



▼  
The Longitudinal Study  
of Australian Children  
[growingupinaustralia.gov.au](http://growingupinaustralia.gov.au)

## THE LONGITUDINAL STUDY OF AUSTRALIAN CHILDREN: AN AUSTRALIAN GOVERNMENT INITIATIVE

# Data user guide

April 2023



Australian Government  
Australian Institute of Family Studies



Australian Government  
Department of Social Services



© Commonwealth of Australia 2023

With the exception of AIFS branding, the Commonwealth Coat of Arms, content provided by third parties, and any material protected by a trademark, all textual material presented in this publication is provided under a [Creative Commons Attribution 4.0 International licence \(CC BY 4.0\)](#). You may copy, distribute and build upon this work for commercial and non-commercial purposes; however, you must attribute the Commonwealth of Australia as the copyright holder of the work. Content that is copyrighted by a third party is subject to the licensing arrangements of the original owner.



The Australian Institute of Family Studies is committed to the creation and dissemination of research-based information on family functioning and wellbeing. Views expressed in its publications are those of individual authors and may not reflect those of the Australian Institute of Family Studies.

*Growing Up in Australia*: The Longitudinal Study of Australian Children is conducted in partnership between the Australian Government Department of Social Services, the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS), with advice provided by a consortium of leading researchers from research institutions and universities throughout Australia.

Australian Institute of Family Studies  
Level 4, 40 City Road, Southbank VIC 3006 Australia  
Phone: (03) 9214 7888 Internet: [aifs.gov.au](http://aifs.gov.au)

Cover photo: © iStockphoto/MicroStockHub

LSAC W9.1C2 Data user guide 2023

## Acknowledgements

*Growing Up in Australia*: The Longitudinal Study of Australian Children (LSAC) is conducted 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 from research institutions and universities throughout Australia.

The Wave 9C data files were prepared by the ABS and AIFS data processing teams and reviewed by DSS. AIFS has updated the current version of the LSAC Data User Guide.

Readers wishing to refer to this document should cite the following:

Mohal, J., Lansangan, C., Gasser, C., Howell, L., Hockey, P., Duffy, J., Renda, J., Scovelle, A., Jessup, K., Daraganova, G. & Mundy, L. (2023). *Growing Up in Australia: The Longitudinal Study of Australian Children – Data User Guide, Release 9.1C2, April 2023*. Melbourne: Australian Institute of Family Studies. doi:10.26193/QR4L6Q

The authors wish to acknowledge:

- the DSS Longitudinal Studies and the ABS for their support in providing feedback
- former AIFS Data Managers (Sebastian Mission, Mark Siphthorp) for their contribution to earlier versions of the data user guide.

## Copyright

Some of the material included or referred to in the Data User Guide is subject to copyright. For more information about copyright permissions, email us at: [aifs-lsac@aifs.gov.au](mailto:aifs-lsac@aifs.gov.au).

The Australian Institute of Family Studies (AIFS) does not guarantee the accuracy, currency or completeness of any copyright information provided in the LSAC study materials.

1. AIFS accepts no legal liability in relation to any use of the material included or referred to in the data user guide.
2. Users should ensure that they seek and obtain the appropriate copyright permissions from the owner of any copyright in this material prior to reproduction or publication of these items. Use of Pearson Test(s), Scoring Program(s) and Social Skills Improvement System (SSIS) Ratings are adapted and reproduced with permission from [NCS Pearson, Inc.](#) All rights reserved. Pearson Intellectual Property (IP) is subject to proprietary rights and prevents the reproduction of individual questions in full.
3. AIFS is not responsible for obtaining or assisting users with copyright permission.

## Data user guide updates

When	Version	Release	Suggested citation
April 2023	9.1 C2	Data User Guide for Release 9C2	Mohal, J., Lansangan, C., Gasser, C., Howell, L., Hockey, P., Duffy, J., Renda, J., Scovelle, A., Jessup, K., Daraganova, G. & Mundy, L. (2023). Growing Up in Australia: <i>The Longitudinal Study of Australian Children – Data User Guide, Release 9.1C2, April 2023</i> . Melbourne: Australian Institute of Family Studies. doi:10.26193/QR4L6Q
June 2022	9.0 C2	Data User Guide for Release 9C2	Mohal, J., Lansangan, C., Gasser, C., Howell, L., Duffy, J., Renda, J., Scovelle, A., Jessup, K., Daraganova, G. & Mundy, L. (2022). Growing Up in Australia: <i>The Longitudinal Study of Australian Children – Data User Guide, Release 9.0C2, June 2022</i> . Melbourne: Australian Institute of Family Studies. doi:10.26193/QR4L6Q
June 2021	9.0 C1	Data User Guide for Release 9C1	Mohal, J., Lansangan, C., Gasser, C., Taylor, T., Renda, J., Jessup, K., & Daraganova, G. (2021). Growing Up in Australia: <i>The Longitudinal Study of Australian Children – Data User Guide, Release 9.0C1, June 2021</i> . Melbourne: Australian Institute of Family Studies. doi:10.26193/BAA3N6
October 2020	8.0	Data User Guide for Release 8	Mohal, J., Lansangan, C., Howell, L., Renda, J., Jessup, K., & Daraganova, G. (2020). Growing Up in Australia: <i>Growing Up in Australia: The Longitudinal Study of Australian Children – Data User Guide, Release 8.0, September 2020</i> . Melbourne: Australian Institute of Family Studies. doi:10.26193/VTCZFF
December 2018	7.0	Data User Guide for Release 7	Australian Institute of Family Studies. (2018). <i>Longitudinal Study of Australian Children Data User Guide – December 2018</i> . Melbourne: Australian Institute of Family Studies.
November 2015	6.0	Data User Guide for Release 6	Australian Institute of Family Studies. (2015). <i>Longitudinal Study of Australian Children Data User Guide – November 2015</i> . Melbourne: Australian Institute of Family Studies.
November 2013	5.0	Data User Guide for Release 5	Australian Institute of Family Studies (2013) Longitudinal Study of Australian Children Data User Guide – November 2013, Melbourne.
August 2011	4.0	Data User Guide for Release 4	Australian Institute of Family Studies (2011) Longitudinal Study of Australian Children Data User Guide
April 2010	3.5	Data User Guide for Release 3.5	Australian Institute of Family Studies (2010) Longitudinal Study of Australian Children Data User Guide
August 2009	3.0	Data User Guide for Release 3	Australian Institute of Family Studies (2009) Longitudinal Study of Australian Children Data User Guide
May 2008	2.5	Data User Guide for Release 2.5	Australian Institute of Family Studies (2008) Longitudinal Study of Australian Children Data User Guide
May 2008	2.0	Data User Guide for Release 2	Australian Institute of Family Studies (2008) Longitudinal Study of Australian Children Data User Guide
November 2006	1.0	Data User Guide for Release 1	Australian Institute of Family Studies (2006) LSAC Data User Guide – Version 2.1, Melbourne.

# Contents

Acknowledgements .....	iii
Copyright .....	iii
Data user guide updates .....	iv
Abbreviations .....	viii
About this guide .....	x
<b>1. Introduction .....</b>	<b>1</b>
1.1 Objectives.....	1
1.2 Who is involved?.....	1
1.3 Timelines.....	1
1.4 Sample design.....	2
1.5 Study informants .....	2
1.6 Mother/Father data.....	3
<b>2. Instruments.....</b>	<b>4</b>
2.1 Child assessments .....	8
2.2 Response rates .....	10
<b>3. The LSAC data release.....</b>	<b>17</b>
<b>4. File structure .....</b>	<b>18</b>
4.1 Main dataset .....	19
4.2 Supplementary files.....	21
4.3 Linked data .....	24
4.4 CheckPoint Health data .....	29
<b>5. Variable naming conventions .....</b>	<b>31</b>
5.1 Questionnaire variables.....	31
5.2 Derived variables .....	35
5.3 Study child household composition variables.....	35
5.4 PLE household composition variables.....	37
5.5 Age invariant indicator variables.....	38
5.6 Indicator variables .....	39
5.7 Variable labelling convention.....	39
5.8 Missing value conventions .....	40
<b>6. Documentation .....</b>	<b>42</b>
6.1 Marked-up instruments.....	42
6.2 Frequency tables .....	43
6.3 Data dictionary .....	43
6.4 Rationale document.....	45
<b>7. Data transformations .....</b>	<b>46</b>
7.1 Transformations to ensure consistency.....	46
7.2 Transformations to update information.....	46
7.3 Summary measures for scales.....	46
7.4 Outcome Index measures.....	47
7.5 Catch-up and compound items .....	47
<b>8. Confidentialisation .....</b>	<b>49</b>
8.1 Restricted release data .....	49
8.2 General release data .....	49
<b>9. Data imputation.....</b>	<b>51</b>
9.1 Virtual roll-forward.....	51
9.2 Longitudinal contradictions.....	51
9.3 Other imputations .....	52

<b>10. Survey methodology</b>	<b>53</b>
10.1 Sample design	53
10.2 Development and testing of survey instruments	54
10.3 Data collection	55
10.4 Fieldwork response	60
10.5 LSAC COVID-19 surveys	62
<b>11. Important issues for data analysis</b>	<b>65</b>
11.1 Weighting and external validity	65
11.2 Unit of analysis	67
11.3 Age at interview	67
11.4 Time between interviews	69
11.5 Cross-cohort comparisons	70
11.6 Sample characteristics	72
<b>12. User support and training</b>	<b>76</b>
12.1 Online assistance	76
12.2 Getting more information	76
<b>References</b>	<b>77</b>
<b>Appendix A: Syntax to derive mother &amp; father items</b>	<b>79</b>
<b>Appendix B: Sample code</b>	<b>81</b>
Example: Derived items from Medicare dataset	81
Example: Sample analysis of time use diary	82
Example: Template for using SURVEYREG and SURVEYMEANS in SAS	84
<b>Appendix C: LSAC variable naming conventions</b>	<b>86</b>
<b>Appendix D: Weighting variables</b>	<b>88</b>
B cohort	88
K cohort	89

## List of figures

Figure 1: The dual cohort cross-sequential design of LSAC .....	2
Figure 2: Marked-up questionnaires .....	42
Figure 3: Wave 2 interview specification .....	42
Figure 4: Example of filtering in Excel .....	44
Figure 5: Month of interview for B cohort study families in Wave 1 to 9C .....	58
Figure 6: Month of interview for K cohort study families in Wave 1 to 9C .....	58
Figure 7: Age distribution of B cohort sample at each wave .....	68
Figure 8: Age distribution of K cohort sample at each wave .....	68
Figure 9: Distribution of time between interviews, B cohort, Wave 1 to 9C .....	69
Figure 10: Distribution of time between interviews, K cohort, Wave 1 to 9C .....	69
Figure 11: Cohort benchmarks by state, part of state and gender (Wave 1) .....	70
Figure 12: Proportion of mothers who completed Year 12, cohort benchmarks by state and part of state (Wave 1) .....	71
Figure 13: Proportion of mothers who speak a language other than English at home, cohort benchmarks by state and part of state (Wave 1) .....	71
Figure 14: Tempogram of children watching TV, video, DVD or movie while eating or drinking by wave and cohort .....	83

## List of tables

Table 1: Data collection modes by Wave .....	6
Table 2: Wave 1 to Wave 9C instrument response .....	10
Table 3: PLE Response rates from Wave 3 to Wave 8 .....	15
Table 4: Summary of RAP field response for Wave 7 .....	16
Table 5: Final response rates for Wave 7.25 .....	16
Table 6: Data release for waves and cohorts .....	18
Table 7: LSAC consents by respondents and the two cohorts across waves .....	24
Table 8: Description of the Centrelink files .....	28
Table 9: Example of variable naming with the child age indicator across waves .....	32
Table 10: Topic indicators and abbreviations .....	32
Table 11: Subject indicators for child care and education variables .....	34
Table 12: Variable names of SDQ <sup>a</sup> conduct and peer problems subscales .....	35
Table 13: Question numbers and household member characteristics .....	36
Table 14: Question numbers used in variable names for PLE household member characteristics .....	38
Table 15: Age indicators .....	39
Table 16: Missing value code frame .....	41
Table 17: Example of the weighted frequency table .....	43
Table 18: Development, pre-testing and pilot periods .....	54
Table 19: Average time in the home by the interviewer .....	55
Table 20: Fieldwork periods .....	57
Table 21: Foreign language interviews .....	60
Table 22: Sample size and response rate for each wave and cohort of LSAC .....	61
Table 23: Response status and reasons for non-response by Wave .....	62
Table 24: Wave 1 sample characteristics compared with ABS data .....	72
Table 25: Response rates for B cohort at Wave 2 to 9C by selected sample characteristics .....	74
Table 26: Response rates for K cohort at Wave 2 to 9C by selected sample characteristics .....	75

## 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

**ANU** – Australian National University

**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

**CAI** – Computer-Assisted Interview

**CASI** – Computer-Assisted Self-Interview

**CATI** – Computer-Assisted Telephone Interview

**CAWI** – Computer-Assisted Web Interview

**CBC** – Centre-Based Carer

**CRP** – Coronavirus Restriction Period. This was a reference period between March and May 2020, which provided a framework to study LSAC respondent experiences to the COVID-19 pandemic.

**CSR** – Child Self-Report

**DSS** – Department of Social Services

**EHC** – Event History Calendar

**EXEC/COGSTATE** – Executive functioning

**FAQs** – Frequently Asked Questions

**F2F** – Face-to-Face Interview

**FCF** – Family Contact Form

**FDC** – Family Day Care

**FDCQA** – Family Day Care Quality Assurance

**FTB** – Family Tax Benefit

**GJT** – Rice Test of Grammaticality Judgement measuring Specific Language Impairment (SLI)

**HBC** – Home-Based Carer

**IOBS** – Interviewer Observations

**LDC** – Long Day Care

**LOTE** – Language Other Than English

**LSAC** – Longitudinal Study of Australian Children

**MBS** – Medicare Benefit Schedule

**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

**P1** – Parent 1

**P1D** – Parent 1 During Interview Questionnaire

**P1L** – Parent 1 Leave-Behind Questionnaire

**P2** – Parent 2

**P2L** – Parent 2 Leave-Behind Questionnaire

**PBS** – Pharmaceutical Benefits 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

**SC** – Study Child

**SEIFA** – Socio-Economic Indexes for Areas

**SRS** – Simple Random Sample

**TIS** – Telephone Interpreter Service

**TQ** – Teacher Questionnaire

**TUD** – Time Use Diary

**WAI** – Who Am I

**WISC-IV** – Wechsler Intelligence Scale for Children, Fourth Edition

**YP** – Young Person

## About this guide

This data user guide is a reference tool for the users of the *Growing Up in Australia: The Longitudinal Study of Australian Children* (LSAC) datasets.

It is intended to provide the necessary information to be able to use the LSAC data. This includes information on 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. Development of the data user guide is ongoing and it is updated at each release of LSAC data to reflect new content, instruments and enhancements.

Additional resources available for users of the LSAC data include:

- questionnaires and interview specifications marked with variable names for Computer Assisted Interviews (CAI), including Computer Assisted Self Interviews (CASI) in the home, Computer Assisted Telephone Interviews (CATI) and Computer Assisted Web Interviews (CAWI) instruments)<sup>1</sup>
- a data dictionary
- technical papers on weighting, non-response and other issues
- data issues papers
- rationale papers.

These resources are all available from the [LSAC website](#).

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.

If you have any feedback, had difficulty understanding any of the data user guide's content or would like us to include additional content, please email us at: [aifs-lsac@aifs.gov.au](mailto:aifs-lsac@aifs.gov.au).

---

<sup>1</sup> 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.

# 1 Introduction

*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.

## 1.1 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.

## 1.2 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 Colmar-Brunton Social Research and I-view/NCS Pearson. Data collection for Waves 2 to 9C was undertaken by the ABS.

## 1.3 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 and one of their parents were invited to participate in the Child Health CheckPoint. This was a clinic appointment or home visit for a comprehensive, one-off physical health and biomarker module, held between Waves 6 and 7. Wave 8 data collection was conducted in 2018.

Due to the COVID-19 pandemic in 2020, the originally scheduled Wave 9 for the B and K cohorts did not go ahead. Face-to-face interviewing could not take place and was replaced with online surveys, which captured key life events and measured the impact of COVID-19 on families in LSAC. These online surveys were completed by all

Young Persons and Parents in the B and K cohorts. The first online survey for Wave 9C was called 9C1. Data for this survey were collected from October to December 2020. The second survey (9C2) began in June 2021, with data collected until September 2021, using an online mode followed by the offering of a telephone interview for those who had not completed online by a certain date. Please refer to section 10.5 for more information on the LSAC COVID-19 survey.

## 1.4 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.

**Figure 1:** The dual cohort cross-sequential design of LSAC

Cohort	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9C	
Year	2004	2006	2008	2010	2012	2014	2016	2018	2020	2021
Infant (B)	0-1 years	2-3 years	4-5 years	6-7 years	8-9 years	10-11 years	12-13 years	14-15 years	16-17 years	17-18 years
Child (K)	4-5 years	6-7 years	8-9 years	10-11 years	12-13 years	14-15 years	16-17 years	18-19 years	20-21 years	21-22 years

**Note:** Data for this survey were collected from October to December 2020. Data for the second survey (9C2) were collected from June to September 2021.

Two cohorts of children were selected from children born within two 12-month periods<sup>2</sup>:

- B cohort (infant cohort): children born March 2003–February 2004
- K cohort (child cohort): children born March 1999–February 2000

A **wave** of data collection refers to the collection of a particular set of questions from the entire sample. In LSAC, each wave of data is collected every two years. For example, Wave 8 refers to the data collected from B and K cohort study children and their parents in 2018.

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](#)).

## 1.5 Study informants

The study collects data from multiple informants:

- Study child (SC) is the cohort child. It is a term used to describe the child as the original subject of the survey.
- Study child (RAP) is the respondent who is living away from the parental home (in Wave 7 only applicable to K cohort children).
- Young person (YP) is a study child who is aged 18 years or older. In Wave 8, Young persons (K cohort only at the moment) were regarded as primary contact and approached prior to their parents for the first time.
- 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.

<sup>2</sup> Date of birth corrections applied to LSAC survey data retrospectively at Wave 7 release. The corrections affected the 12-month selection period when children of two cohorts were selected in LSAC. Refer to Section 14, LSAC data issues paper for more information.

From Wave 8, the K cohort study child is referred to as a young person replacing the terminology of study child. From 9C1 onwards, both 'YP' and 'SC' are used interchangeably but they represent the child as the original subject of the survey.

In addition, we have data about the partner of the young persons. However, the YP partner is not an informant as these data were provided by the YP.

LSAC data are also linked to the data files from the National Childcare Accreditation Council, Medicare Australia, Australian Immunisation Register (AIR), ABS Census, the National Assessment Program – Literacy and Numeracy (NAPLAN), MySchool, Australian Early Development Census (AEDC) and Centrelink. Section 4.3 of this Data User Guide contains more information about linked data.

## 1.6 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. In Wave 7, RAPs and P1 RAPs were interviewed separately. For Wave 8 this classification of participants as RAP is no longer needed as all parent interviews were conducted separately to the young person interviews for the K cohort; therefore, the distinction between young persons living with and away from their parents is not needed.

For the K cohort in Wave 8 P1, P2 and PLE were invited to complete the parent CATI, resulting in up to three parental figures being interviewed for each young person. The three parental figures retained the same parental roles as in Wave 7 (i.e. P1, P2, PLE) or the last participating wave regardless of whether their living arrangements in relation to the study child had changed (e.g. a P1 no longer living with the study child would still be referred to as P1).

For the B cohort, this change was made for 9C1; that is, the three parental figures retained the same parental roles as in Wave 8, or the last participating wave, regardless of whether their living arrangements in relation to the study child had changed.

The mother and father items are no longer available in the development of LSAC datasets in Wave 9C. When three parent figures (P1, P2, and PLE) were interviewed with identical content in 9C1, assigning mother and father roles became difficult. However, only P1 was interviewed in 9C2; therefore, it is not possible to create father items based on gender longitudinally. When a YP reaches adulthood and leaves home, the role of a mother or father decreases in importance or influence. Even though a subset of the sample will be a biological parent and a step-parent, mum/dad are the resident parents in earlier waves where the YP resided with parents. Because the content for PLE often differed from P1/P2, the mother and father items were viable until Wave 8. If data users want to analyse parent attributes by gender, they may find the computation to create mother and father items in Appendix A.

## 2 Instruments

The following data collection instruments are used to collect the LSAC study data.

