: Met Xmet
- TAS: Met Xmet
- NT: Met Xmet
- ACT: Met

Note: There are no respondents from non-metropolitan ACT.
12.6 Sample characteristics

To assist in the assessment of the representativeness of the Wave 1 sample, selected characteristics were compared with ABS estimates: gender, state and region were compared with the ABS September 2004 Estimated Resident Population figures; the other characteristics were compared with (previously unpublished) population data from the ABS 2001 Census of Population and Housing (see Table 17).

Table 17: Wave 1 sample characteristics compared with ABS data

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>LSAC %</th>
<th>ABS %</th>
<th>LSAC %</th>
<th>ABS %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>51.2</td>
<td>51.3</td>
<td>50.9</td>
<td>51.3</td>
</tr>
<tr>
<td>Female</td>
<td>48.8</td>
<td>48.7</td>
<td>49.1</td>
<td>48.7</td>
</tr>
<tr>
<td><strong>Family type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two resident parents/guardians</td>
<td>90.7</td>
<td>88.1</td>
<td>86.0</td>
<td>82.0</td>
</tr>
<tr>
<td>One resident parent/guardian</td>
<td>9.3</td>
<td>11.9</td>
<td>14.0</td>
<td>18.0</td>
</tr>
<tr>
<td><strong>Siblings</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only child</td>
<td>39.5</td>
<td>36.2</td>
<td>11.5</td>
<td>12.1</td>
</tr>
<tr>
<td>One sibling</td>
<td>36.8</td>
<td>35.6</td>
<td>48.4</td>
<td>45.9</td>
</tr>
<tr>
<td>Two or more siblings</td>
<td>23.7</td>
<td>28.2</td>
<td>40.1</td>
<td>42.0</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study child Indigenous</td>
<td>4.5</td>
<td>4.3</td>
<td>3.8</td>
<td>4.3</td>
</tr>
<tr>
<td>Mother speaks a language other than English at home</td>
<td>14.5</td>
<td>16.8</td>
<td>15.7</td>
<td>17.6</td>
</tr>
<tr>
<td><strong>Educational status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother completed Year 12</td>
<td>66.9</td>
<td>56.6</td>
<td>58.6</td>
<td>48.3</td>
</tr>
<tr>
<td>Father completed Year 12</td>
<td>58.5</td>
<td>50.2</td>
<td>52.7</td>
<td>45.3</td>
</tr>
<tr>
<td><strong>State</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New South Wales</td>
<td>31.6</td>
<td>34.1</td>
<td>31.6</td>
<td>33.7</td>
</tr>
<tr>
<td>Victoria</td>
<td>24.5</td>
<td>24.6</td>
<td>25.0</td>
<td>23.8</td>
</tr>
<tr>
<td>Queensland</td>
<td>20.6</td>
<td>19.3</td>
<td>19.8</td>
<td>19.7</td>
</tr>
<tr>
<td>South Australia</td>
<td>6.8</td>
<td>6.8</td>
<td>6.8</td>
<td>7.2</td>
</tr>
<tr>
<td>Western Australia</td>
<td>10.4</td>
<td>9.9</td>
<td>10.2</td>
<td>10.1</td>
</tr>
<tr>
<td>Tasmania</td>
<td>2.2</td>
<td>2.3</td>
<td>2.7</td>
<td>2.5</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>1.7</td>
<td>1.4</td>
<td>1.7</td>
<td>1.6</td>
</tr>
<tr>
<td>Australian Capital Territory</td>
<td>2.1</td>
<td>1.7</td>
<td>2.3</td>
<td>1.3</td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital city statistical division</td>
<td>62.5</td>
<td>63.7</td>
<td>62.1</td>
<td>62.1</td>
</tr>
<tr>
<td>Balance of state</td>
<td>37.5</td>
<td>26.3</td>
<td>37.9</td>
<td>37.9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>5,047</td>
<td>4,983</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: ABS data comes from the 2001 Census for families for 0 and 4 year olds, except where indicated with a *, where it is based on the September 2004 Estimated Resident Population for families of 0 and 4 year olds.

For most characteristics, the Wave 1 sample is only marginally different to the ABS data. The largest difference is in the educational status of the parents. Children with mothers who have completed Year 12 are over-represented in the sample, with proportions 10% higher than in the 2001 Census.

Other differences in the Wave 1 sample include:
- Children in lone-parent families are under-represented.
- Children with two or more siblings are under-represented and only children are over-represented in the B cohort.
- Children from an ATSI background are under-represented for the K cohort, and marginally over-represented for the B cohort.
- Children with mothers who speak a language other than English at home are under-represented.
- Children in New South Wales are under-represented.

Table 18 shows the number of children in the Wave 1 sample with selected characteristics, and gives the Waves 2–7 response rates for children with these characteristics. As can be seen in the table, the greatest sample loss has been from Indigenous families and families where P1 speaks a language other than English at home.