- The face-to-face interview for P1 (F2F) consisted of an interviewer administered paper form in Wave 1 and a Computer Assisted Interview (CAI) for Waves 2–7. In Wave 1, P2 could complete some sections if this was more convenient. Some P1 interviews might be completed over the telephone; for example, with participating families in remote areas (see section 10.3.7). In Waves 8 and 9C1 there are no face-to-face interviews for P1.
- The P1 during interview questionnaire (PID) consisted of a self-complete paper form with items for which it was considered important to achieve high response rates and/or were considered sensitive. From Wave 4 this in-interview self-complete component was administered via 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 from Wave 4 and the CATI in Wave 8 for P1 of K cohort children. This questionnaire was not used at all in 9C1.
- 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. This questionnaire was not used at all in 9C1.
- In Wave 8, the Parent CATI was administered to all parents. This instrument replaced the CAI and CASI for P1 and the P2L for P2. PLEs had received a CATI in Waves 3–7; however, the Wave 8 CATI was designed to be administered to all parents and not just PLEs.
- The Child self-report interview (CSR) consists of survey questions answered by the study child/young person and administered by an interviewer. As part of the interview, physical measurements are taken and other assessments (such as measures of cognition or achievement) are administered. The CSR has been administered to study children in all waves from Waves 2 to 8 for the K cohort and Waves 4 to 8 for the B cohort, with the exception of Wave 4 for the K cohort and Wave 6 for the B cohort when survey questions were only administered via self-complete methods rather than interviewer administered. Where it was identified that undertaking an interview with a young person required long-distance travel, a Telephone Interview was an alternative option for the K cohort in Wave 8.
- The Audio Computer-Assisted Self-Interview (ACASI) was introduced in Wave 4 for the K cohort and Wave 6 for the B cohort. The Audio component was removed for the K cohort in Wave 7, thus the instrument was renamed the Computer-Assisted Self-Interview (CASI). The study child completes an Audio Computer-Assisted Self-Interview (ACASI) or a Computer-Assisted Self-Interview (CASI) by themselves, allowing for private completion of sensitive content.
- The Computer-Assisted Web-Interview (CAWI) was introduced in Wave 8 for the K cohort respondents and could be completed prior to their home visit. If respondents did not complete the CAWI before the home visit, they could complete an in-home version of the instrument, the Computer-Assisted Web Self-Interview (CAWSI), during home interview. This method allows sensitive content to be answered by the child in total anonymity.
- A 30-minute CAWI was the only mode used in 9C1 for all Parents (P1, P2 and PLE) and Young Persons, as the COVID-19 pandemic prevented interviewers visiting households. The YP and parents completed their web surveys independently of each other. The CAWI was also used for P1 and YP in 9C2. A telephone interview (in the form of a CATI) with a subset of the content collected in the CAWI instruments was offered as an alternate strategy to maximise the 9C2 response.

- 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. From Wave 4, the study child (K cohort only) was asked to complete one TUD. In Wave 6 the TUD was also completed by the B cohort study child. From Wave 7, the TUD was only completed by the B cohort study child. 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). In Wave 8 the K PLE was administered the same CATI instrument as the P1 and P2.
- The RAP study child is the study child respondent living away from parents (for Waves 7 and 8 of the 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. In Wave 8 there was no teacher form for the K cohort.
- 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.
- 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.

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

**Table 1:** Data collection modes by Wave

Questionnaire		Mode	Completed by	Indicator variable	W1	W2	W3	W4	W5	W6	W7	W8	W9C
													9C1 9C2
Face-to-face interview (F2F)		Paper	Parent 1	N/A	BK	-	-	-	-	-	-	-	-
Computer Assisted Interview - (CAI)		Computer	Parent 1	N/A	-	BK	BK	BK	BK	BK	BK	B	-
Parent 1 during interview (P1D)		Paper	Parent 1	[*]p1dd	BK	BK	BK	-	-	-	-	-	-
Parent 1 during interview (CASI)		Computer	Parent 1	[*]p1dd	-	-	-	BK	BK	BK	BK	B	-
Parent CATI (P1,P2, PLE)	Telephone		Parent 1	[*]p1cati	-	-	-	-	-	-	-	-	BK
			Parent 2	[*]p2cati (NA in W9C2)	-	-	-	-	-	-	-	-	-
			PLE	[*]plecati (NA in W9C2)	-	-	-	-	-	-	-	-	-
Young Person CATI		Telephone	Young Person	[*]jscati	-	-	-	-	-	-	-	-	BK
Parent 1 leave behind (P1L)		Paper	Parent 1	[*]p1scd	BK	BK	BK	-	-	-	-	-	-
Parent 2 leave behind (P2L)		Paper	Parent 2	[*]p2scd	BK	BK	BK	BK	BK	BK	BK	B	-
Computer Assisted Interview (CAI)#		Computer	Study child	[*]csrd & [*]jd40d	-	K	K	B	BK	BK	BK	BK	-
Computer assisted web-interview (CAWI)	Computer		Young Person	[*]jscawi	-	-	-	-	-	-	-	-	BK
			Parent 1	[*]p1cawi	-	-	-	-	-	-	-	-	BK
			Parent 2	[*]p2cawi (NA in W9C2)	-	-	-	-	-	-	-	-	-
			PLE##	[*]plecawi (NA in W9C2)	-	-	-	-	-	-	-	-	-
Computer assisted web self-interview (CAWSI)		Computer	Young Person		-	-	-	-	-	-	-	-	-
Audio computer-assisted self-interview (ACASI)		Computer	Study child	Need consent from: P1 [*]jd40e & Study Child (SC) [*]jd40f	-	-	-	K	K	K	B	B	-
Computer assisted self-interview (CASI)		Computer	Study child	Need consent from: P1 [*]jd40e & SC [*]jd40f	-	-	-	-	-	-	-	-	-
Time use diary (TUD)		Paper	Parent 1	N/A	BK	BK	BK	-	-	-	-	-	-
Time use diary (TUD)		Computer	Study child	Need consent from: P1 [*]jd40i & SC [*]jd40j	-	-	-	K	K	BK	B	B	-
Parent living elsewhere (PLE)		Paper - mailed out	PLE	[*]plescd	-	BK	-	-	-	-	-	-	-
Parent living elsewhere (PLE CATI)		Computer/ Telephone	PLE	[*]plescd	-	-	BK	BK	BK	BK	BK	B	-
Home-based carer (HBC)		Paper	Carer	[*]hbccbc	B	B	-	-	-	-	-	-	-
Centre-based carer (CBC)		Paper	Carer	[*]hbccbc	B	B	-	-	-	-	-	-	-

Table continued on next page →



Questionnaire	Mode	Completed by	Indicator variable	W1	W2	W3	W4	W5	W6	W7	W8	W9C 9C1 9C2
Teacher questionnaire (TQ)	Paper	Teacher	[*]tcd	K	K	BK	BK	BK	BK	B	B	-
Physical measurements (PM)	Computer	Study child	Need consent from: P1 [*]jd30d & SC [*]jd30e	BK	BK	BK	BK	BK	BK	BK	BK	-
Who am I? (WAI)	Computer	Study child	cid44a1	K	-	B	-	-	-	-	-	-
PPVT assessment (PPVT)	Computer	Study child	[*]jppvtd	K	K	BK	B	B	-	-	-	-
Matrix reasoning (MR)	Computer	Study child	[*]jd44a1	-	K	K	BK	B	B	-	-	-
Study child blood pressure (BP)	Computer	Study child	Need consent from: P1 [*]jd47a & SC [*]jd47b	-	-	-	K	K	B	B	-	-
Interviewer observations (IOBS)	Computer	Interviewer		BK	BK	BK	BK	BK	BK	BK	BK	-
Executive functioning (EXEC/CogSTATE)	Computer	Study Child	[*]jd40m	-	-	-	-	-	K	-	B	-
		Parent 1	[*]jd40n	-	-	-	-	-	-	K	-	-
Event history calendar (EHC)	Computer	Study Child	Employment :[*]jd40s7 Study :[*]jd40s8 Residential :[*]jd40s9	-	-	-	-	-	-	K	K	-

**Notes:** The indicator variable can be used to see if data is present or not for a particular instrument in the data dictionary (for more information see sections 8.6 & 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. # This was previously referred to as Child self-report (CSR). ## For Wave 8, CAWI was only given to the Young Person but for 9C1, CAWI was given to Parent 1, Parent 2, PLE and the Young Person.

## 2.1 Child assessments

### 2.1.1 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 Body Mass Index (BMI) bathroom scales (180 kg x 100 g). In Waves 4, 5, 6, 7 and 8, 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, 7 and 8 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, 7 and 8, 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.

#### 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.

### 2.1.2 'Who am I?' (WAI)<sup>3</sup>

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). For more details about the Rasch Modelling used to score the WAI, refer to the data issues paper available from the [LSAC website](#).

<sup>3</sup> The 'Who Am I?' is copyrighted by the Australian Council for Educational Research, Melbourne, 1999.

### 2.1.3 Peabody Picture Vocabulary Test (PPVT)<sup>4</sup>

A short form of the Peabody Picture Vocabulary Test, the Peabody Picture Vocabulary Test, 3rd Edition (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.

Various 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. For more details, refer to the data issues paper available from the LSAC website.

### 2.1.4 Matrix Reasoning<sup>5</sup>

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 LSAC 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. For more details, refer to the data issues paper available from the LSAC website.

### 2.1.5 Executive functioning (EXEC/CogState)<sup>6</sup>

The executive functioning of children in the K cohort was tested at Wave 6 using the [Groton Maze Learning Test](#). In Wave 7, executive functioning was also collected from the P1 of K cohort children. In Wave 8, executive functioning was only collected from the study children in the B cohort.

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 through 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.

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. Each column represents demographic information or an outcome variable. Further information about the instruments used is available in the 'Instruments' section of this guide, and in LSAC Technical Paper No. 19, *Executive Functioning – Use of CogState measures in the Longitudinal Study of Australian Children*.

<sup>4</sup> The Peabody Picture Vocabulary Test, Third Edition (PPVT-III) Form IIA is copyright by Lloy Dunn, Leota Dunn, Douglas Dunn, & American Guidance Service, Inc., 1997, and published exclusively by AGS Publishing. Permission to adapt and create a short form for LSAC was granted by the publisher. The PPVT-III – LSAC Australian Short-form was developed by S. Rothman, Australian Council for Educational Research (ACER), Melbourne, from the Peabody Picture Vocabulary Test, Third Edition (PPVT-III), Form IIA, English edition.

<sup>5</sup> The Wechsler Intelligence Scale for Children, Fourth Edition is copyrighted by Harcourt Assessment, Inc., 2004.

<sup>6</sup> 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](http://Cogstate.com)

## 2.1.6 Rice Test of Grammaticality Judgement (GJT/SLI)<sup>7</sup>

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 Receiver Operating Characteristic (ROC) of approximately 0.85.

## 2.2 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 the LSAC website.

The eligible cases in Wave 9C is different in Wave 1 to 8. 'Eligible' in the earlier waves is including the full recruited sample from Wave 1, whereas Wave 9C defines 'eligible' as the starting sample which removes those who have refused and/or withdrawn for the last three consecutive waves or were hard refusal at Wave 8 were removed from the starting sample under these criteria in earlier waves.

**Table 2:** Wave 1 to Wave 9C instrument response

Wave 1 instrument <sup>a</sup>	B cohort			K cohort		
	Eligible <sup>b</sup>	Actual <sup>c</sup>	%	Eligible <sup>b</sup>	Actual <sup>c</sup>	%
F2F	5,107	5,107	100	4,983	4,983	100
P1L	5,107	4,341	85	4,983	4,229	85
P2L	4,630	3,696	80	4,286	3,388	79
TUD 1	5,107	4,031	79	4,983	3,867	78
TUD 2	5,107	3,751	73	4,983	3,582	72
WAI	-	-	-	4,983	4,880	98
PPVT	-	-	-	4,983	4,382	88
HBC	788	342	43	-	-	-
CBC	436	233	53	-	-	-
TQ	-	-	-	4,761	3,276	69
AEDC	-	-	-	1,366	720	53
W1.5	5,061	3,573	71	4,935	3,594	73

Table continued on next page →

<sup>7</sup> Test of Early Grammatical Impairment. United States: The Psychological Corporation, A Harcourt Company.

Wave 2 instrument <sup>a</sup>	B cohort			K cohort		
	Eligible <sup>b</sup>	Actual <sup>c</sup>	%	Eligible <sup>b</sup>	Actual <sup>c</sup>	%
F2F <sup>d</sup>	5,107	4,606	90	4,983	4,464	90
P1D	4,606	4,504	98	4,464	4,358	98
P1L	4,606	3,536	77	4,464	3,495	78
P2L	4,099	3,128	76	3,804	2,949	78
TUD 1	4,606	3,477	75	4,464	3,446	77
TUD 2	4,606	3,459	75	4,464	3,460	78
PPVT	-	-	-	4,464	4,409	99
MR	-	-	-	4,464	4,402	99
PLE mail-out	400	96	24	612	199	33
HBC	791	533	67	-	-	-
CBC	1,672	1,144	68	-	-	-
TQ	-	-	-	4,447	3,632	82
W2.5	5,107	3,246	64	4,983	3,252	65
Wave 3 instrument <sup>a</sup>	B cohort			K cohort		
	Eligible <sup>b</sup>	Actual <sup>c</sup>	%	Eligible <sup>b</sup>	Actual <sup>c</sup>	%
F2F <sup>d</sup>	5,107	4,386	86	4,983	4,331	87
P1D	4,386	3,831	87	4,331	3,807	88
P2L	3,900	2,753	71	3,707	2,680	72
TUD 1	4,386	2,959	67	4,331	2,961	68
TUD 2	4,386	2,950	67	4,331	2,963	68
PPVT	4,386	4,266	97	4,331	4,273	99
WAI	4,386	4,197	96	-	-	-
MR	-	-	-	4,331	4,270	99
PLE CATI	346	272	77	510	403	79
TQ	4,114	3,395	83	4,275	3,643	85
Wave 4 instrument <sup>a</sup>	B cohort			K cohort		
	Eligible <sup>b</sup>	Actual <sup>c</sup>	%	Eligible <sup>b</sup>	Actual <sup>c</sup>	%
F2F <sup>d</sup>	5,107	4,242	82	4,983	4,164	84
CASI	4,242	4,210	99	4,164	4,116	99
P2L	3,706	2,677	72	3,512	2,645	75
CSR	4,242	4,181	99	-	-	-
ACASI	-	-	-	4,169*	4,094	99
TUD	-	-	-	4,169*	3,994	96
PPVT	4,242	4,185	99	-	-	-
MR	4,242	4,180	99	4,169*	4,103	99
PLE CATI	439	377	86	572	493	86
TQ	4,143	3,427	83	4,025	3,352	83

Table continued on next page →

Wave 5 instrument <sup>a</sup>	B cohort			K cohort		
	Eligible <sup>b</sup>	Actual <sup>c</sup>	%	Eligible <sup>b</sup>	Actual <sup>c</sup>	%
F2F <sup>d</sup>	5,107	4,085	80	4,983	3,956	79
CASI	4,077	4,010	98	3,952	3,857	98
P2L	3,512	2,444	70	3,277	2,333	71
CSR	4,026*	4,014	100	3,872	3,850	99
ACASI	-	-	-	3,873*	3,844	99
TUD	-	-	-	3,871*	3,649	94
PPVT	4,026	3,977	99	-	-	-
MR	4,027	3,985	99	-	-	-
PLE CATI	537	404	75	614	464	76
TQ	4,021	3,490	87	3,857	3,225	84
Wave 6 instrument <sup>a</sup>	B cohort			K cohort		
	Eligible <sup>b</sup>	Actual <sup>c</sup>	%	Eligible <sup>b</sup>	Actual <sup>c</sup>	%
F2F <sup>d</sup>	5,107	3,764	74	4,983	3,537	71
CASI	3,759	3,668	98	3,526	3,376	96
P2L	3,197	2,311	72	2,904	2,212	76
CSR	-	-	-	3,388	3,317	98
ACASI	3,648*	3,597	99	3,386*	3,313	98
TUD	3,649*	3,460	95	3,387*	3,071	91
EXEC	-	-	-	3,386*	3,333	98
GJT	-	-	-	3,386*	3,281	97
MR	3,648*	3,585	98	-	-	-
PLE CATI	559	398	71	554	420	76
TQ	3,678	3,102	84	3,422	2,698	79
Wave 7 instrument <sup>a</sup>	B cohort			K cohort		
	Eligible <sup>b</sup>	Actual <sup>c</sup>	%	Eligible <sup>b</sup>	Actual <sup>c</sup>	%
F2F <sup>d</sup>	5,107	3,381	66	4,983	3,089	62
P1 CASI	3,374	3,287	97	3,048	3,003	99
P2L	2,794	1,999	72	2,467	1,775	72
CSR	3,238	3,224 <sup>e</sup>	100	-	-	-
SC ACASI/CASI	3,238	3,213	99	2,978	2,941	99
W 7.25 CATI	441	55	13	451	13	3
CAI <sup>f</sup>	-	-	-	2,978	2,954	99
TUD	3,238	3,059	95	-	-	-
EXEC	-	-	-	2,995	2,624	88
PLE CATI	508	325	64	488	270	56
TQ or TCHB	3,160	2,567	81	-	-	-
EHC - Employment	-	-	-	2,978	2,931	98
EHC - Resident Living Away	-	-	-	2,978	2,915	98
EHC - Study	-	-	-	2,978	2,931	98

Table continued on next page ➔

Wave 8 instrument <sup>a</sup>		B cohort			K cohort		
		Eligible <sup>b</sup>	Actual <sup>c</sup>	%	Eligible <sup>b</sup>	Actual <sup>c</sup>	%
F2F <sup>d</sup>		5,107	3,127	61	4,983	3,037	61
P1 CASI		3,123	3,086	99	–	–	–
P2L		2,573	1,854	72	–	–	–
CSR		3,022	3,018	100	–	–	–
ACASI		3,036	3,011	100	–	–	–
CASI		–	–	–	2,708	2,656	98
CAWI		–	–	–	2,708	1,908	70
CAWSI		–	–	–	800	596	74
CAI		–	–	–	2708	2682	99
Parent CATI <sup>e</sup>	P1	–	–	–	3,015	2,635	87
	P2	–	–	–	2,424	1,681	69
	PLE	–	–	–	473	317	67
TUD		3,036	2,827	98	–	–	–
EXEC		3,027	2,995	99	–	–	–
GJT		3,015	3,007	100	–	–	–
PLE CATI		521	319	61	–	–	–
TQ		3,059	2,318	76	–	–	–
EHC – Employment		–	–	–	2,704	2,673	99
EHC – Residential		–	–	–	2,704	2,673	99
EHC – Study		–	–	–	2,704	2,674	99
Wave 9C, Survey C1 instrument <sup>a</sup>		Eligible <sup>b</sup>	Actual <sup>c</sup>	%	Eligible <sup>b</sup>	Actual <sup>c</sup>	%
YP CAWI		3,849	1,595	41	3,809	1,361	36
P1 CAWI		3,844	1,296	34	3,110	975	31
P2 CAWI		2,542	770	30	2,020	541	27
PLE CAWI		442	130	29	413	110	27
Wave 9C, Survey C2 instrument <sup>a</sup>		Eligible <sup>b</sup>	Actual <sup>c</sup>	%	Eligible <sup>b</sup>	Actual <sup>c</sup>	%
YP <sup>h</sup>		3,716	2,228	60	3,742	1,960	52
P1 <sup>h</sup>		3,710	2,199	59	3,135	1,944	62

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

<sup>a</sup> Questionnaire acronyms are detailed above in section 3, Table 1: Data collection modes by wave.

<sup>b</sup> '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).

<sup>c</sup> 'Actual' means the number of respondents for whom a form was returned.

<sup>d</sup> Response rates for Waves 2 to 7 as proportion of Wave 1 families.

<sup>e</sup> 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. Also in Wave 7 an 'in-between' wave activity was conducted to address the increase in refusals, hence W7.25 was developed.

<sup>f</sup> Introduced first time in K cohort.

<sup>g</sup> Parent CATI was introduced in Wave 8 K cohort and was the only mode of survey administration for all parents.

<sup>h</sup> CAWI was initially administered to eligible respondents. After a month of administering CAWI, both CATI and CAWI modes were used. The actual responding sample of 9C2 includes 9% B cohort and 10% K cohort YP CATI cases, 18% B cohort and 20% K cohort P1 CATI cases.

## 2.2.1 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 P1D form *must* be completed when they were in the home (resulting in a high response rate). The P1L 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 P1L 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.

In early waves of LSAC the status of a parent as either a P1, P2 or PLE was decided according to living arrangements and other circumstances at that wave. The status was then fixed at the end of Wave 7 for the K cohort or Wave 8 for the B cohort, so that these labels would not change in subsequent waves regardless of changes in living arrangements or other circumstances. This means that the Young Person may no longer be living with the P1 or even have the most significant relationship with them, but the P1 is still the parent for whom we have the most historical data and contact information.

With the young person interviewed independently in Wave 8, new procedures were implemented for collecting parent data (for K cohort only). For the first time, information was collected from the P1 and P2 via CATI with the 87% and 69% response rates respectively. During the young person's interview, the young person was asked to provide contact details for their parents (P1, P2 and PLE). The parent's data for the B cohort continued to be collected via CASI and CAI in Wave 8, and the response rates increased slightly from 97% in Wave 7 to 99% in Wave 8 based on eligible interviews.

All Parents (P1, P2 and PLE) in the B and K cohorts were asked to complete the same 30-minute CAWI in 9C1. The instrument was available for completion for a two-month period. An SMS reminder was sent to non-respondents after three weeks, and then a telephone reminder call was made to households in which there was at least one non-respondent after five weeks. The response rates were much lower than in previous waves at 33% for P1s, 29% for P2s and 28% for PLEs. P1 in the B and K cohorts completed either a 30-minute CAWI or 45-minute CATI in 9C2. SMS and email reminders were sent at two, four and 12 weeks of enumeration. The response rate was 49% for the CAWI and 12% for the CATI in 9C2.

## 2.2.2 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. In Wave 8, the CATI replaced the leave-behind form for the P2 of the young person (K cohort).



### 2.2.3 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. Table 3 summarises the PLE response rates from Waves 3 to 8.

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.

In Wave 8, the young person (K Cohort) was asked to provide the most up-to-date contact details for each of their parents, including parent living elsewhere. Refer to Wave 8 instrument table for eligible and responding PLE for K cohort.

**Table 3:** PLE Response rates from Wave 3 to Wave 8

		PLE identified during P1 interview	Eligible PLE*
<b>Wave 3</b>	B Cohort	578	346
	K Cohort	837	510
	<b>Total</b>	<b>1,415</b>	<b>856</b>
<b>Wave 4</b>	B Cohort	674	439
	K Cohort	878	572
	<b>Total</b>	<b>1,552</b>	<b>1,011</b>
<b>Wave 5</b>	B Cohort	773	537
	K Cohort	911	614
	<b>Total</b>	<b>1,684</b>	<b>1,151</b>
<b>Wave 6</b>	B Cohort	778	559
	K Cohort	817	554
	<b>Total</b>	<b>1,595</b>	<b>1,113</b>
<b>Wave 7</b>	B Cohort	732	508
	K Cohort	756 **	486 **
	<b>Total</b>	<b>1,488</b>	<b>994</b>
<b>Wave 8</b>	B Cohort	751	521
	<b>Total</b>	<b>751</b>	<b>521</b>

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.

## 2.2.4 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, 14 (58%) parents undertook an interview, while one parent (4%) refused, eight parents (17%) were not contactable, and the remaining record was finalised as having machine problems. Table 4 summarises the final RAP response rates for Wave 7.

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

	Study child		Parent	
	<i>n</i>	%	<i>n</i>	%
Responding	27	35.5	13	54.2
Refusal*	4	5.3	3	12.5
Non-contact	45	59.2	8	33.3
<b>Total</b>	<b>76</b>	<b>100</b>	<b>24</b>	<b>100</b>

**Note:** \* Includes avoidance

## 2.2.5 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.

The final response rates for Wave 7.25 are shown in Table 5.

**Table 5:** Final response rates for Wave 7.25

Field response	B cohort		K cohort		Total	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Fully responding	55	12.5	13	2.9	68	7.6
Part responding*	131	29.7	154	9.3	285	32.0
Refusal**	84	19.1	42	9.3	126	14.1
Non-contact	171	38.8	242	53.7	413	46.3
<b>Total</b>	<b>441</b>	<b>100.0</b>	<b>451</b>	<b>100.0</b>	<b>892</b>	<b>100.0</b>

**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.

# 3 The LSAC data release

Each time LSAC data are collected from the entire sample, it is considered a wave; this occurs every two years. The repeated collection of data on the responding sample results in multiple waves that allow researchers to measure change over time. Each wave of LSAC data is numbered in sequential order, with Wave 9 being the most recent.

A new *release* of the LSAC datasets is generated as additional information becomes available after each wave of data collection. It contains datasets for the new wave in addition to all previous waves (e.g. Release 9.0C2 of LSAC includes data for 9C2 as well as Wave 1 to 9C1). Each data release is given a unique number that corresponds to the wave of data being collected (e.g. Release 9.0C2 reflects data assets up to and including 9C2). With a new release, there may be some changes or enhancements to earlier waves; for example, the correction of errors, changes to naming or labelling conventions, or the addition of derived variables. Because of this, it is important to refer to the data release used when publishing because precise replication may not be possible if using earlier or later releases.

An *update* occurs when edits or additions are made to an existing release. An update to a release would result in it being reissued as a new version. You do not need to reapply in order to receive an update. If you are an authorised user, you will receive an email notification and will be able to download the updated dataset. Information about the new or changed data will be included in the notification.

AIFS, in partnership with the Australian Data Archive (ADA), is using [Dataverse](#) to facilitate access to the LSAC datasets. Dataverse is an online platform that enables the user to:

- access LSAC datasets (current and previous releases), once approved
- access LSAC data documentation, such as the Data User Guide, Data Dictionary, questionnaires and Data Issues Paper.

The LSAC datasets are available free of charge for download by approved data users from the ADA in SAS, STATA and SPSS formats.

A main wave dataset is provided combining data from all questionnaires for each wave of data. Other confidentialised information is available in the dataset at the unit record level, including demographic information, geographical characteristics, recruitment area and household identifiers, area-level variables, and meta-data related to participation in each wave of data collection. Personal or identifying information of LSAC participants is not available.

More details about how to access the data can be found in the DSS Longitudinal Studies [Data Access and Use Guidelines](#). This guide outlines the requirements for data users.

## 4 File structure

A dataset naming convention was developed to ensure that the name of the file easily signifies the data product.

Due to the range of different LSAC data products, a mixed naming convention has been applied. Main survey datasets follow a standard naming convention that includes: a reference to main survey data ('LSAC'), followed by release type (General Release 'GR' or Restricted Release 'RR'), followed by cohort ('B' or 'K'), followed by age group of study child at wave of data collection. For example:

- LSACGRB12 LSAC Main dataset, General Release, B Cohort, SC age (12 years at Wave 7)
- LSACGRK16 LSAC Main dataset, General Release, K Cohort, SC age (16 years at Wave 7)

For other data products such as household datasets and linked datasets, a mnemonic convention has been applied. File names for these products will include a mix of information relating to type of dataset, type of respondent and/or cohort and type of release. For example:

- HHGRK Household, General Release, K Cohort
- PLEHHGRB8 PLE, Household, General Release, B Cohort, SC age (8 years at Wave 5)
- TUDB10 Time Use Diary, B Cohort, SC age (10 years at Wave 6)
- MBSSC Medicare Benefit Schedule, Study Child

The following is a general representation of data files available in the general release version, with the exception of distance to coast and AEDC data.

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

Description of datasets	Main dataset for each wave		Data type
	B cohort	K cohort	
Main datasets for each wave and cohort	lsacgrb0*, lsacgrb2, lsacgrb4, lsacgrb6, lsacgrb8, lsacgrb10, lsacgrb12, lsacgrb14, lsacgrb16, lsacgrb17	lsacgrk4*, lsacgrk6, lsacgrk8, lsacgrk10, lsacgrk12, lsacgrk14, lsacgrk16, lsacgrk18, lsacgrk20, lsacgrk21	Main
Study child household	hhgrb	hhgrk	Supplementary
Parental household		phhgrk	Supplementary
P1 RAP household		p1raphhgrk16	Supplementary
PLE household	plehhgrb6, plehhgrb8, plehhgrb10, plehhgrb12, plehhgrb14	plehhgrk10, plehhgrk12, plehhgrk14, plehhgrk16	Supplementary
Event history calendar		ehcegrk16, ehcrgrk16, ehcsgrk16, ehcegrk18, ehcrgrk18, ehcsgrk18	Supplementary
Executive functioning	execbsc	execksc, execkp1	Supplementary

Table continued on next page →

Description of datasets	Main dataset for each wave		Data type
	B cohort	K cohort	
Time use diary	tudb10, tudb12, tudb14 one cleaned data file with problematic cases deleted for each cohort for Waves 1, 2 and 3 (diaryb0, diaryb2, etc.) 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.) 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.)	tudk10, tudk12, tudk14 one cleaned data file with problematic cases deleted for each cohort for Waves 1, 2 and 3 (diaryk4, diaryk6, etc.) one data file with the cases deleted from the above files after cleaning for each cohort for Waves 1, 2 and 3 (poortudsk4, poortudsk6, etc.) one data file with all cases and no data cleaning performed on them for each cohort for Waves 1, 2 and 3 (ucdiaryk4, ucdiaryk6, etc.)	Supplementary
Wave 2.5	lsacgrb3	lsacgrk7	Supplementary
Wave 3.5	lsacgrb5	lsacgrk9	Supplementary
Distance to coast^	lsacbgeodtc	lsackgeodtc	Supplementary
Child Health CheckPoint^	lsacgrcp		Substudy
AEDC^	aedc		Linked
Centrelink welfare^		isp_summary, ftb_summary, concession_cards	Linked
Description of datasets	Main dataset for each wave (each dataset contains both B and K cohorts)		Data type
Medicare Australia	mbssc, pbssc, mbsp1, mbsp2, pbsp1, pbsp2, acir		Linked
NAPLAN	lsacnaplan		Linked
MySchool	lsacmyschool		Linked

**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. ^Additional approval applies to access the data.

## 4.1 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)

Derived variables are included in the output dataset alongside the raw responses. Additionally, the main datasets contain status variables (e.g. date of interview, whether each type of form was returned, etc.), ABS Population Census and National Childcare Accreditation Council (NCAC) data, and weights.

### 4.1.1 ABS 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-based characteristics are provided at Young person main household level throughout the study, whereas indexes of Socio-Economic Indexes for Areas (SEIFA) are provided for young person and their parents household from Wave 8 onwards.

The census 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 are either linked at the Statistical Local Area (SLA) (before 2011) or the Australian Statistical Geography Standard (ASGS) level (from 2011). Where this wasn't available, the census data were linked at 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:

$$estimate = 2001Data + (2006Data - 2001Data) \times \frac{time\_since\_census}{time\_between\_census}$$

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

$$estimate = 5.6\%$$

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 which censuses were used (2001, 2006, 2011, 2016 or all of them).

### 4.1.2 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 Family Day Care Quality Assurance (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 Quality Improvement and Accreditation System (for Long Day Care centres) (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-school and after-school care arrangements were assessed by the guidelines laid out in the Outside School Hours Care Quality Accreditation (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.

Users can refer to the topic 'NCAC linked data' in the LSAC data dictionary to identify the variables in the main wave data files.

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).

## 4.2 Supplementary files

### 4.2.1 Household composition data

Household information was collected for Waves 1 to 8 of data collection detailing the family composition of each household.

- *Main household:* For Waves 1 to 8 of data collection, detailed information about every member of the household where the study child resides was collected. Information was 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 was usually collected from Parent 1 only. However, in Wave 7, if a study child has moved out of the parental household, this information was collected directly from the study child. Parent 1 was 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 'hgrb/k'.

As the study children grow older, they leave parental households to live independent lives. As the young person is the main respondent of the study, the young person is treated as the main resident of the household. All other household members are treated as people who enter or leave the household, regardless of who is reporting on the composition of the household. When the young person reported on household composition in Wave 8, the information was recorded in the main household file. The data file structure remained longitudinal across waves and one record per young person detailed the young person's household composition. The member number within the young person's household file is given for life to enable longitudinal tracking of old and new household members. The file structure allows data users to track the parental household in which the young person grew up.

For example, in Wave 8 John reported living with his girlfriend and two friends (living away from parents), then the following member numbers are assigned and information on the relevant variables will be recorded as missing for non-resident members (striketrough).

John - m1  
~~P1 - m2~~  
~~P2 - m3~~  
~~Sibling 1 - m4~~  
~~Sibling 2 - m5~~  
~~Grandparent - m6~~  
~~Aunty - m7~~  
~~Uncle - m8~~  
 Girlfriend - m9  
 Friend 1 - m10  
 Friend 2 - m11

- *Parental household:* In Wave 8, household information for K cohort families was available for up to three parents: Parent 1 regardless of whether he/she lives with the young person at the time of the interview or not; Parent 2 regardless of whether he/she separated from P1 at Waves 7 or 8; and Parent Living Elsewhere (PLE). A parental household file was introduced from Wave 8 K cohort that merges Parent 1, Parent 2 and PLE household. It is a cross-sectional file that contains the combined non-longitudinal household data for parents who no longer reside with the young person at the time of the interview. The development of this file follows the same rules as the development of the PLE household file in Waves 4-7. There are up to three parents' records (i.e. P1, P2, PLE) per young person in parental household data where available. There is no historical data provided in the parent's household file (e.g. data for those parents who were not living with the study child in Waves 1-7).
- *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. From Wave 8 onwards, PLE household information (K cohort) was integrated into Parental household data, which accommodates up to three parents, including P1/P2 and PLE.
- *P1 RAP household:* Another household composition data file available for Wave 7 of 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 'p1raphhgrb/k'. The P1 RAP household is a parental household of study children who were living away from P1 during the Wave 7 interview.

The usual detailed household data could not be collected in 9C1 as the questions are designed to be asked in a face-to-face interview and cannot be reliably translated into online questions. In previous waves, the household form could be asked reasonably quickly, as the data were rolled forward and, for many of the questions, the respondent was just asked about changes since the previous wave. As data were not rolled forward in 9C1, this confirmatory approach could not be used. Instead, respondents were asked a few key household questions, focusing on the respondent rather than other household members. The household information was included in the main wave data file for 9C1. The questions asked of Young Persons included how many people they lived with and their relationship to them, whether they have children living with them or elsewhere and the age ranges of their children, whether the household composition changed during COVID-19 restrictions, whether they or a household member has a disability, carer activities, partner status and postcode during the Coronavirus Restriction Period (CRP). Household questions for parents were similar but included a question on how the study Young Person is related to them.

## 4.2.2 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 (for example, Wave 8: ehcrgrk18 - residential, ehcegk18 - employment and ehcsgrk18 - 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. In Wave 8, the recall period was since the Wave 7 interview, or if the respondent was not interviewed in Wave 7, the two years preceding the date of the Wave 8 interview.



### 4.2.3 Executive functioning

Executive functioning data were collected from K cohort study children in Wave 6, the parents (P1) of K cohort study children in Wave 7 interviews and study children in the B cohort in Wave 8. This information is available in three separate data files:

- execksc – with the keyword KSC representing study children of K cohort
- execkp1 – with the keyword KP1 representing parents of K cohort
- execbsc – with the keyword BSC representing study children of B cohort.

The first letter of variable names in both of these data files represents the Wave-specific/child age indicator.

Further information about the CogState data collection is available in LSAC Technical Paper No. 19, *Executive Functioning: Use of CogState Measures in the Longitudinal Study of Australian Children*.

### 4.2.4 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 8* for details, available on the [LSAC website](#)). 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, 11 and 13 include detailed discussions of issues that should be considered when using the time use data. The technical papers are available on the [LSAC website](#).

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 4–5 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 6, both the K and B cohorts completed the TUD. From Wave 7, the TUD was collected only for B cohort.

Waves 4–8 had the form of an 'ABS Activity Episode' diary. These data are stored as a long file, as opposed to the wide files the previous diaries were stored as. An example of analysis using the TUD is provided in Appendix B.

### 4.2.5 Wave 2.5 data

Unlike Wave 1.5 in relation to Wave 1, families that responded to Wave 2.5 did not necessarily respond to Wave 2. Therefore, the data from the Wave 2.5 mailout is included in two separate datasets, and not merged with the Wave 2 dataset.

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. For more information, refer to the LSAC Data Issues Paper.

## 4.2.6 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.

## 4.2.7 Distance to coast data

Distance to coast has been generated for every residential address in Wave 1 to 9C 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](#), providing information on distance calculation and confidentialisation strategy. Distance to coast data are only available with restricted release data files.

## 4.3 Linked data

Over the years the LSAC data have been linked to different types of national administrative data including:

- Medicare Benefits Schedule (MBS)
- Pharmaceutical Benefits Scheme (PBS)
- Repatriation Schedule of Pharmaceutical Benefits (RPBS)
- Australian Childhood Immunisation Register (ACIR), now known as the Australian Immunisation Register (AIR)
- National Assessment Program – Literacy and Numeracy (NAPLAN)
- Australian Early Development Census (AEDC)
- Australian Curriculum Assessment and Reporting Authority (ACARA) (also known as MySchool)
- Centrelink Welfare (CLNK).

These databases are described in more detail in the following sections. Table 7 provides summary information on LSAC consents to administrative data linkage collected by respondents across waves.

**Table 7: LSAC consents by respondents and the two cohorts across waves**

Wave	Respondent	Consent for	B cohort	K cohort
1	Parent 1	Study child	MBS, PBS, ACIR	MBS, PBS, ACIR
2	New Parent 1 <sup>a</sup>	Study child	MBS, PBS, ACIR	MBS, PBS, ACIR
3	New Parent 1 Parent 1	Study child Study child	MBS, PBS, ACIR –	MBS, PBS, ACIR NAPLAN
4	New Parent 1 Parent 1	Study child Study child	MBS, PBS, ACIR NAPLAN, AEDC	MBS, PBS, ACIR NAPLAN
5	New Parent 1	Study child	MBS, PBS, ACIR, NAPLAN	MBS, PBS, ACIR, NAPLAN
6	New Parent 1 Study child	Study child Themselves	MBS, PBS, ACIR, NAPLAN –	MBS, PBS, ACIR, NAPLAN MBS, PBS
7	New Parent 1 Study child Parent 1/Parent 2	Study child Themselves Themselves	MBS, PBS, ACIR, NAPLAN – MBS, PBS, RPBS	MBS, PBS, ACIR CLNK MBS, PBS, RPBS, CLNK
8	Study child	Themselves	MBS, PBS	CLNK <sup>b</sup>

**Notes:** <sup>a</sup>Parent 1 (and/or Parent 2) might change between waves, and any new parent (new Parent 1) may join in subsequent waves. <sup>b</sup>Who did not participate or incorrectly completed Centrelink consent form in Wave 7.

### 4.3.1 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 Benefits Schedule (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.

Study children (14–15 years) of K cohort in Wave 6 and B cohort in Wave 8 were asked to consent for the first time to link their information to MBS, PBS and RPBS. In Wave 7, Parent 1 and Parent 2 themselves consented to their data linkage for the MBS, PBS and RPBS.

Linkage was successful for 93% of children (incomplete consent forms resulted in data not being released for about 400 children). Although the consent to link PBS information in Wave 1 of the LSAC study were high, limited cases were extracted. There could be several reasons for that including no records of PBS as participants may have received medicines that are not on the PBS, doctor may have provided medicines within the clinic, medications provided within a hospital is not on the PBS and relying on the pharmacy to submit the scripts through the appropriate channel to be recorded on the history of the participant PBS information.

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.

#### 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
- date of receipt of payment
- date of immunisation.

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 2022 for study children. For parents of study children, records are currently available for services between January 1996 and early 2022. 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 2022. 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.

There are simple techniques in SAS, SPSS and STATA to summarise across multiple records to create derived items from the Medicare datasets. There is some sample code provided in Appendix B.

## 4.3.2 AEDC data

Every three years since 2009, the Department of Education has undertaken a census of all children in their first year of full-time schooling. The data from the Australian Early Development Census (AEDC) is managed by the Social Research Centre. The AEDC data for B cohort children were obtained. The data contain no variable labels or value labels but these can be found in the AEDC Data Dictionary. The AEDC Data Dictionary and more information about the census can be obtained from the [AEDC website](#).

The data users are advised to refer to this [LSAC technical paper](#) for further information about the linkage process between the LSAC and AEDC data. The paper describes the process of obtaining consent, the eligible sample for data linkage and the results of the data matching.

## 4.3.3 NAPLAN data

NAPLAN tests are undertaken by all students in Years 3, 5, 7 and 9.