**Table 18: Response rates at Waves 2–7 by selected sample characteristics**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Wave 1 n</th>
<th>% responding at Wave 2</th>
<th>% responding at Wave 3</th>
<th>% responding at Wave 4</th>
<th>% responding at Wave 5</th>
<th>% responding at Wave 6</th>
<th>% responding at Wave 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>5,107</td>
<td>4,983</td>
<td>90.2</td>
<td>89.6</td>
<td>85.9</td>
<td>86.9</td>
<td>83.1</td>
</tr>
<tr>
<td>Study child male</td>
<td>2,610</td>
<td>2,537</td>
<td>90.0</td>
<td>89.8</td>
<td>86.2</td>
<td>87.2</td>
<td>83.9</td>
</tr>
<tr>
<td>Study child female</td>
<td>2,497</td>
<td>2,446</td>
<td>90.3</td>
<td>89.4</td>
<td>85.5</td>
<td>86.6</td>
<td>82.2</td>
</tr>
<tr>
<td>Study child Indigenous</td>
<td>230</td>
<td>187</td>
<td>78.3</td>
<td>81.8</td>
<td>64.8</td>
<td>66.3</td>
<td>63.0</td>
</tr>
<tr>
<td>Mother speaks language other than English</td>
<td>740</td>
<td>778</td>
<td>83.9</td>
<td>83.8</td>
<td>75.0</td>
<td>76.6</td>
<td>72.0</td>
</tr>
<tr>
<td>Mother did not complete Year 12</td>
<td>1,688</td>
<td>2,044</td>
<td>84.8</td>
<td>86.5</td>
<td>78.8</td>
<td>81.7</td>
<td>74.4</td>
</tr>
<tr>
<td>Father did not complete Year 12</td>
<td>1,890</td>
<td>2,016</td>
<td>90.0</td>
<td>90.1</td>
<td>85.9</td>
<td>87.0</td>
<td>83.6</td>
</tr>
<tr>
<td>New South Wales</td>
<td>1,615</td>
<td>1,573</td>
<td>90.3</td>
<td>90.1</td>
<td>84.4</td>
<td>86.3</td>
<td>81.8</td>
</tr>
<tr>
<td>Victoria</td>
<td>1,251</td>
<td>1,245</td>
<td>88.4</td>
<td>86.3</td>
<td>85.1</td>
<td>86.0</td>
<td>81.9</td>
</tr>
<tr>
<td>Queensland</td>
<td>1,054</td>
<td>988</td>
<td>91.4</td>
<td>90.8</td>
<td>88.0</td>
<td>87.2</td>
<td>84.3</td>
</tr>
<tr>
<td>South Australia</td>
<td>347</td>
<td>339</td>
<td>91.1</td>
<td>89.4</td>
<td>88.2</td>
<td>86.7</td>
<td>85.9</td>
</tr>
<tr>
<td>Western Australia</td>
<td>533</td>
<td>507</td>
<td>89.7</td>
<td>91.5</td>
<td>83.9</td>
<td>87.6</td>
<td>81.6</td>
</tr>
<tr>
<td>Tasmania</td>
<td>113</td>
<td>136</td>
<td>90.3</td>
<td>94.1</td>
<td>92.0</td>
<td>91.2</td>
<td>92.9</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>87</td>
<td>82</td>
<td>90.8</td>
<td>89.0</td>
<td>83.9</td>
<td>87.8</td>
<td>80.5</td>
</tr>
<tr>
<td>Australian Capital Territory</td>
<td>107</td>
<td>113</td>
<td>97.2</td>
<td>94.7</td>
<td>95.3</td>
<td>94.7</td>
<td>93.5</td>
</tr>
<tr>
<td>Capital city statistical division</td>
<td>3,194</td>
<td>3,095</td>
<td>90.6</td>
<td>89.3</td>
<td>86.2</td>
<td>86.8</td>
<td>82.9</td>
</tr>
<tr>
<td>Balance of state</td>
<td>1,913</td>
<td>1,888</td>
<td>89.5</td>
<td>90.0</td>
<td>85.4</td>
<td>87.2</td>
<td>83.3</td>
</tr>
</tbody>
</table>
User training sessions are offered by AIFS to further develop the information provided in the user manual and to allow users to interact with the LSAC Data Management team and benefit from their knowledge and experience with the data. These sessions consist of an introduction to LSAC and the newly released datasets including:

- study methodology
- introduction to the datasets
- issues for data analysts (e.g. weighting, clustering, confidentialisation)
- variable naming
- user resources (e.g. data dictionary, labelled questionnaires).

See the LSAC website for details on when training sessions are being offered.

If you have questions about the data files or variables, please submit your query via Dataverse. Click on the email icon button part-way down the Dataverse page on the right or the ‘Support’ option at the top right of the page.

**13.1 Online assistance**

An email alert list is used to convey key information and updates to users. Important information distributed via the email alert list is also stored in the data access area of the Growing Up in Australia website. This area contains:

- all reference material made available to users (in downloadable form)
- Excel data dictionary
- critical updates and alerts as distributed through the email alert list
- updates on data-user workshops.