Parents of the B and K cohort study children were asked to provide consent to link to their child's NAPLAN data for the duration of the study. Consent was obtained from parents of K cohort children in Wave 3. Parents who did not give consent, or who did not participate at Wave 3, were asked again at Wave 4. Parent 1 of B cohort children were also asked to consent at Wave 4 on behalf of the study child. Overall, the consent rate was 96% for K cohort and 90% for B cohort.

In 2011, students were required to complete a persuasive writing task for the first time. This is a change from previous years (2010 and prior) when students were required to write a narrative or story. Due to this change in genre, post-2010 writing results should not be compared to previous years.

In 2020, NAPLAN testing was cancelled due to COVID-19. Students in Years 3, 5, 7 and 9 in 2020 did not undertake the NAPLAN assessment that year and there was no form of 'catch-up' test issued the following year. By 2020, all study children in the K cohort and close to all in the B cohort had completed their final NAPLAN

assessment. A limited number of cases (<5) in the B cohort were in Year 9 in 2020. These students would not have results for Year 9 NAPLAN unless they repeated Year 9 in 2021.

A final mop-up activity to source historical missing NAPLAN data points was undertaken in 2019 (after Wave 8) for the K cohort, and in 2021 (after Wave 9) for the B cohort. The mop-up activity aimed to backfill gaps in data and repeat test information across year levels (with appropriate consent and participation at a given wave). NAPLAN data files will not be refreshed in future waves.

The NAPLAN data linkage process and data issues are discussed in the [LSAC Technical Paper 8](#) (*Using National Assessment Program – Literacy and Numeracy (NAPLAN) data in the Longitudinal Study of Australian Children (LSAC)*). This paper should be considered when using the LSAC NAPLAN data, and is available from the [LSAC website](#).

### 4.3.4 ACARA MySchool data

The LSAC MySchool data are compiled using multiple school-level characteristics data files received from Australian Curriculum, Assessment and Reporting Authority (ACARA). ACARA is responsible for collating NAPLAN data received from Australian schools, collecting school characteristics and managing the MySchool website. Wave 8 contains year level information up to 2019.

The MySchool data linked to LSAC participants include detailed information about school performance in NAPLAN and school demographics (e.g. the school type, student population, staff numbers and financial information). School data about the schools LSAC participants attend have been linked onto the LSAC survey datasets and are available to data users. See the [LSAC website](#) for data structures, confidentialisation treatment and data consideration in analyses and interpretation of LSAC. The MySchool data files were last refreshed in Wave 8 for the B cohort and included school-level information up to 2018. Data files will not be refreshed in future waves as all respondents have now completed their NAPLAN assessment year levels.

### 4.3.5 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. Furthermore, the young person who did not take part or incorrectly completed the consent form in Wave 7 was given a catch-up consent form in Wave 8 to gain consent to access their Centrelink data. Centrelink consent flags for Parent 1, Parent 2 and SC of K cohort are available in the main wave data. For the K cohort, 81% of study children, 85% of Parent 1 and 59% of Parent 2 provided consent to link income support administrative data in Wave 7.

The data include 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 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.

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. Below is a description of the data available in the two versions of the Centrelink files.

**Table 8:** Description of the Centrelink files

Release	Centrelink files	Description	Information included
General Release	ISP_Summary	The ISP_Summary file contains data for income support payments receipt (ISP) 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.	<ul style="list-style-type: none"> <li>benefit type received by the participant for the greatest duration during the year</li> <li>number of days that the participant received an income support payment and duration they received the primary benefit type</li> <li>duration in receipt of rent assistance, home ownership status and rent type</li> <li>number of days the participant received other income while in receipt of an income support payment</li> <li>number of days the participant was partnered</li> <li>indicators for receipt of carer allowance payment and low income card</li> </ul>
General Release	FTB_Summary	The FTB_Summary file 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 are considered 'mature'; that is, the vast majority have tax data against which their entitlement can be reconciled.	<ul style="list-style-type: none"> <li>number of days the participant was eligible for FTB (in total), FTB-A and FTB-B</li> <li>number of days the participant was eligible for an ISP while eligible for FTB</li> <li>number of days customer was partnered with a primary partner while eligible for FTB</li> <li>number of days the participant was partnered with ex-partners while eligible for FTB</li> <li>count of children assessed as FTB children</li> <li>total validated adjusted taxable income (customer + primary partner + ex-partners)</li> </ul>
General Release	Concession_card	The Concession_cards file 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, this file may have overlapping episodes of concession cards for a participant.	<ul style="list-style-type: none"> <li>benefit type that qualified them for a concession</li> <li>concession card type</li> <li>number of dependent children</li> </ul>
Restricted Release	ISP_Episodic	The ISP_Episodic file holds the information for each episode of ISP receipt.	<p>In addition to the variables in the ISP_Summary file the following information is provided:</p> <ul style="list-style-type: none"> <li>entitlement rate</li> <li>activity requirements</li> <li>reason for end of payment</li> <li>earnings amount and work hours</li> <li>educational details – student status, course level and type, highest educational level before episode</li> <li>rent amount</li> <li>homelessness</li> <li>medical conditions (currently a binary indicator pending confidentialisation) and impairment rating</li> <li>vulnerability indicator</li> </ul>

Table continued on next page →

Release	Centrelink files	Description	Information included
Restricted Release	FTB_Customer_Reconciled	The FTB_Customer_Reconciled file has the same structure as the FTB_Summary file.	<p>In addition to the variables in the FTB_Summary file the following information is provided:</p> <ul style="list-style-type: none"> <li>• age, citizenship, Indigenous indicator, overseas indicator, preferred written language, remoteness area</li> <li>• number of days eligible for FTB-A (by rate type)</li> <li>• number of days eligible for of FTB-B</li> <li>• FTB-A and FTB-B pre-reconciliation eligibility amounts (paid and notional)</li> <li>• FTB-A and FTB-B post-reconciliation entitlement amounts</li> <li>• maintenance income and amount of FTB-A not paid due to MI test</li> <li>• number of days overseas</li> <li>• count of FTB shared care children</li> <li>• number of days also eligible for an ISP</li> <li>• adjusted taxable income broken down by components</li> </ul>
Restricted Release	FTB_Child_Reconciled	The FTB_Child_Reconciled file holds the reconciled data for the FTB children for which a participant received FTB payments in an entitlement year. The data contain one observation for each FTB customer; a 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.	<ul style="list-style-type: none"> <li>• age, gender, overseas indicator and duration</li> <li>• post-reconciliation durations for FTB-A and FTB-B</li> <li>• regular and shared care durations</li> <li>• FTB-A supplement amount</li> </ul>

## 4.4 CheckPoint Health 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 a 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.

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
- 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 from March to September 2014. Permission for contact was received from 3,513 families (93% of Wave 6 families and 69% of the original cohort).

The following derived items are available in the latest CheckPoint data:

- retinal photography (a non-intrusive measure of the cardiovascular system's small vessels)

- telomere length (a measure of accelerated cell division associated with age-related diseases)
- metabolomics (228 metabolic biomarkers including lipids, amino acids, and fatty acids)
- children's handwritten stories about their expected life at age 25 (including measures of vocabulary, grammar, and text content)
- new data derived from various photos and bio-samples taken during the assessment
- additional cardiovascular measures derived from retinal photos (scored using Singapore 'I' Vessel Assessment (SIVA) software)
- approximately 60 dental variables, such as the number of teeth visible, soft tissue assessment, visible dentures and caries, generated from 2D and 3D oral photos
- scores for horizontal facial measurements, facial symmetry, angles and indices, from 3D photos of the face
- high sensitivity C-Reactive Protein (hsCRP), an acute-phase protein in inflammatory processes.

Further information about Child Health CheckPoint is available from the [study website](#).



# 5 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 guide is provided in Appendix C to assist users with the variable naming convention.

## 5.1 Questionnaire variables

Most variable names follow the standard naming convention, with the exception being derived items and household composition variables.

The standard format is A tt xxxxx, where:

A = child age indicator

tt = topic indicator

xxxxx = specific question identifier.

Variable names are kept to the standard length of eight characters where possible. Variable names may exceed this length due to exceptions in the rules for child age, topic, and/or specific question identifier, as described below.

### 5.1.1 Child age indicator (alpha)

The child age indicator is the first character in the variable name and indicates the child's age. This allows for comparisons between the B and K cohorts where data have been collected for both cohorts at that age. For example:

*a* indicates the child is aged 0–1 years (B cohort in Wave 1)

*b* indicates the child is aged 2–3 years (B cohort in Wave 2)

*c* indicates the child is aged 4–5 years (B cohort in Wave 3, and the K cohort in Wave 1)

*d* indicates the child is aged 6–7 years (B cohort in Wave 4, and the K cohort in Wave 2)

*e* indicates the child is aged 8–9 years (B cohort in Wave 5, and the K cohort in Wave 3)

*f* indicates the child is aged 10–11 years (B cohort in Wave 6, and the K cohort in Wave 4)

*g* indicates the child is aged 12–13 years (B cohort in Wave 7, and the K cohort in Wave 5)

*h* indicates the child is aged 14–15 years (B cohort in Wave 8, and the K cohort in Wave 6)

*i* indicates the child is aged 16–17 years (K cohort in Wave 7)

*j* indicates the child is aged 18–19 years (K cohort in Wave 8)

*z* indicates any variable that is common across all ages.

Due to two surveys within Wave 9, a different approach was used to maintain the child age indicators in variable names. The indicators *i1/k1* and *i2/k2* were used to represent 9C1 and 9C2, respectively. For example:

*i1* indicates the child is aged 16–17 years (B cohort in 9C1)

*i2* indicates the child is aged 17–18 years (B cohort in 9C2)

*k1* indicates the child is aged 20–21 years (K cohort in 9C1)

*k2* indicates the child is aged 21–22 years (K cohort in 9C2)

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.

Table 9 shows how the child age indicator is used for the variable 'Region of residence'.

**Table 9:** Example of variable naming with the child age indicator across waves

Wave/Survey	B cohort	K cohort
1	aregion	cregion
2	bregion	dregion
3	cregion	eregion
4	dregion	fregion
5	eregion	gregion
6	fregion	hregion
7	gregion	iregion
8	hregion	jregion
9C1	i1region	k1region
9C2	i2region	k2region

## 5.1.2 Topic indicator (alpha)

The second and third characters in the variable name represent the topic indicator of the corresponding question. For example: apa01a (P1 rating of self-efficacy) has 'pa' as the second and third letters as its topic is 'Parenting'; and zhs03a (Birth weight of study child) has 'hs' as the second and third letter as its topic is 'Health status'. A list of topic indicators and their abbreviations is provided in Table 10.

**Table 10:** Topic indicators and abbreviations

Abbrev.	Topic	Scope
bl	Bullying	Information on the study child or young person's bullying victimisation and bullying perpetration
ce	Centrelink data	Statistical information about payments and services
fd	Family demographics	Demographic information relating to the family such as education, ethnicity and religion
fn	Finances	Financial information such as income and use of government benefits
ed	Education	Scales that measure the effect of study on parenting
fp	Future plans	Information on the study child or young person's future plans such as activity since leaving school, family hopes and expectations, plans to return to study, post-school plans, and relationship hopes and expectations
gd	General development	Scales that contain items from multiple domains of child development
hb	Health behaviour and risk factors	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.
he	Home education environment	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
ho	Housing	Information on housing such as number of bedrooms, tenure type and payments
hs	Health status	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
id	Identifiers	Questionnaire process variables such as sequence guides, consents and details of proxy respondents

Table continued on next page →

Abbrev.	Topic	Scope
lc	Learning and cognition outcomes	Information on the child's development in the areas of learning and cognition including language, literacy and numeracy
op	Offspring program characteristics	Information on current non-parental care for the young person's offspring such as type of child care, time in care, reason for child care and satisfaction with child care
pa	Parenting	Information on parenting styles and other information affecting parenting such as self-efficacy
pc	Program characteristics	Characteristics of the educational or child care program such as type of program, number of days or hours the child attends and staff satisfaction
pe	Parent living elsewhere	Details of the child's PLE such as the relationship to study child, interactions with resident parents and child support
pl	Parental leave in Australia	Data from the Parental Leave in Australia Survey – a nested study
pw	Paid work	Information on work status such as employment, occupation and work/family interactions
re	Relationships and study child's romantic relationships	Information on the quality of relationships primarily focused on the relationship between Parent 1 and Parent 2, but also on broader family harmony. Information on the young person's romantic relationships included their relationship status, degree of happiness in their relationship and the seriousness of their relationship.
sc	Social capital	Information on social capital such as attitudes to neighbours, neighbourhood, use of services and social support such as the extent to which the young person's need for support was met since the beginning of coronavirus restrictions, the type of support provided to the young person by the parent or close family member and the young person's need for support during coronavirus restrictions.
se	Social and emotional outcomes	Information relevant to the social and emotional development of the child such as temperament, behaviour and emotional states
tp	Teaching practices	Practices employed by teachers and child care workers in their work such as time use, use of resources and general philosophies

### 5.1.3 Specific question identifier (alphanumeric)

The specific question identifier (if required) is usually represented by the fourth to last characters in the variable name. These characters and numbers contain the necessary information to uniquely identify each item. Question identifiers are usually five digits long but may exceed this if required. Each variable has an arbitrary two-digit question number, which is 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 character of the variable name can also be an informant or subject indicator where a question is asked of or about more than one person. This indicator is usually a single character but may be two characters. The informant or subject indicators used are:

- a Parent 1
- b Parent 2
- c Study child
- f Father (or family home for census data)
- i In-between waves respondent
- m Mother
- p Parent living elsewhere
- t Teacher/Carer
- w Study child's partner
- x\_ Other biological parent of study child offspring (xa-Other biological parent of 1st Child, xb-Other biological parent of 2nd Child and xc-Other biological parent of 3rd Child)
- y\_ Study child offspring (ya-1st offspring, yb-2nd offspring and yc-3rd offspring)
- o Other adults aged 15 and over other than P1 or P2 in the household
- \_v Parent's partner who is not a study parent (av-P1's partner, mv-mother's partner, etc.).

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 areas of child care and education. These variables can be identified by the topic indicators of Program Characteristics (pc) and Teaching Practices (tp). In these cases, the prefixes a, b, c, d and e are used as the sixth character in the variable name to represent different options available at each wave depending on the child's age. This is explained further in Table 11.

**Table 11: Subject indicators for child care and education variables**

Age	Indicator					
	a	b	c	d	e	o
0–1 years	1st child care	2nd child care	3rd child care			
2–3 years	1st child care	2nd child care	3rd child care	Other child care		Any extra care
4–5 years	Main educational program	1st child care	2nd child care	3rd child care		Any extra care
6–7 years	Main educational program	Before school care	After school care		Program child would attend if attending school	Any extra care
8–9 years	Main educational program	Before school care	After school care	Child care at other times	Program child would attend if attending school	
10–11 years	Main educational program	Before school care	After school care		Program child would attend if attending school	Any extra care
12–13 years	Main educational program	Before school care	After school care	Other child care		
14–16 years	Main educational program			Other child care		
16–17 years	Main educational program					

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. Subscale or sub-subscale indicators are limited to two digits where possible.

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 12). 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 12:** Variable names of SDQ<sup>a</sup> conduct and peer problems subscales

	Wave 1 Parent 1 K cohort name	Wave 1 Teacher K cohort name	Wave 2 Parent 1 K cohort name	Wave 2 Teacher K cohort name
<b>Conduct problems</b>				
Often loses temper	cse03a4a	cse03t4a	dse03a4a	dse03t4a
Generally, well behaved, usually does what adults request	cse03a4b	cse03t4b	dse03a4b	dse03t4b
Often fights with other children or bullies them	cse03a4c	cse03t4c	dse03a4c	dse03t4c
Often argumentative with adults	cse03a4d	cse03t4d	N/A	N/A
Can be spiteful to others	cse03a4e	cse03t4e	N/A	N/A
Often lies or cheats	N/A	N/A	dse03a4f	dse03t4f
Steals from home, school or elsewhere	N/A	N/A	dse03a4g	dse03t4g
<b>Peer problems</b>				
Rather solitary, tends to play alone	cse03a5a	cse03t5a	dse03a5a	dse03t5a
Has at least one good friend	cse03a5b	cse03t5b	dse03a5b	dse03t5b
Generally liked by other children	cse03a5c	cse03t5c	dse03a5c	dse03t5c
Picked on or bullied by other children	cse03a5d	cse03t5d	dse03a5d	dse03t5d
Gets on better with adults than with other children	cse03a5e	cse03t5e	dse03a5e	dse03t5e

**Note:** <sup>a</sup> The SDQ is copyrighted by Robert Goodman, UK, 1999.

When necessary, subscale or sub-subscale indicators may exceed two-digits. Some examples of when this may occur are:

- Further subscales exist for sub-subsubscales. In this case the eighth character indicates the sub-subscale and the final character uniquely identifies the item (e.g. gpc44c**9c1**, gpc44c**9c2**).
- The number of subscales for a question number exceeds 9. In this case the subscale indicator becomes two digits followed by the unique item indicator (e.g. jpc44c**10a**, jpc44c**10b**).

## 5.2 Derived variables

The first to third characters of derived variables following the standard variable naming convention. That is, the first character is the age indicator, followed by the two character informant or subject indicator. The remaining characters are a mnemonic that relates to the subject matter of the derived item.

For example, the variable name for 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.

## 5.3 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 format A f ##xmmm, where:

A = Child age indicator

f = f, for family

## = Question number

x = Sub-question indicator (optional)

mmm = person identifier.

The child age indicator is the first character in the variable name and indicates the child's age. The second character is a 'f', indicating the household composition. The question number and sub-question indicator describe the question being responded to.

The person identifier indicates the member number within each household. 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 and above. Each household member retains the same member number throughout the study, even if they leave and re-enter the study child's home. Member 1 is denoted by 'm1' in the above convention, member 2 as 'm2' and so on as required.

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.

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 (age indicator = b), while 'cmmn' gives the member number of the mother when the child is aged 4-5 (age indicator = c).

Some further examples are provided below:

- 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)
- cf01m3 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 13:** Question numbers and household member characteristics

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

Table continued on next page →

Question number	Question
17	Has a condition or disability for six months or more (excluding mental illness) (as per Wave 2 question)
17a	Has sight problems (as per Wave 2 question)
17b	Has hearing problems (as per Wave 2 question)
17c	Has speech problems (as per Wave 2 question)
17d	Has blackouts, etc. (as per Wave 2 question)
17e	Has difficulty learning (as per Wave 2 question)
17f	Limited use of arms or fingers (as per Wave 2 question)
17g	Difficulty gripping (as per Wave 2 question)
17h	Limited use of legs and feet (as per Wave 2 question)
17i	Other physical condition (as per Wave 2 question)
17j	Other disfigurement (as per Wave 2 question)
17k	None of the above conditions (as per Wave 2 question)
17l	Mental illness
17z	Condition/disability for 6+ months (W5)(inc. mental illness)
18	Restricted in everyday activities
18a	Has difficulty breathing (as per Wave 2 question)
18b	Has chronic pain (as per Wave 2 question)
18c	Has nervous condition requiring treatment (as per Wave 2 question)
18d	Has mental illness requiring supervision (as per Wave 2 question)
18e	Has head injury (as per Wave 2 question)
18f	Has other long-term condition (as per Wave 2 question)
18g	Has other condition requiring treatment (as per Wave 2 question)
18h	None of the above restrictions (as per Wave 2 question)
19	Date began living with the study child
20	Household member was in the household for at least three months but moved in and left between current and previous waves
21	Person type
22	Young carer activities
23	Migration status

## 5.4 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. They have the format 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

The child age indicator is the first character in the variable name and indicates the child's age. The second character is a 'f', indicating the household composition. The question number and sub-question indicator describe the question being responded to.

The person identifier includes the constant 'ple' to indicate that it is the PLE household, followed by the household 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 member when a study child is 10–11 years old (age indicator is f). 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 14 shows the information that is available for each PLE.

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.

**Table 14:** Question numbers used in variable names for PLE household member characteristics

Question number	Question
01	Present for Wave
02	Gender
03	Age
04	Date of birth
05	Temporarily away from home (as per Wave 1 question)
06a	Relationship to PLE
07	Relationship to Parent 2
08	Relationship to study child
09	Country of birth
10	Year of first arrival in Australia
11	Language other than English spoken at home
12	ATSI status

## 5.5 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:

- hcid – unique identifier assigned when child was selected by Medicare Australia
- cohort – with B or K cohort
- Wave – numerical value indicating Wave 1 through to 8
- 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.

### 5.5.1 Study child unique identifier

Each study child has a single, unique identification variable to enable 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 and 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). That is:

- 1–4 indicate Infant cohort
- 5–8 indicate Child cohort.

The second digit indicates the state the child was selected from (1=NSW, 2=Vic, 3=Qld, 4=SA, 5=WA, 6=Tas, 7=NT, 8=ACT). The third digit 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.

## 5.6 Indicator variables

There are indicator variables in the main data files that show which parts of an interview were incomplete. These variables were created to flag (through yes/no values) that no data, or only partial data, exists for an instrument (e.g. 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
- the interview was only partially completed
- an individual chooses not to answer certain questions. In the CAWI, individuals can just move onto the next screen without providing an answer to each question.

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).

The primary purpose of interview characteristics variables (for example, \*p1cati, \*p2cati and \*plecati without age indicator) was to flag to data users the number of respondents (Code frame=1) responding to the survey instruments/the informant added to a survey data file. These types of variables require careful interpretation when summing responses yielding the eligible population for each informant (Code frame 0 and 1). There is an underlying issue with these types of items, as there won't be an accurate reflection of eligible number because of the enumeration methodology. For example, the Wave 8 data files contain 3,037 records, whereas the active sample eligible for the interview was 3,943. In Wave 8, the interview characteristics indicators were outputted with those who responded to the instrument/the number of data observations added to the data file. The treatment was applied to parents only (coded with values of 1 and missing) and not applied retrospectively to previous waves.

Other indicator variables include:

- '\*tcd' identify cases where a teacher form was not returned
- '\*partresp' to identify cases that were incomplete due to an interview stopping halfway as opposed to just certain sections being refused
- '\*hhresp' to identify cases where the household interview was completed.

Data users are encouraged to investigate the reasons for data being incomplete through these indicator variables. Note that the indicator variables do not follow the general variable naming convention but can be identified in the data dictionary under the topic 'Identifiers'.

## 5.7 Variable labelling convention

The variable labels in the LSAC dataset generally take the following format:

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

### 5.7.1 Age

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

**Table 15: Age indicators**

Age indicator	a	b	c	d	e	f	g	h	i	i1	i2	j	k1	k2
Age range	0/1	2/3	4/5	6/7	8/9	10/11	12/13	14/15	16/17	16/17	17/18	18/19	20/21	21/22
Wave (B cohort)	1	2	3	4	5	6	7	8	-	9C*	9C**	-	-	-
Wave (K cohort)	-	-	1	2	3	4	5	6	7	-	-	8	9C*	9C**

Note: Wave 9C, Surveys \*9C1, \*\*9C2

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.

## 5.7.2 Informant/subject

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.

## 5.7.3 Questionnaire position

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.

## 5.7.4 Construct label

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 - P1L D1 - Global Health Measure'. (0/1 is the age indicator, P1 is the informant/subject indicator, P1L 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).

## 5.8 Missing value conventions

Missing values occur when the data value is not stored for a variable and this may happen for a number of reasons. In the CAWI, respondents did not have to answer any individual question beyond the consent screen. They could simply skip a question, or press the 'next' button to move onto the next screen even if they left the whole screen blank. They did not have to indicate whether this is because they did not know the answer or whether it was because they would prefer not to answer. It is important to understand the reasons for missing values, as they can have a significant effect on any conclusions drawn from the data. The following missing value code frame has been implemented in the LSAC data.

Table 16: Missing value code frame

Code	Description
-1	Not applicable (when explicitly available as an option in the questionnaire)
-2	Don't know
-3	Refused or not answered
-4	Section refused
-5	Item non-response on web form, and reason is unknown
-9	Not asked due to one of the following reasons: <ul style="list-style-type: none"> <li>• 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).</li> <li>• 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 study child 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).</li> <li>• 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 where an informants refused to participate, a data user can use an indicator variable).</li> <li>• 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 cases where a form was partially completed, a data user can use an indicator variable).</li> </ul>
-99	Specific code for the one of the following reasons: <ul style="list-style-type: none"> <li>• Negative income (loss)</li> <li>• Before baby's birth-SC age when stopped living with PLE</li> <li>• No set amount for expected child support</li> </ul>
.	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).

## 6 Documentation

There are a number of products available to assist the user in navigating the LSAC dataset. These include the marked-up instruments, frequency tables, online data dictionary and rationale documents.

### 6.1 Marked-up instruments

The associated variable name has been added beside each question in the questionnaires and/or interview specifications. An example is shown in Figure 2.

**Figure 2:** Marked-up questionnaires

<p><b>8</b> Sometimes family members may have difficulty getting along with one another. They do not always agree and they may get angry. In general, how would you rate your family's ability to get along with one another?</p> <p>Excellent ..... <input type="checkbox"/> 1</p> <p>Very good ..... <input type="checkbox"/> 2</p> <p>Good ..... <input type="checkbox"/> 3</p> <p>Fair ..... <input type="checkbox"/> 4</p> <p>Poor ..... <input type="checkbox"/> 5</p>	<p><b>9</b> Which best describes the degree of happiness, all things considered, in your relationship?</p> <p>Perfectly happy ..... <input type="checkbox"/> 7</p> <p>Extremely happy ..... <input type="checkbox"/> 6</p> <p>Very happy ..... <input type="checkbox"/> 5</p> <p>Happy ..... <input type="checkbox"/> 4</p> <p>A little unhappy ..... <input type="checkbox"/> 3</p> <p>Fairly unhappy ..... <input type="checkbox"/> 2</p> <p>Extremely unhappy ..... <input type="checkbox"/> 1</p> <p>Not in a relationship <input type="checkbox"/> -1 ➔ Go to Question 11</p>
--	--

Labelled questionnaires have been generated for the instruments used in Wave 2 onwards. An example of this is shown in Figure 3.

**Figure 3:** Wave 2 interview specification

<p><u>B4). hs13c</u></p> <p>Show Prompt Card 9.</p> <p><b>In general, how would you say &lt;NameSC's&gt; current health is?</b></p> <p>1 Excellent</p> <p>2 Very good</p> <p>3 Good</p> <p>4 Fair</p> <p>5 Poor</p> <p>-2 Don't know</p>	<p><u>B5.4). hs14d</u></p> <p><b>Does &lt;NameSC&gt; need or use more medical care, mental health or educational services than is usual for most children of the same age?</b></p> <p>1 Yes</p> <p>2 No</p> <p>-2 Don't know</p> <p>If SC needs or uses more medical care, mental health or educational services than is usual for most children of the same age [If hs14d=1]</p>
--	---

## 6.2 Frequency tables

Weighted frequency tables have been produced for each wave of LSAC using the survey data. They contain a listing of the response categories for every variable and are useful for simple queries to particular questions. 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. Table 17 provides an example of a frequency table for the variable 'hhs55c'.

**Table 17:** Example of the weighted frequency table

14/15 - SC - ACASB 32.1.3 - Sought help from Parent (Wave 8 B Cohort)				
hhs55c	Frequency	Percentage (%)	Cumulative frequency	Cumulative percentage (%)
-9	130.191	4.16	130.191	4.16
-3	23.72135	0.76	153.9124	4.92
No	1,018.624	32.58	1,172.537	37.50
Yes	1,954.463	62.50	3,127.00	100.00

## 6.3 Data dictionary

The LSAC data dictionary contains a detailed listing of all variables, including those that have been derived or calculated. The variables are listed in the order that they appear in the dataset, starting with Wave 1.

The data dictionary is available as an online version and as an Excel spreadsheet. Therefore, the data can be easily sorted, filtered using the drop-down menus or searched according to the user's requirements.

Each record describes a single variable and includes the following fields:

- variable name
- variable name without age indicator
- topic number
- question id (i.e. variable name without age or subject/informant)
- file (each of the main datasets is 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
- 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.

The Excel main wave data dictionary also includes worksheets that provide information about EHC, TUD, Medicare, NAPLAN and ACARA data. There are separate data dictionaries for the Centrelink, Medicare, AEDC and Child Health CheckPoint data, which are available from the [study website](#) and [LSAC Dataverse](#).

### 6.3.1 Excel data dictionary

The Excel data dictionary contains two spreadsheets for the main data dictionary, one with the complete detailed listing of variable attributes and another with a shorter listing in a print-ready format. The print-ready format contains the variable name, question, responses and population fields, but other fields could easily be added by the data user if required.

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 Isacgr6 file (K cohort at Wave 2) first click on the drop-down menu in the 'File' field as shown in Figure 4 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 4) 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 4:** Example of filtering in Excel

	A	B	C	D	E	F
1	File Order	File	Wave	Cohort	Variable Name	Without age
18981	18980			B	did37b	id37b
18982	18981			B	did37b1	id37b1
18983	18982			B	did37b2	id37b2
18984	18983			B	did37b3	id37b3
18985	18984			B	did37b4	id37b4
18986	18985			B	did37b5	id37b5
18987	18986			B	did12a	id12a
18988	18987			B	did12a2	id12a2

### 6.3.2 Using wildcards for filtering

An 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 used by both these programs:

\* represents any combination of letters and characters

? represents any single character.

Some examples of the use of these wildcard characters are:

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.

### 6.3.3 Navigating the data dictionary

The following are some useful tips for navigating the data dictionary:

- Only items currently on the main datasets are included in the data dictionary.<sup>8</sup>
- 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.
- The introduction page for the data dictionary contains a list of topics and constructs that can be used for finding the information you want.
- 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 '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.

## 6.4 Rationale document

The LSAC rationale documents have been developed to assist data users by providing contextual information on the scales and items included in the LSAC datasets. The Waves 1 to 9C1 integrated rationale document presents background information on scales introduced throughout all waves of LSAC. Each of the items and scales have information on the waves they were used in, the source reference, and the scoring for the item (if applicable). If changes have been made to the item between waves, or from the original item source, these are also specified within the rationale document. The rationale documents also contain information on item and scale selection for the study and describe the reasoning for their inclusion in the study (rationales). These rationales have an included reference list that can be accessed within the [rationale report](#).

<sup>8</sup> The data dictionary reflects the variables that are included in the main datasets (i.e. Isacgrb0, Isacgrb2, Isacgrb4, Isacgrb6, Isacgrb8, Isacgrb10, Isacgrb12, Isacgrk4, Isacgrk6, Isacgrk8, Isacgrk10, Isacgrk12, Isacgrk14, Isacgrk16). Items from the study child household and the PLE household modules, the NAPLAN items and the Medicare items are not in the data dictionary.

## 7 Data transformations

The data from many of the responses to questions have been transformed to assist data users.

### 7.1 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, income 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 has been added.
- 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.

### 7.2 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 cumulative number of all the homes the study child has ever lived in.

### 7.3 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.

Some scales have different sets of items for children at different ages. In these cases, multiple versions of the same scale have been calculated, each based on the common items shared. For example, the parenting hostility scale began as a five-item measure for children aged 0-1 years 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 children aged 0-1 years, 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, data users comparing hostility between the ages of 0 and 1 year should use



the five-item version, data users comparing hostility between the ages of 0 and 7 years should use the four-item version, and data users comparing hostility between the ages of 0 and 9 years should use the three-item version.

Data users are advised to refer to the rationale document for further information about how scale items are calculated, interpretation and appropriate references.

## 7.4 Outcome Index measures

A unique component of the derivation and analysis work was the development and derivation of the LSAC Outcome Index, which is 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.

Wherever possible, the LSAC Outcome Index 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. This is in contrast to some other indices that focus on problems or negative outcomes.

The Outcome Index is only calculated for Waves 1 to 3.

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.

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*. This technical paper also contains important information about the correct use of the Outcome Index.

## 7.5 Catch-up and compound items

Catch-up and compound items are types of derived items. Catch-up items were introduced in Wave 2. They are only asked of current wave respondents who did not participate in the previous wave. Catch-up question data collected in the current wave are combined with the same data from the previous wave to create a new, more complete item. Because they ask about two years ago, catch-up items have the child's previous wave's age in the Child age column of the data dictionary, and the variable name starts with the previous wave's age indicator. Therefore, they can be identified using either the Child Age or Variable name columns in the data dictionary for a wave.

An example of a catch-up question is 'Before school care two years ago' that can be found in the data dictionary through the column 'Measure'. If a family was not interviewed in the previous wave, the instrument sequences them through a series of catch-up questions. This information, which is collected in the current wave and asks about the previous wave, is combined with previous wave data during data processing to create an item regarding the person who provided care for the child two years ago. Often, a similar question will also be asked of the current wave; for example, 'Before school care?' These questions are output with the current wave age indicator and can be identified using the wave column in the data dictionary. Please see example below of catch-up item for variables gpc64v and hpc64v from B Cohort Wave 8 data file (B14).

Example	Wave	Variable Name	Child's age	Variable Label	Topic	Construct	Measure	Question
Catchup - B Cohort Wave 8	7+8	gpc64v	12/13	12/13 - P1B CAI C2.2.1+W7 - Before-school: Parent 1	Program characteristics	Type of program	Before school care two years ago	Who provides care for child before school? I do.
Wave specific - B Cohort Wave 8	8	hpc64v	14/15	14/15 - P1B CAI C4.1.1 - Before-school: Parent 1	Program characteristics	Type of program	Before school care	Who provides care for child before school? I do.

Compound items combine data from previous waves with data from the present wave. As these combine data from multiple waves, compounding items can be identified using the Wave column in the data dictionary. Instead of referencing a single wave, for example, in Wave 8 compounding items reference multiple waves is '1+2+3+4+5+6+7+8'.

An example of a compound question is 'How many homes has study child lived in since he/she was born?', which can be searched in the data dictionary by filtering the column for 'Question' or by measure 'Number of homes Study Child has lived in since birth' or by searching for the variable 'ho03a2a'. Instead of asking this question every wave, respondents are asked whether the Study Child has moved homes since the last interview and, if so, how many homes has the child lived in since the last interview. If the child has not moved homes since the last wave, then the number of homes that the child has lived in has not changed and their answer from the previous wave is used or rolled in. If the child has moved homes since the last interview, then the number of homes lived in since the last interview is added to the previous wave's answer to calculate the total number of homes the child has lived in since birth. Below is an example of a compound item in B Cohort Wave 8.

Example	Wave	Cohort	Variable Name	Variable Label	Topic	Construct	Measure	Question
Compound - B Cohort Wave 8	1+2+3+4+ 5+6+7+8	B	hho03a2a	14/15 - P1B CASI G1.3+W1-7 - No. homes SC lived in since birth	Housing	Previous housing	Number of homes Study Child has lived in since birth	How many homes has study child lived in since he/she was born?

# 8 Confidentialisation

Confidentialisation was undertaken at different levels for the LSAC datasets. To increase availability of information while minimizing disclosure risks, a data sharing framework to differentiate the user's access level was implemented. This resulted in two datasets for each wave being generated with different levels of confidentialisation – General release and Restricted release.

## 8.1 Restricted release data

A lower level of confidentialisation is applied to the LSAC restricted release dataset, with all initial information preserved. The only information not included in this dataset is name, address and other contact details for the child, family, child care agency, educational institution, teacher and/or carer.

Access to the restricted release datasets may only be granted where data users are able to demonstrate a genuine need for the additional data and that they meet the necessary additional security requirements.

## 8.2 General release data

The general release dataset has undergone additional data confidentialisation in order to reduce the risk of re-identification of participants. In addition to the information removed for the Restricted release dataset, further confidentialisation for the general release dataset includes:

- additional items being removed
- transforming some variables
- collapsing some response categories
- top and/or bottom coding some response categories (i.e. recoding outlying values to a less extreme value).

For a complete list of confidentialised variables, users should consult the LSAC data dictionary, where these variables have been flagged in the 'Confidentialisation' column. It is important for data users to be aware that these items are eligible for confidentialisation if required but not all items may require confidentialisation in a given wave.

Confidentialisation of general release data is detailed below.

The following items have been removed:

- qualitative data provided by respondents
- census and postcode data for the location of carers and schools.

The following items have been transformed:

- postcode and local government area-assigned with an indicator so that all respondents selected in the same postcode/local government area can be identified
- date left hospital after birth – this has been transformed into the number of days between birth and hospital 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
  - rounded off to the nearest five if ANU (Australian National University) four ratings of occupational prestige
  - output at two-digit Australian Standard Classification of Occupations (ASCO) Code
  - removed if codes are in Australian Socioeconomic Index 2006 (AUSEI06)
- occupation in previous job – output at two-digit ANZSCO level
- future work expectations and desired occupation of SC – 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 have had top/bottom coding applied:

- 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
- number of household members.

LSAC assessed disclosure risk assessment of study child offspring information available in the B and K cohorts (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 the restricted release dataset whereas the information has been suppressed and is presented with -9 code frame in the general release dataset. The K cohort Young Person has also been asked about their current gender identity for the first time in Wave 8. Currently, the number of Young Persons not identifying as male or female is low and therefore some gender variables have been suppressed in the general release dataset.

In 9C1, 30% of the questions were related to COVID-19 or the CRP. Not all questions were sensitive in nature and items such as those related to behaviour during the CRP were recommended for general release. However, there were some sensitive topics where the risk of identification was high. This included questions on whether respondents had tested positive to COVID-19. Very few respondents answered 'yes' to these questions. Due to risk of disclosure, this information is only available in Restricted Release datasets. There were less than 10 cases from Wave 9C who responded that they had tested positive for COVID.

# 9 Data imputation

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.

## 9.1 Virtual roll-forward

'Roll-forward' is the term in CAI/CASI design 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 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 was 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 that from Wave 3 onwards there is a greater use of roll-forward, which reduced the number of situations where such conflicts could occur.

For privacy reasons, data collected at previous waves could not be rolled forward into the online surveys used in 9C1 and 9C2. Similarly, data collected in 9C1 could not be rolled forward into 9C2. Only the name of the respondent was rolled forward and additionally, the name of the Young Person was rolled forward into the Parent instrument. This was to confirm that the correct person was completing the instrument and that parents were referring to the correct Young Person when answering questions about them. One of the design considerations for both 9C1 and 9C2 was to minimise questions that required prior knowledge. For example, it was not possible to ask follow-up questions about a partner that had been mentioned in previous waves.

## 9.2 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 but 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).

## 9.3 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](#).
- **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:** 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.

Further details of these imputations are given in the Data Issues Paper available from the LSAC website.

# 10 Survey methodology

LSAC employs a cross-sequential design that follows two cohorts of children:

- initially aged 0–1 years in 2004 (B cohort)
- initially aged 4–5 years in 2004 (K cohort).

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, weighting and non-response analysis are given in various LSAC technical papers available from the LSAC website.

## 10.1 Sample design

A two-stage clustered sample design was employed, first selecting postcodes and then children. The clustered design allowed analysis of children within communities and produced 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)
- children born March 1999–February 2000 (K cohort).

### 10.1.1 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 a 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.

## 10.2 Development and testing of survey instruments

### 10.2.1 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.

**Table 18:** Development, pre-testing and pilot periods

Wave or Survey	Development began	Pre-testing	Pilot
1	March 2002	Small-scale pre-testing occurred in September 2002 to October 2002	A pilot test with about 50 families from each cohort was conducted in March to April 2003
2	July 2004	Small-scale pre-testing occurred in September 2004 to October 2004	A pilot test with 86 families was conducted in April 2004
3	March 2006	Pre-testing occurred in a number of stages from mid 2006 to March 2007	No pilot test was required
4	February 2008	Pre-testing occurred in a number of stages from mid-August 2008 to June 2009	No pilot test was required
5	February 2010	Pre-testing occurred in a number of stages from mid-June 2009 to March 2010	No pilot test was required
6	May 2012	Pre-testing occurred in a number of stages from August 2012 to September 2013	No pilot test was required
7	May 2014	Pre-testing occurred in a number of stages from August 2014 to September 2014	No pilot test was required
8	February 2016	Cognitive testing was conducted in August and September 2016	No pilot test was required
9C	January 2019 – 9C1	Cognitive testing was conducted in April 2019 for the planned Wave 9 instruments. Although this wave never went to field, some of the findings of the cognitive testing informed the development of the subsequent 9C1 CAWI instrument.	No pilot test was required
	October 2020 – 9C2	Limited and informal cognitive testing was conducted in April 2021.	No pilot test was required

### 10.2.2 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: August – November 2003 (526 families interviewed)
- Wave 2: September – November 2005 (423 families interviewed)
- Wave 3: July – October 2007 (420 families interviewed)
- Wave 4: July – October 2009 (387 families interviewed)
- Wave 5: July – August 2011 (451 families interviewed)
- Wave 6: June – August 2013 (351 families interviewed)



- Wave 7: June–September 2015 (309 families interviewed)
- Wave 8: June–September 2017 (269 families interviewed)
- Wave 9 (B Cohort): June–September 2019 (149 families interviewed)
- Wave 9 (K Cohort): February–March 2020 (56 YPs interviewed prior to ceasing further interviews due to COVID-19)

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

Due to the short lead time, it was not possible to conduct a dress rehearsal for Wave 9C, and many aspects of the enumeration strategy and content remained untested until fieldwork began. Skirmish testing was conducted in August 2020 with seven Parents and 15 Young People aged 15–22, all of whom were members of the public and not study participants. The purpose of the skirmish testing was to investigate how easy it was to navigate through the questionnaire and answer the questions, suggestions for improvement, whether there were any uncomfortable topics, comprehension of certain key concepts and other general comments.