**13.2 Getting more information**

There are several other ways to get more information about the LSAC Survey data:

- Go to the LSAC website (growingupinaustralia.gov.au/data-and-documentation) – copies of all survey instruments and various discussion and technical papers can be viewed and downloaded. You will also find the order forms for the datasets along with a growing bibliography of research papers that use the LSAC Survey data.
- If you have questions about the data files or variables, please submit your query via Dataverse. Click on the email icon button part-way down the Dataverse page on the right or the ‘Support’ option at the top right of the page.
- You can also send your queries directly to aifs-lsac@aifs.gov.au or by contacting the:
  LSAC Data Manager
  Australian Institute of Family Studies
  HWT Tower, Level 4/40 City Road
  Southbank VIC 3006
  Tel: +61 3 9214 78888
  Fax: + 61 3 9214 7839
The following publications provide more information on techniques for the analysis of longitudinal and survey data:


Information on related studies:


Appendix: LSAC variable naming conventions

This is a help sheet for data users.

### Standard input variables – attnnsxx

<table>
<thead>
<tr>
<th>a</th>
<th>Child age</th>
<th>tt</th>
<th>Topic</th>
<th>nn</th>
<th>Arbitrary number within topic</th>
<th>s</th>
<th>Subject/informant (optional)</th>
<th>xx</th>
<th>Sub-numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a = 0-1 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a = parent 1</td>
<td></td>
<td>As required for grouping of like items.</td>
</tr>
<tr>
<td></td>
<td>b = 2-3 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b = parent 2</td>
<td></td>
<td>See examples below.</td>
</tr>
<tr>
<td></td>
<td>c = 4-5 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c = study child</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d = 6-7 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p = PLE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(e = 8-9 years) etc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>m = mother</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>z = does not change with age of child</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>f = father</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t = teacher/carer</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>i = between waves respondent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**As required for grouping of like items.**

**See examples below.**

**Examples:**
- bhs13a = (b) 2–3 year old child; (hs) health status topic; (13) rating of own health status; (a) P1 is respondent
- bhs23c1, bhs23c2, bhs23c3 = 2–3 year old child’s height, weight, girth
- cse03a4a, cse03a4b = 4–5 year old child, se topic SDQ, reported by P1, 2 of the conduct subscale items

### Derived items – asm

<table>
<thead>
<tr>
<th>a</th>
<th>Child age</th>
<th>s</th>
<th>Up to 6-character mnemonic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(optional)</td>
<td>Subject/informant</td>
<td>e.g. vocab = MCDI vocabulary measure score</td>
</tr>
</tbody>
</table>

**Examples:**
- aaemp = P1 employment status when child aged 0–1 years
- bbemp = P2 employment status when child aged 2–3 years
## Household composition variables – aFnnxmmm

### Values

<table>
<thead>
<tr>
<th>Child age</th>
<th>F</th>
<th>nn</th>
<th>x</th>
<th>mmm</th>
</tr>
</thead>
<tbody>
<tr>
<td>a = 0–1 years</td>
<td>f</td>
<td></td>
<td>E.g. a</td>
<td>m1 = study child</td>
</tr>
<tr>
<td>b = 2–3 years</td>
<td></td>
<td></td>
<td>E.g. b</td>
<td>m2 = W1 P1</td>
</tr>
<tr>
<td>c = 4–5 years</td>
<td></td>
<td></td>
<td>E.g. c</td>
<td>m3 = W1 P2</td>
</tr>
<tr>
<td>d = 6–7 years</td>
<td></td>
<td></td>
<td>E.g. d</td>
<td>m4-15 = other hh members</td>
</tr>
<tr>
<td>(e = 8–9 years) etc.</td>
<td></td>
<td></td>
<td>etc.</td>
<td>1-6 = temporary hh members</td>
</tr>
<tr>
<td>z = does not change with age of child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example:**
- zf02m1 = gender (zf02) of study child (m1)
- bf01m2 = whether the W1 P1 (m2) is present (f01) when study child is aged 2–3 years (b)
- af08am = relationship to study child (f08) of mother (m) when child aged 0–1 years (a)
- df01cp1 = whether P1 (p1) when child aged 4–5 years (c) is present (f01) when child was 6–7 years (d)

### OR

<table>
<thead>
<tr>
<th>cpp</th>
<th>c = child’s age</th>
<th>and pp is</th>
</tr>
</thead>
<tbody>
<tr>
<td>m = mother</td>
<td>f = father</td>
<td>p1 = P1</td>
</tr>
<tr>
<td>p2 = P2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Parent identifier variables – apMN

These take values of 1-15 or missing.

<table>
<thead>
<tr>
<th>a</th>
<th>p</th>
<th>MN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child age</td>
<td>Parent</td>
<td>Same letters in all variables</td>
</tr>
<tr>
<td>As above</td>
<td>m = mother</td>
<td>mn</td>
</tr>
<tr>
<td>f = father</td>
<td>p1 = P1</td>
<td></td>
</tr>
<tr>
<td>p2 = P2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example:**
- ammn = member number (mn) for mother (m) when child aged 0–1 years (a)
- bplmn = member number for P1 when child aged 2–3 years