## 10.3 Data collection

### 10.3.1 Interview length

Table 19 indicates the average time allowed for time in the home by the interviewer. It also includes the actual time required for the interviews with both B and K cohort for each wave.

**Table 19:** Average time in the home by the interviewer

Wave	Average time allocated for 'time in the home'	Actual time	
		B cohort	K cohort
1	126 minutes	90 minutes	150 minutes
2	90 minutes	66 minutes	85 minutes
3	110 minutes	91 minutes	98 minutes
4	110 minutes	102 minutes	108 minutes
5	110 minutes	98 minutes	98 minutes
6	110 minutes	108 minutes	116 minutes
7	110 minutes	114 minutes	115 minutes
8	110 minutes	110 minutes	113 minutes

In Wave 9C, a 30-minute (roughly) online survey per respondent was planned for two rounds of surveys. As an alternative to the online form in 9C1, young persons in 9C2 were provided a 45-minute phone interview (30-minute in the case of Parent 1). In 9C2, young persons spent an average of three minutes above and six minutes below, whereas Parent 1 spent six and 12 minutes below for online and telephonic interviews, respectively, than the allocated interview time.

### 10.3.2 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.

For Wave 8, due to the differences in methodology across the two cohorts, separate training sessions were held for the B and K cohort interviews. B cohort training was conducted between 27 February and 20 September 2018, and 207 interviewers attended one of five three-day training sessions. For the K cohort, initially four three-day training sessions were held between 26 March and 6 April 2018 and were attended by 100 interviewers. Where possible in these training sessions and in the practice activities, new LSAC interviewers were paired with experienced LSAC interviewers. An additional six K cohort training sessions were conducted between 16 April and 18 May 2018 and were attended by 98 interviewers who were newly recruited to the ABS for the purpose of conducting Wave 8 K cohort interviews. Due to not having previous experience of LSAC or ABS procedures these training sessions included an additional day and were thus conducted over four days. Experienced LSAC interviewers attended these training sessions and assisted with delivering some of the training modules. In addition, these new interviewers were assigned a mentor, an experienced LSAC interviewer, who could provide information and support throughout fieldwork, and given the opportunity to observe an interview being conducted. ABS staff conducted the training.

There were no interviews in 9C1, as data collection was all completed online. However, the second reminder was in the form of a telephone call conducted by ABS interviewers. These interviewers, recruited to conduct the reminder calls, received detailed Interviewer Instructions, including a comprehensive set of Frequently Asked Questions (FAQs), that helped them to respond to any questions from respondents and to help with any technical difficulties experienced. There was no face-to-face training for 9C1. For 9C2, 60 interviewers from the ABS were trained in a series of two-day training courses. Because of COVID restrictions, ABS staff were unable to travel for face-to-face training. Instead, in July 2021, training was conducted via Skype, with ABS staff in the Canberra office and interviewers in the Geelong office or in their homes. Interviewers who had not worked on LSAC previously were given background training in LSAC before the two-day course commenced.

### 10.3.3 Fieldwork periods

In 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 then 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 for Wave 1 continued throughout 2004.

In Wave 2, 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.

Fieldwork for Wave 3 was organised the same as for Wave 2.

From Wave 4 onwards, the focus was more on the location of the sample and interviewers with less emphasis given to following interviews within the set phases. This change was implemented to assist the efficiency of work allocations to interviewers.

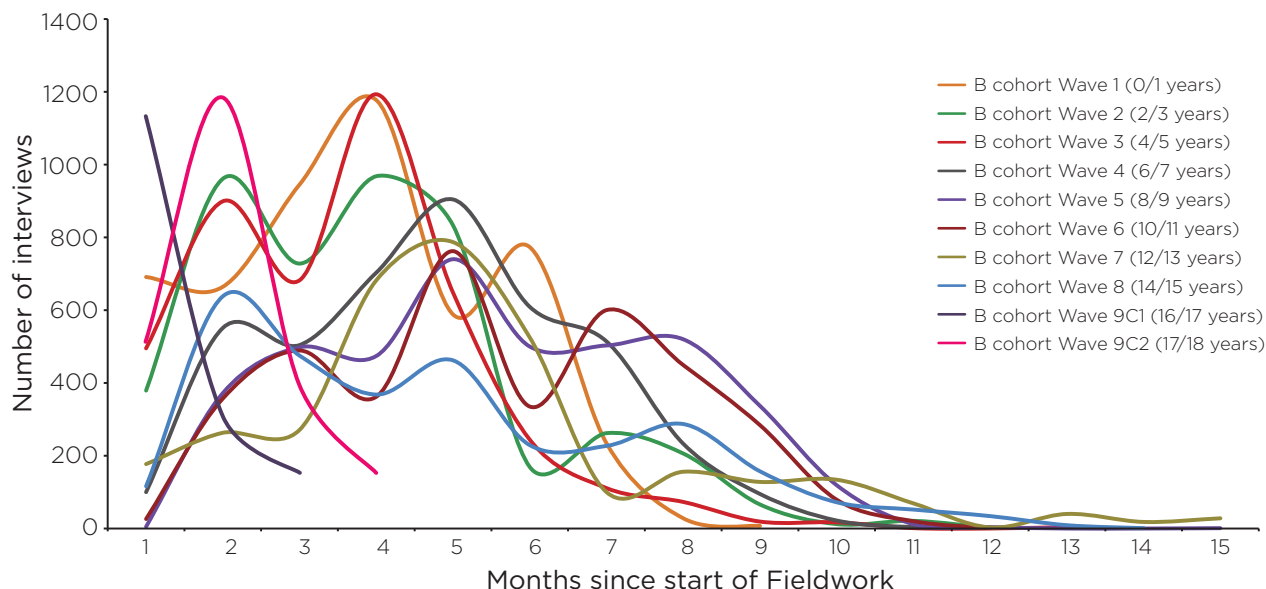
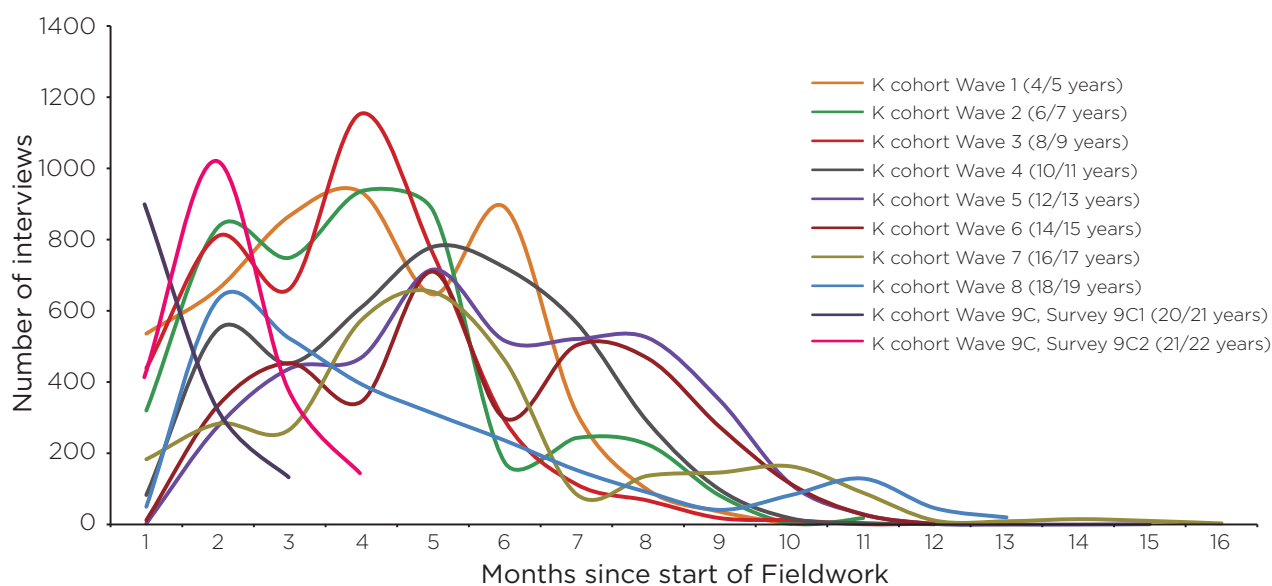
Table 20 indicates the fieldwork time period for each cohort and wave. Figures 5 and 6 show the distribution of interviews over time for each cohort and wave.

The figures show that the distribution of interviews for Wave 7 decreased greatly in September (six months from the start of fieldwork for this wave). This can be mostly attributed to the ABS Census Post Enumeration Survey priorities during this time and, as a result, fieldwork for Wave 7 was extended beyond the originally planned end in December 2016.

Wave 8 fieldwork was organised based on the location of the sample and interviewers. The online survey for 9C1 was open for two months, from early October to early December 2020. The online survey for 9C2 was open for three months, from late June to late September 2021. These timelines were the same for both the B and K cohorts.

**Table 20: Fieldwork periods**

Wave or Survey		B cohort		K cohort	
		Period	Length	Period	Length
1		Mar 2004–Nov 2004	9 months	Mar 2004–Jan 2005	11 months
2		Mar 2006–Mar 2007	12 months	Apr 2006–Feb 2007	11 months
3		Apr 2008–Apr 2009	13 months	Apr 2008–Apr 2009	13 months
4		Mar 2010–Feb 2011	12 months	Mar 2010–Feb 2011	12 months
5		Mar 2012–May 2013	15 months	Mar 2012–May 2013	15 months
6		Mar 2014–Feb 2015	12 months	Mar 2014–Feb 2015	12 months
7		Apr 2016–Jun 2017	15 months	Apr 2016–Jul 2017	16 months
8		Mar 2018–Apr 2019	14 months	Apr 2018– Apr 2019	13 months
9C	9C1	Oct 2020–Dec 2020	2 months	Oct 2020–Dec 2020	2 months
	9C2	Jun 2021–Sep 2021	3 months	Jun 2021–Sep 2021	3 months

**Figure 5:** Month of interview for B cohort study families in Wave 1 to 9C**Figure 6:** Month of interview for K cohort study families in Wave 1 to 9C

### 10.3.4 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.

## Wave 9C

Pre-interview packages were sent to all eligible informants. Following this, an email, containing the same information as the pre-interview package, was sent to all those for whom the ABS had an email address.

For 9C1, the first reminder was an SMS sent two weeks into the fieldwork period to all those for whom there was a mobile phone number. The second and final reminder was a telephone call made between week 5 and week 7 of the field period. For 9C2, three reminders were sent via SMS or email approximately two, four and 12 weeks into the fieldwork period. The final contact at week 12 of enumeration was an SMS or email contact, with a one week follow-up via telephone for non-responses.

## 10.3.5 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 (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.

### Waves 3-8

The details around foreign language interviews for Waves 1-8 are provided in Table 21.

**Table 21:** Foreign language interviews

Wave <sup>a</sup>	Interviews needing an interpreter	Number of languages	Family or friends assisted	Method used	
				ABS interpreter	TIS interpreter
1	310	50+	138	76	96
2	110	23	58	37	15
3	97	24	58	31	8
4	93	26	50	29	14
5	81	18	47	24	10
6	64	17	42	18	4
7	55	19	31	21	3
8	53	14	20	31	2

**Note:** <sup>a</sup> There were no interpretation facilities for Wave 9C.

### 10.3.6 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. For more details, refer to the weighting and non-response technical papers on the LSAC website.

### 10.3.7 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
- 49 (0.76%) in Wave 7
- 44 (0.75%) in Wave 8.

## 10.4 Fieldwork response

### 10.4.1 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](#).

## 10.4.2 Response in later waves

Table 22 summarises the response from families in later waves, using the Wave 1 sample and 'available' sample as the bases for comparisons.

**Table 22:** Sample size and response rate for each wave and cohort of LSAC

	B cohort			K cohort			Total		
	<i>n</i>	Resp. rate of Wave 1 (%)	Resp. rate of available sample (%)	<i>n</i>	Resp. rate of Wave 1 (%)	Resp. rate of available sample (%)	<i>n</i>	Resp. rate of Wave 1 (%)	Resp. rate of available sample (%)
<b>Main waves or survey</b>									
Wave 1 original	5,107	100.0		4,983	100.0		10,090	100.0	
Wave 2 available <sup>a</sup>	5,047	98.8		4,913	98.6		9,960	98.7	
Wave 2 responding <sup>b</sup>	4,606	90.2	91.2	4,464	89.6	90.9	9,070	89.9	91.1
Wave 3 available	4,971	97.3		4,829	96.9		9,800	97.1	
Wave 3 responding	4,386	85.9	88.2	4,331	86.9	89.7	8,717	86.4	89.0
Wave 4 available	4,929	96.5		4,774	95.8		9,703	96.2	
Wave 4 responding	4,242	83.0	86.0	4,169	83.7	87.3	8,411	83.4	86.7
Wave 5 available	4,884	96.6		4,735	95.0		9,619	95.3	
Wave 5 responding	4,085	80.0	91.1	3,956	79.4	83.5	8,041	79.7	83.6
Wave 6 available	4,483	87.8		4,395	88.2		8,878	88.0	
Wave 6 responding	3,764	73.7	84.0	3,537	71.0	80.5	7,301	72.4	82.2
Wave 7 available	4,318	84.6		4,176	83.8		8,494	84.2	
Wave 7 responding	3,381	66.2	78.3	3,089	62.0	74.0	6,470	64.1	76.2
Wave 8 available	4,030	78.9		3,943	79.1		7,973	79.0	
Wave 8 responding	3,127	61.2	77.6	3,037	60.9	77.0	6,164	61.1	77.3
9C1 available	3,849	75.4		3,809	76.4		7,658	75.9	
9C1 responding	2,017	30.4	52.4	1,789	35.9	47.0	3,806	37.7	49.7
9C2 available	3,716	72.8		3,742	75.1		7,458	73.9	
9C2 responding	2,688	52.6	72.3	2,463	49.4	65.8	5,151	51.1	69.1

Table continued on next page →

	B cohort			K cohort			Total		
	<i>n</i>	Resp. rate of Wave 1 (%)	Resp. rate of available sample (%)	<i>n</i>	Resp. rate of Wave 1 (%)	Resp. rate of available sample (%)	<i>n</i>	Resp. rate of Wave 1 (%)	Resp. rate of available sample (%)
<b>Between waves</b>									
Wave 1.5 sent	5,061	99.1		4,935	99.0		9,996	99.1	
Wave 1.5 returned	3,573	70.0	70.6	3,584	71.9	72.6	7,157	71.0	71.6
Wave 2.5 sent	4,859	95.1		4,712	94.6		9,571	94.9	
Wave 2.5 returned	3,268	63.5	64.0	3,287	65.5	66.0	6,555	63.4	65.0
Wave 3.5 sent	4,772	93.4		4,641	93.1		9,413	93.3	
Wave 3.5 returned	3,012	59.0	63.1	2,972	59.6	64.0	5,984	59.3	63.6

**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. <sup>a</sup> Available sample excludes those who opted out of the study between waves. Some additional families also opted out permanently during the fieldwork process. <sup>b</sup> Those who had a home visit.

**Table 23:** Response status and reasons for non-response by Wave

Response status	Wave 2		Wave 3		Wave 4		Wave 5		Wave 6		Wave 7		Wave 8 <sup>a</sup>	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Responding	9,070	91.1	8,717	89.0	8,411	86.6	8,041	83.6	7,301	82.2	6,470	76.2	5,835	73.2
Refusal	284	2.8	436	4.4	637	6.6	774	8.0	938	10.6	1,118	13.2	1,483	18.6
Non-contact	540	5.4	552	5.6	526	5.4	715	7.4	555	6.3	803	9.5	578	7.2
Away entire enumeration period	61	0.6	93	1.0	135	1.4	88	0.9	39	0.4	34	0.4	72	0.9
Death of study child	5	0.1	1	0.01	0	0	1	0.01	3	0.0	2	0.0	3	0.0
Other <sup>b</sup>	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	43	0.5	67	0.8	2	0.0
<b>Total starting sample</b>	<b>9,960</b>	<b>100.0</b>	<b>9,799</b>	<b>100.0</b>	<b>9,709</b>	<b>100.0</b>	<b>9,619</b>	<b>100.0</b>	<b>8,879</b>	<b>100.0</b>	<b>8,494</b>	<b>100.0</b>	<b>7,973</b>	<b>100.0</b>

**Notes:** Families are reported for Waves 2-7. <sup>a</sup> Wave 8 figures are reported for families for B cohort and Young Person for K cohort. <sup>b</sup> Includes OH&S and machine problems.

In Wave 9C, of Young Persons (B and K cohorts), 2,958 (38.6%) responded to 9C1 and 4,188 (56.2%) responded to 9C2. Two hundred and seventy-four (3.6%) Young Persons in 9C1 and 545 (16.7%) in 9C2 did not respond, with a reason received for non-response. Additionally, 4,426 (57.8%) Young Persons in 9C1 and 2,725 (83.3%) in 9C2 did not respond with no reason provided. Overall, there were 3,806 responding households (49.7%) in 9C1 and 5,151 responding households (69.0%) in 9C2. This is the number of households in which at least one person responded whether they live in the same household or not. In 849 of these for 9C1 and 963 of these for 9C2, a parent responded but not the Young Person.

## 10.5 LSAC COVID-19 surveys

As mentioned in section 1.3, Wave 9C had two online surveys: 9C1 and 9C2. Data for 9C1 were collected from October to December 2020. Data for 9C2 were collected from June to September 2021. 9C2 was designed to be a follow-up to 9C1, which would capture shorter-term impacts of the coronavirus pandemic and recent natural disasters on key indicators.

Fieldwork for the Wave 9 K cohort dress rehearsal and Main Wave 9 B cohort was due to start in April, around the time that state and federal government restrictions aimed at stopping the spread of COVID-19 were implemented. This necessitated the immediate cessation of all fieldwork activities, without any knowledge of when these activities could be reinstated.



As both cohorts were undergoing rapid transitions in their development, it was considered vital to collect information on key life events from the participating families at around this time, rather than waiting two years when the next wave would be due. There was also clear value in measuring the impact of the pandemic on the Young People, as well as the impacts of the recent natural disasters and extreme weather events, such as the 2019/20 bushfires.

After considering different options, it was decided that the data could be collected most effectively and efficiently through a 30-minute online survey (CAWI), administered to all Young Persons and their parents.

There were two separate instruments developed for 9C1 – one for Parents and one for Young Persons. Although the instruments included some sequencing to ensure respondents only saw questions that were relevant to them, both cohorts received the same instrument. For each Young Person, up to three parental figures were sent the study information in 9C1. These were the same three parental figures as identified in Wave 8.

Following a low response rate in 9C1, a CATI mop-up interview was also introduced in 9C2. Young Persons and Parent 1s were invited to participate. Parent 2s and Parents Living Elsewhere were omitted to streamline the sample review process and devote more resources to tracking the Young Persons and Parent 1s. As with 9C1, survey content focused on key measures and following the impacts of COVID-19.

The original planned Wave 9 interview was to last approximately 75 minutes for B cohort parents, 20 minutes for K cohort parents and 90 minutes for YPs from both cohorts. Due to the requirement to reduce the interview significantly and add items pertaining to COVID-19 and natural disasters/extreme weather events, only key measures could be included.

The COVID-19 material focused on:

- Changes in living arrangements
- Support required from parents/other family members
- Support or help received
- Accessing medical services
- Partner relationships
- Relationship with study child (parent only)
- Social media use (young person only)
- Stresses and difficulties
- Sleeping (young person only)
- Changes in carer activities
- Ability to study
- Working from home
- Home schooling
- Employment/self-employment
- Government payments
- Housing
- Experience of restriction
- COVID-19 test
- Time in quarantine/self-isolation (9C2).

For 9C1, when answering questions related to the coronavirus restrictions, respondents were asked to consider the initial restriction period, which was approximately March–May 2020. For 9C2, when answering questions related to the coronavirus restrictions, respondents were asked to consider the time since December 2020 (roughly the period since their last interview). A small number of catch-up questions were also included for respondents who missed 9C1. Respondents who answered catch-up questions were asked to consider the initial restriction period, which was approximately March–May 2020.

Other topics asked not in relation to COVID-19 included:

- Household harmony
- Social support
- Trust (young person only)
- General health

- Engagement in community activities
- Professional help or hospital admission for physical or mental health problems
- Disability (self or household members)
- Physical activity
- Smoking and alcohol consumption (young person only)
- Illicit drug use (young person only)
- Mental health (young person only)
- Bullying (young person only)
- Gambling
- Current educational pathways and future plans (young person only)
- Safety at work
- Availability for work
- Gig economy activities (young person only)
- Income
- Exposure to extreme weather events and natural disasters
- Stressful life events
- Health behaviours
- Optimism.

Due to sensitivity of the items or difficulty administering over the phone, the CATI instrument in 9C2 included a reduced set of questions on relationships, while questions about illicit drug use, gambling, bullying and homelessness were omitted altogether.

# 11 Important issues for data analysis

An LSAC data issues paper for Waves 1 to 9C is available and is available from the LSAC website. This paper provides details of all issues that have been identified over the course of the study.

## 11.1 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.

### 11.1.1 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.

### 11.1.2 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.

### 11.1.3 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 the LSAC 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 Appendix D.

### 11.1.4 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.

### 11.1.5 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).

### 11.1.6 Software

There is now a range of software available 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.

Appendix B provides a template for using SURVEYREG and SURVEYMEANS procedures in SAS software.

## 11.2 Unit of analysis

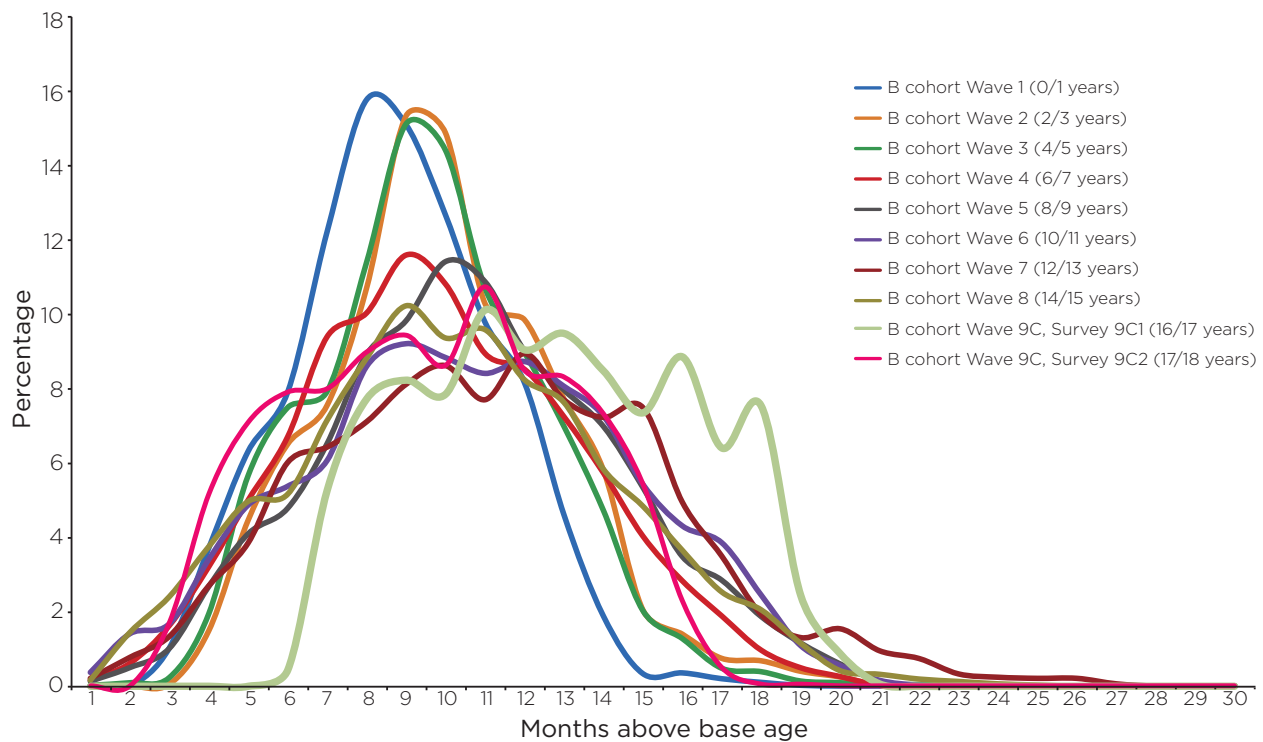
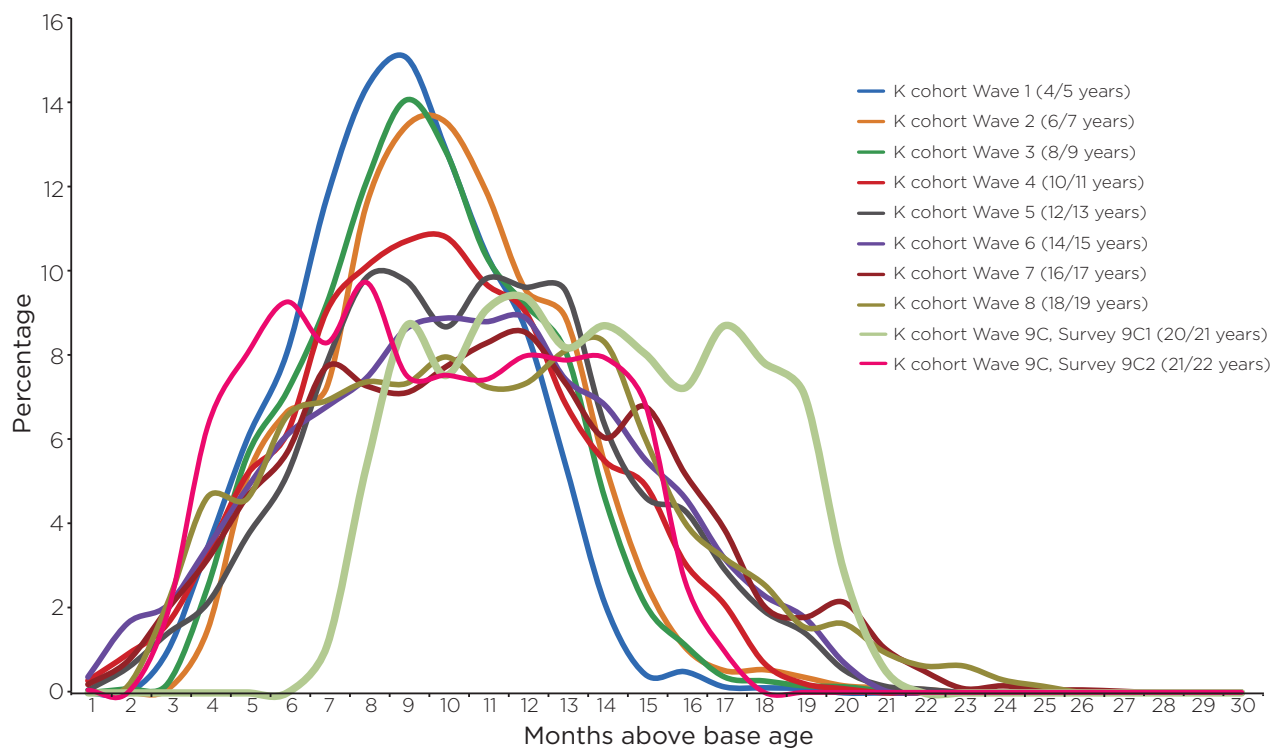
The child is the unit of selection in LSAC and estimates produced from this survey are of children, not of parents or families. It is important that this point is understood when producing population estimates from this survey.

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.

## 11.3 Age at interview

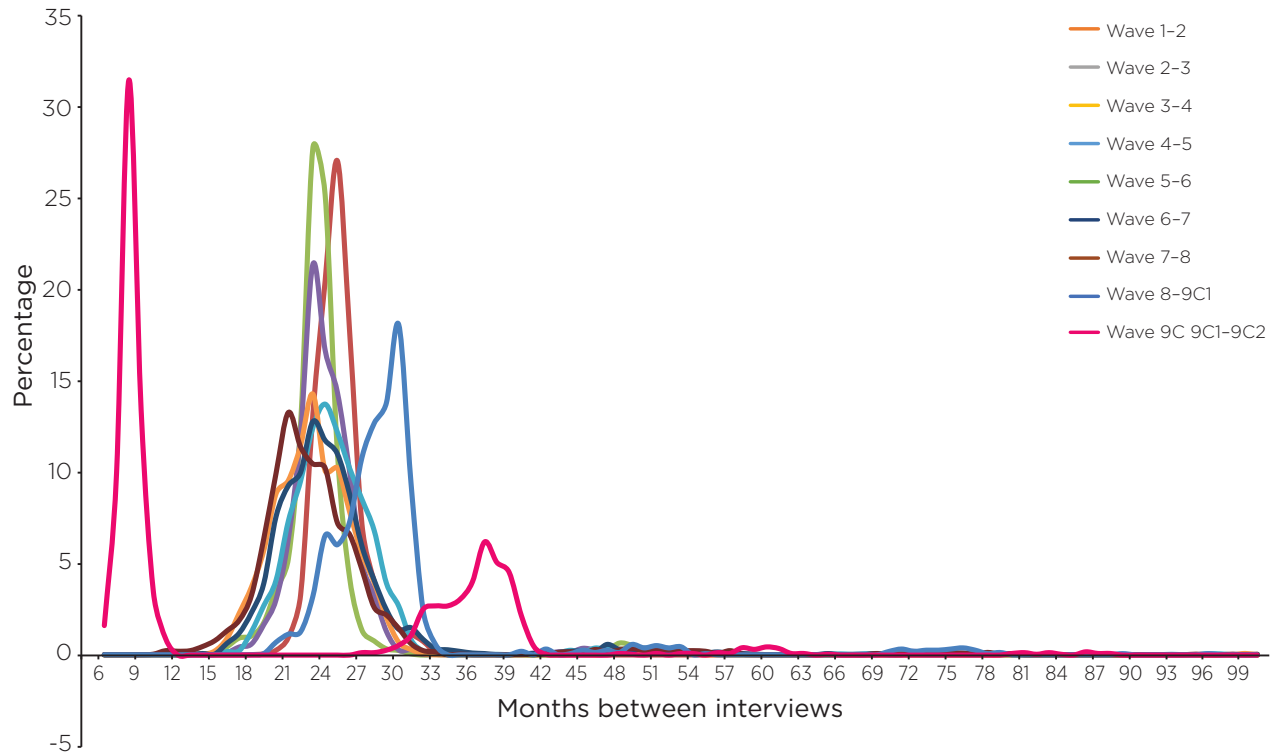
Different ages of children should be accounted for in any analyses focused on age-dependent measures such as cognitive and motor development. Figures 7 and 8 show the age distribution of the two cohorts at each wave. The figures show the age of the study child as a base figure (i.e. 0, 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20 years) plus a number of months. For example, a B cohort study 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

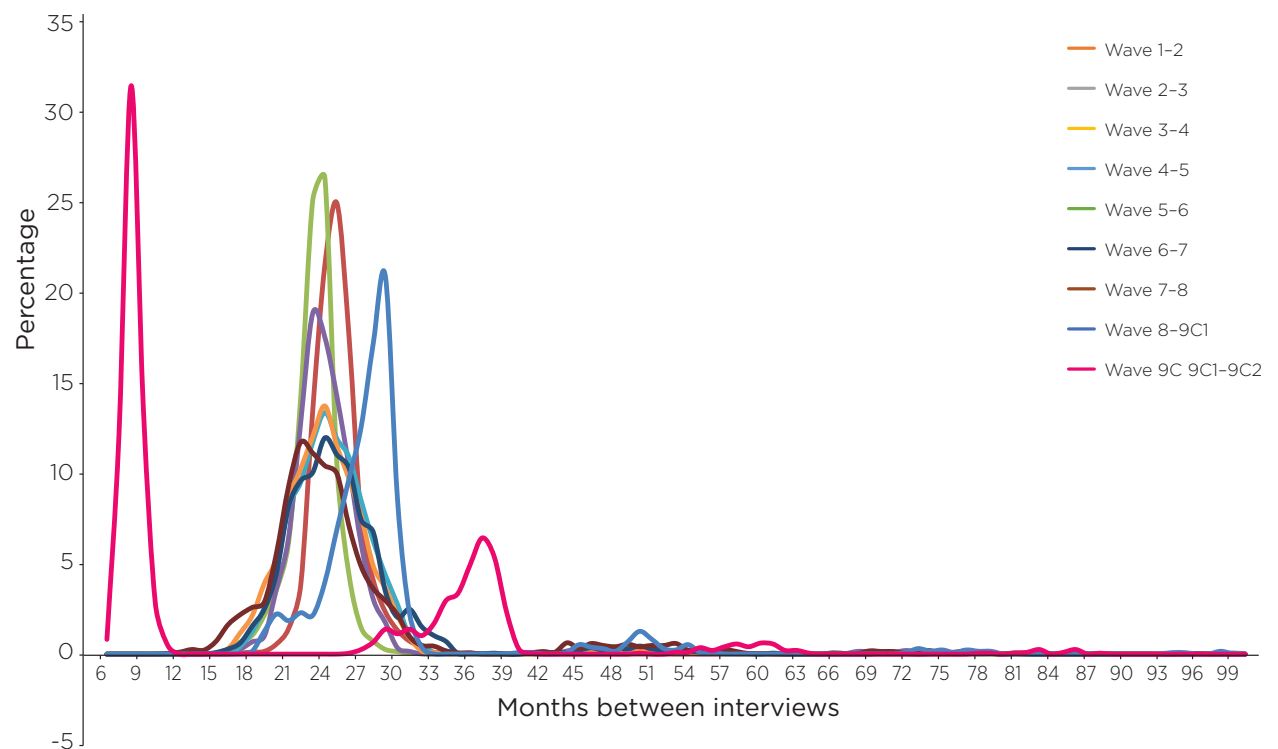
## 11.4 Time between interviews

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, Wave 1 to 9C



**Figure 10:** Distribution of time between interviews, K cohort, Wave 1 to 9C



## 11.5 Cross-cohort comparisons

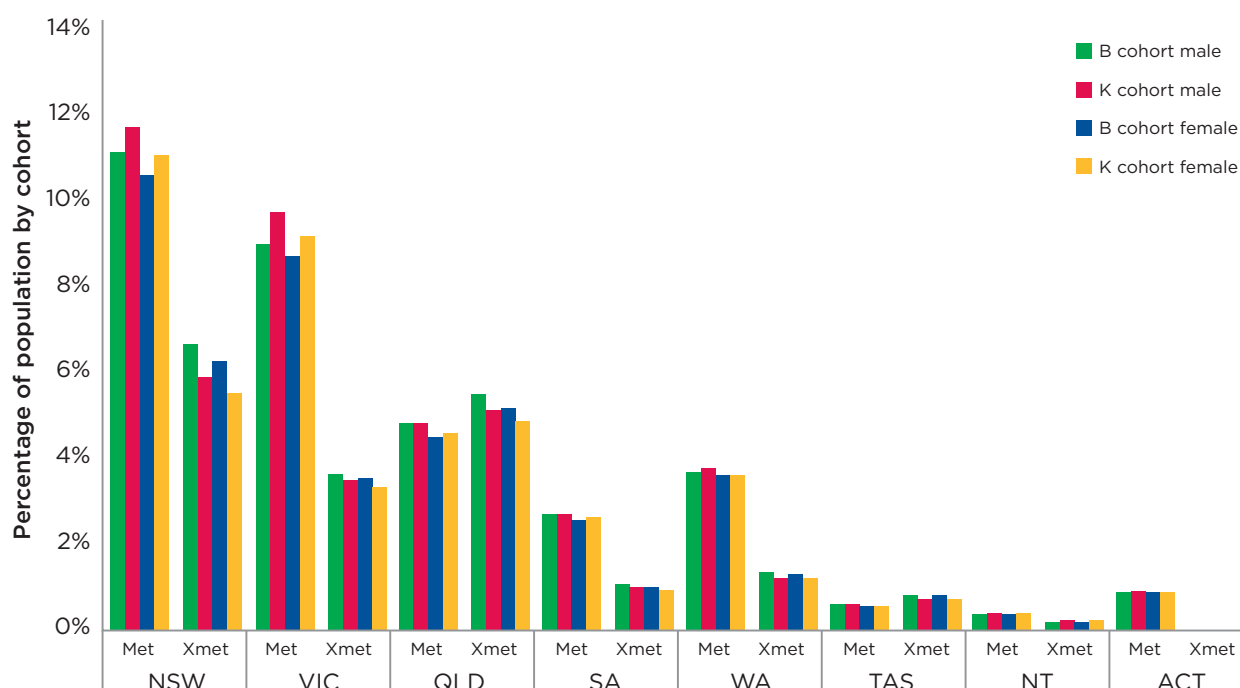
It should be noted that the two cohorts of LSAC were selected and weighted to represent similar but different populations. For the B cohort, the reference population is '0-year-old children in Australia in 2004 excluding those from certain remote postcodes', while for the K cohort the reference population 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.

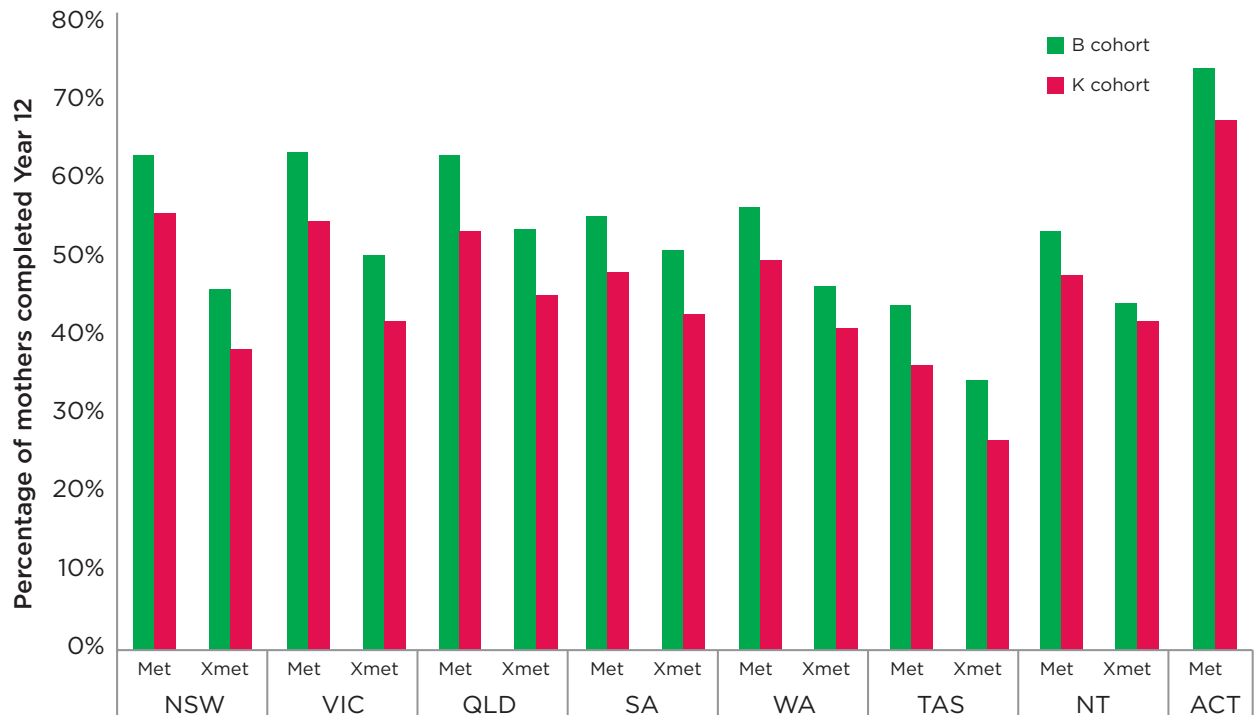
It is important to note that the Wave 7 B cohorts and Wave 9C K cohorts are comparable in terms of age range, but there are likely considerable differences in terms of interview period effects, modal effects and non-response. Wave 9C was collected during the COVID pandemic, when restrictions were in place, potentially introducing a period effect; data collection was all done online, potentially introducing modal effects; and the sample size was much smaller than usual, so patterns of non-response may differ from those previously established. As a result, data users need to consider the differential composition of the sample during analyses, interpretation and discussion of the results.

**Figure 11:** Cohort benchmarks by state, part of state and gender (Wave 1)

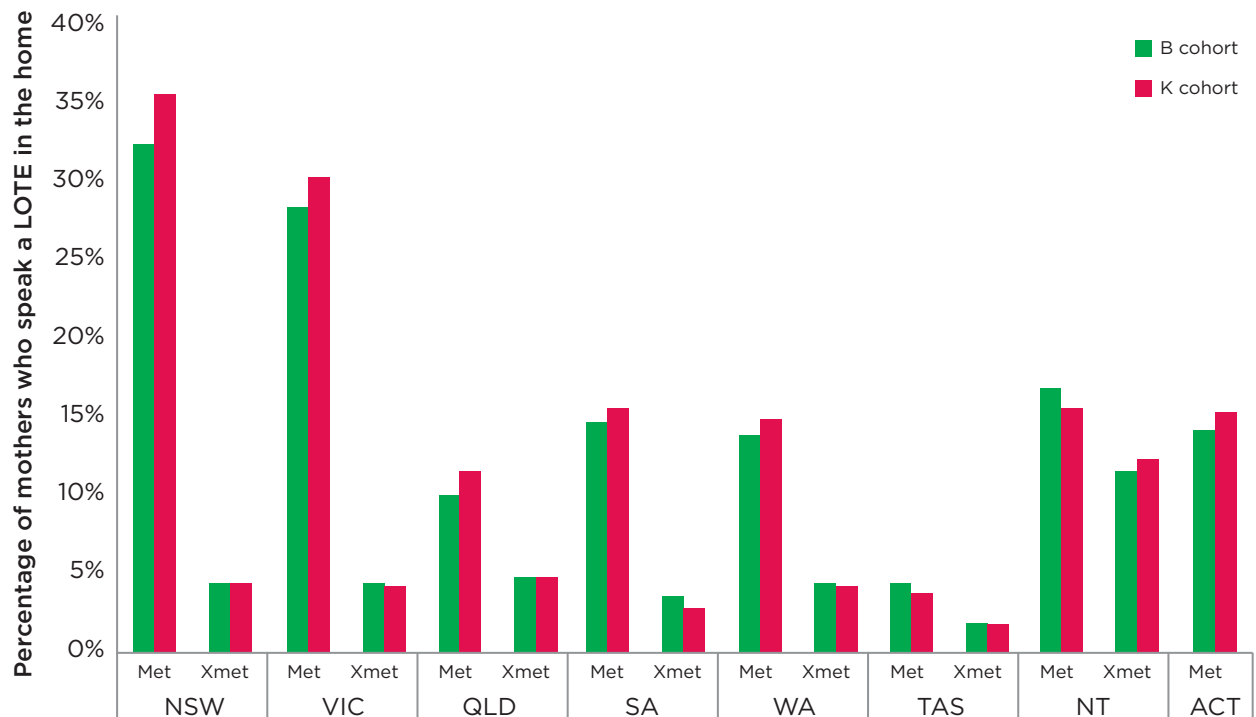


**Note:** There are no respondents from non-metropolitan ACT.



**Figure 12:** Proportion of mothers who completed Year 12, cohort benchmarks by state and part of state (Wave 1)

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 (Wave 1)

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

## 11.6 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 24).

**Table 24: Wave 1 sample characteristics compared with ABS data**

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

**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.

Tables 25 and 26 show the number of children in the Wave 1 sample with selected characteristics, and give the Waves 2 to 9C response rates for children with these characteristics. As can be seen in the tables, the greatest sample loss has been from Indigenous families, families where the mother speaks a language other than English at home and families where the mother did not complete Year 12.

**Table 25:** Response rates for B cohort at Wave 2 to 9C by selected sample characteristics

Characteristics	Wave 1 <i>n</i>	% responding at Wave 2	% responding at Wave 3	% responding at Wave 4	% responding at Wave 5	% responding at Wave 6	% responding at Wave 7	% responding at Wave 8	% responding at 9C1 <sup>a</sup>	% responding at 9C2 <sup>a</sup>
<b>Full sample</b>	<b>5,107</b>	<b>90.2</b>	<b>85.9</b>	<b>83.1</b>	<b>80.0</b>	<b>73.7</b>	<b>73.4</b>	<b>61.2</b>	<b>39.5</b>	<b>52.6</b>
Study child male	2,610	90.0	86.2	83.9	80.3	73.9	73.8	61.6	37.5	51.1
Study child female	2,497	90.3	85.5	82.2	79.6	73.4	73.0	60.9	41.6	54.2
Study child Indigenous	230	78.3	64.8	63.0	60.4	46.1	48.3	34.3	13.9	24.3
Mother speaks language other than English	740	83.9	75.0	72.0	68.6	61.1	61.4	49.1	30.4	40.5
Mother did not complete Year 12	1,688	84.8	78.8	74.4	70.1	61.4	62.5	46.6	25.7	37.6
Father did not complete Year 12	1,890	90.0	85.9	83.6	79.7	73.0	70.9	58.1	33.9	48.6
New South Wales	1,615	90.3	84.4	81.8	79.8	71.2	70.6	59.4	38.0	52.4
Victoria	1,251	88.4	85.1	81.9	76.6	71.5	71.5	58.9	41.9	52.0
Queensland	1,054	91.4	88.0	84.3	82.4	75.4	76.5	62.5	35.6	50.7
South Australia	347	91.1	88.2	85.9	81.0	76.1	73.1	59.4	41.2	52.4
Western Australia	533	89.7	83.9	81.6	78.6	75.0	75.3	63.2	40.0	50.5
Tasmania	113	90.3	92.0	92.9	91.2	88.5	88.2	76.1	44.2	63.7
Northern Territory	87	90.8	83.9	80.5	81.6	79.3	77.2	66.7	49.4	63.2
Australian Capital Territory	107	97.2	95.3	93.5	89.7	85.0	78.8	78.5	52.3	74.8
Capital city statistical division	3,194	90.6	86.2	82.9	79.9	74.1	73.2	61.7	41.6	54.4
Balance of state	1,913	89.5	85.4	83.3	80.1	73.0	73.8	60.4	36.0	49.7

**Notes:** <sup>a</sup> The methodology and content of 9C1 differs substantially from Waves 1–8. Therefore, data users are advised to cautiously interpret trends for key sample characteristics.

**Table 26:** Response rates for K cohort at Wave 2 to 9C by selected sample characteristics

Characteristics	Wave 1 <i>n</i>	% responding at Wave 2	% responding at Wave 3	% responding at Wave 4	% responding at Wave 5	% responding at Wave 6	% responding at Wave 7	% responding at Wave 8	% responding at 9C1 <sup>a</sup>	% responding at 9C2 <sup>a</sup>
<b>Full sample</b>	<b>4,983</b>	<b>89.6</b>	<b>86.9</b>	<b>83.7</b>	<b>79.4</b>	<b>71.0</b>	<b>71.0</b>	<b>60.9</b>	<b>35.9</b>	<b>49.4</b>
Study child male	2,537	89.8	87.2	84.1	79.7	70.9	70.9	61.2	32.5	46.8
Study child female	2,446	89.4	86.6	83.2	79.1	71.0	71.0	60.7	39.4	52.1
Study child Indigenous	187	81.8	66.3	63.1	60.4	44.4	44.4	31.0	13.4	19.3
Mother speaks language other than English	778	83.8	76.6	71.1	66.1	58.5	58.5	50.0	29.4	42.7
Mother did not complete Year 12	2,044	86.5	81.7	78.1	72.6	62.2	62.2	50.6	26.0	37.7
Father did not complete Year 12	2,016	90.1	87.0	84.9	80.9	71.6	71.6	60.2	31.9	46.2
New South Wales	1,573	90.1	86.3	81.8	78.2	70.2	70.2	61.4	35.9	50.8
Victoria	1,245	86.3	86.0	83.1	76.7	68.1	68.1	59.0	36.4	49.9
Queensland	988	90.8	87.2	84.0	80.9	71.9	71.9	59.5	31.7	43.9
South Australia	339	89.4	86.7	83.2	79.6	70.5	70.5	56.6	35.4	46.6
Western Australia	507	91.5	87.6	86.0	81.1	73.0	73.0	64.3	37.5	50.9
Tasmania	136	94.1	91.2	90.4	87.5	83.1	83.1	72.1	43.4	53.7
Northern Territory	82	89.0	87.8	89.0	86.6	72.0	72.0	57.3	36.6	51.2
Australian Capital Territory	113	94.7	94.7	92.0	89.4	82.3	82.3	76.1	52.2	69.0
Capital city statistical division	3,095	89.3	86.8	82.8	78.7	70.4	70.4	61.4	37.7	51.0
Balance of state	1,888	90.0	87.2	85.0	80.5	72.0	72.0	60.2	32.9	46.9

Notes: <sup>a</sup> The methodology and content of 9C1 differs substantially from Waves 1-8. Therefore, data users are advised to cautiously interpret trends for key sample characteristics.

# 12 User support and training

User training sessions are offered by AIFS to provide more detailed information than in the data user guide. This training will allow users to interact with the AIFS staff and benefit from their in-depth knowledge and experience with the LSAC data.

These sessions consist of an introduction to LSAC data, and any newly released datasets, including:

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

The LSAC website will provide further details as to when the training sessions are being offered.

## 12.1 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; and
- updates on data-user workshops.

## 12.2 Getting more information

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

- [LSAC website](#)
- Check the [Frequently Asked Questions](#)
- If you have questions about the data files or variables, please submit your query via [Dataverse](#). Click on the email icon or the 'Support' option on the Dataverse page.
- Send your queries directly to AIFS at: [aifs-lsac@aifs.gov.au](mailto:aifs-lsac@aifs.gov.au) and [datamanager@aifs.gov.au](mailto:datamanager@aifs.gov.au)

# References

- Australian Institute of Family Studies (AIFS). (2018). *Event History Calendar Wave 7: Business Rules*. (Final report in progress). Melbourne: AIFS.
- Australian Bureau of Statistics (ABS). (1996). *Women's safety Australia: User guide*. Canberra: ABS.
- Baker, K., Maguire, B., Daraganova, G., & Siphthorp, M. (2016). *Using My School Data in the Longitudinal Study of Australian Children* (LSAC Technical Paper No. 16). Melbourne: Australian Institute of Family Studies.
- Baxter, J. (2007). *Children's time use in the Longitudinal Study of Australian Children: Data quality and analytical issues in the 4-year-old cohort* (LSAC Technical Paper No. 4). Melbourne: Australian Institute of Family Studies.
- Chambers, R. L., & Skinner, C. J., (Eds.). (2003). *Analysis of survey data*. Chichester: Wiley.
- Cusack, B., & Defina, R. (2013). *Wave 5 weighting and non response* (LSAC Technical Paper No. 10). Canberra: Australian Bureau of Statistics.
- Daraganova, G., & Siphthorp, M. (2011). *Wave 4 weights* (LSAC Technical Paper No. 9). Melbourne: Australian Institute of Family Studies.
- Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376–382.
- Deville, J. C., Särndal C. E., & Sautory, O. (1993). Generalised raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013–1020.
- Kalton, G. (1983). *Compensating for missing survey data* (Research report series). Ann Arbor, MI: Institute for Social Research, University of Michigan.
- Lepkowski, J. M. (1989). Treatment of wave nonresponse in panel surveys. In D. Kasprzyk, G. Duncan, G. Kalton & M. P. Singh (Eds.), *Panel Surveys* (pp. 348–374). New York: Wiley.
- Levy, P. S., & Lemeshow, S. (1999). *Sampling of populations: Methods and applications* (3rd ed.). New York: Wiley.
- Misson, S., & Siphthorp, M. (2007). *Wave 2 weighting and non-response* (LSAC Technical Paper No. 5). Melbourne: Australian Institute of Family Studies.
- Murdoch Children's Research Institute, The Royal Children's Hospital (MCRI). (2017). *CheckPoint Health Data: Data User Guide*. Melbourne: MCRI.
- National Childcare Accreditation Council (NCAC). (2003). *OSHCQA quality practices guide* (1st ed.). Sydney: NCAC.
- NCAC. (2003). *QIAS validation report* (2nd ed.). Sydney: NCAC.
- NCAC. (2004). *FDCQA quality practices guide* (2nd ed.). Sydney: NCAC.
- NCAC. (2005). *FDCQA quality practices guide* (3rd ed.) Sydney: NCAC.
- NCAC. (2006). *QIAS quality practices guide* (1st ed.). Sydney: NCAC.
- Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society Series B*, 60, 23–40.
- Rice M. L., Hoffman L., & Wexler K. (2009). Judgments of omitted BE and DO in questions as extended finiteness clinical markers of specific language impairment (SLI) to 15 years: A study of growth and asymptote. *Journal of Speech, Language, and Hearing Research*, 52, 1417–1433.
- Rowe, K. (2006). *The measurement of composite variables from multiple indicators: Applications in quality assurance and accreditation systems—childcare*. Background paper prepared for the National Childcare Accreditation Centre, Sydney.
- Sanson, A., Misson, S., & The LSAC Outcome Index Working Group. (2006). *Summarising children's wellbeing: The LSAC Outcome Index* (LSAC Technical Paper No 2). Melbourne: Australian Institute of Family Studies.
- Sanson, A., Nicholson, J., Ungerer, J., Zubrick, S., Wilson, K., Ainley, J. et al. (2002). *Introducing the Longitudinal Study of Australian Children* (LSAC Discussion Paper No. 1). Melbourne: Australian Institute of Family Studies.
- Siphthorp, M., & Misson, S. (2009). *Wave 3 weighting and non-response* (LSAC Technical Paper No. 6). Melbourne: Australian Institute of Family Studies.
- Skinner, C. J., & Holmes, D. J. (2003). Random effects models for longitudinal survey data. In R. L. Chambers & C. J. Skinner (Eds.), *Analysis of Survey Data* (pp. 205–218). Chichester: Wiley.

- Soloff, C., Lawrence, D., & Johnstone, R. (2005). *Sample design* (LSAC Technical Paper No. 1). Melbourne: Australian Institute of Family Studies.
- Soloff, C., Lawrence, D., Misson, S., & Johnstone, R. (2005). *Wave 1 weighting and non-response* (LSAC Technical Paper No. 3). Melbourne: Australian Institute of Family Studies.
- Tabachnick, B. G., & Fidell, L. S. (1989). *Using Multivariate Statistics* (2nd ed.). New York: Harper and Row.
- Wolter, K. (1984). *Introduction to Variance Estimation*. New York: Springer.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
- Yu, M., & Daraganova, G. (2017). *Executive Functioning-Use of Cogstate measures in the Longitudinal Study of Australian Children* (LSAC Technical Paper No. 19). Melbourne: Australian Institute of Family Studies.



# Appendix A: Syntax to derive mother & father items

Syntaxes for calculating the mother and father items in SAS, STATA and SPSS applications. Identify the variable names for Parent 1 and Parent 2 from LSAC data dictionary and substitute in the syntax below.

zf02hp1 – Parent 1 Sex

zf02hp2 – Parent 2 Sex

P1/P2 variables – Variables for Parent 1 and Parent 2 (refer to the Data Dictionary for the variable name & person label)

## A. SAS Syntax

```

if zf02hp1=1 then do;
    mother=P2 variable;
    father=P1 variable;
end;
else if zf02hp1=2 then do;
    mother=P1 variable;
    father=P2 variable;
end;
else if zf02hp1=-9 then do;
    mother=-9;
    father=-9;
end;
else if zf02hp1 = zf02hp2 then do;
    mother=P1 variable;
    father=P2 variable;
end;

```

## B. STATA Syntax

```

generate mother = P2 variable if zf02hp1==1
replace mother = P1 variable if zf02hp1==2
replace mother = -9 if zf02hp1== -9
replace mother = P1 variable if zf02hp1== zf02hp2

generate father = P1 variable if zf02hp1==1
replace father = P2 variable if zf02hp1==2
replace father = -9 if zf02hp1== -9
replace father = P2 variable if zf02hp1== zf02hp2

```

## C. SPSS Syntax

```
compute Mother= 0.  
if (zf02hp1=1) mother =P2 variable.  
if (zf02hp1=2) mother =P1 variable.  
if (zf02hp1 = zf02hp2) mother =P1 variable.  
if (zf02hp1=-9) mother=-9.  
Compute father = 0.  
if (zf02hp1=1) father =P1 variable.  
if (zf02hp1=2) father =P2 variable.  
if (zf02hp1 = zf02hp2) father =P2 variable.  
if (zf02hp1=-9) father =-9.
```

# Appendix B: Sample code

## Example: Derived items from Medicare dataset

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 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

```
proc means data=m.pbs nway sum;
  class hcid;
  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 hcid;

  if medicare=1 and ben07=. then ben07=0;
run;
```

### SPSS

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

## STATA

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

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

## Example: Sample analysis of time use diary

### 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.

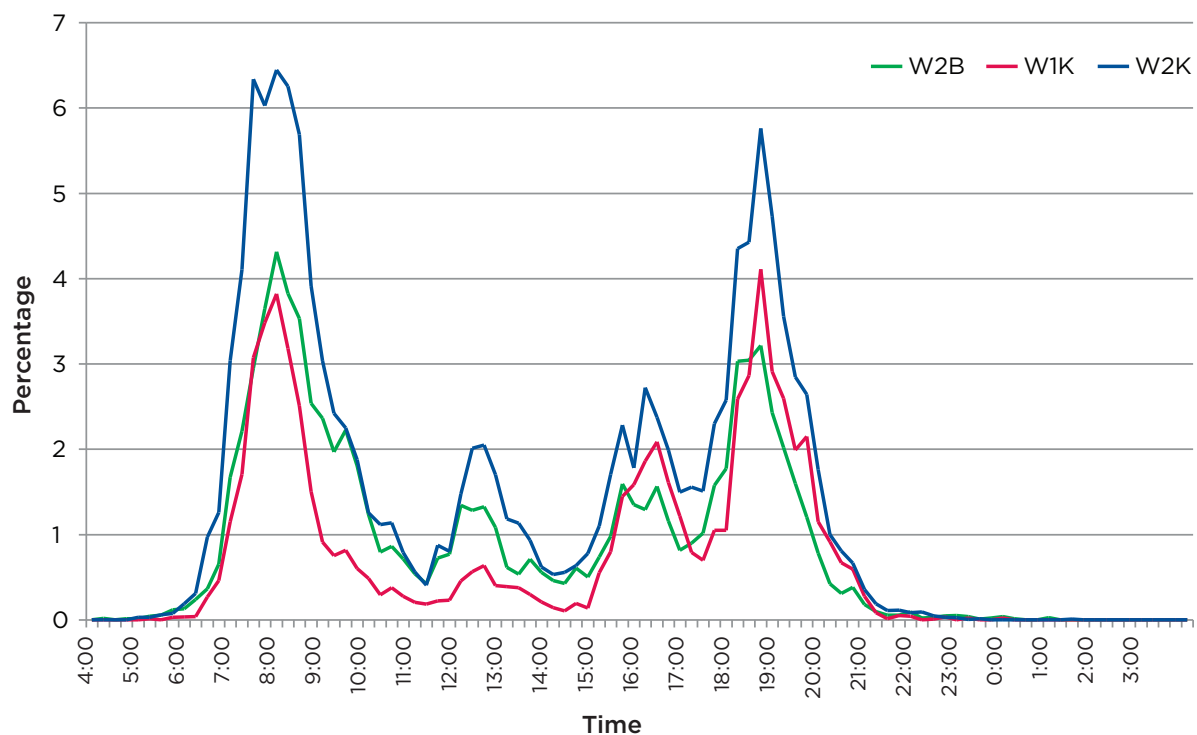
Statements 10–13 produce the means of the variables in the 'Tveat' array (which SAS gives the names Tveat1 to Tveat96 by default). The mean here will be the percentage of children from whom an activity was known that ate or drank in front of the TV, etc., at each time period. Line 12 uses the day weight 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 14 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 14:** 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:

```
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:

```
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)
}
```

## Example: Template for using SURVEYREG and SURVEYMEANS in SAS

The following code 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

$\text{stratum} = \text{int}(\text{mod}(\text{hiciid}, 10000000) / 100000);$

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.

```
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 4  
74 1  
81 23  
; run;
```

# Appendix C: LSAC variable naming conventions

## Standard input variables – attnsxx

a	tt	nn	s	xx
Child age	Topic	Arbitrary number within topic	Subject/informant (optional)	Sub-numbering
<b>Values</b>				
a = 0–1 years b = 2–3 years c = 4–5 years d = 6–7 years e = 8–9 years f = 10–11 years g = 12–13 years h = 14–15 years i = 16–17 years i1 = 16–17 years i2 = 17–18 years j = 18–19 years k1 = 20–21 years k2 = 21–22 years z = does not change with age of child	(examples) fd = family demographics hs = health status se = social and emotional outcomes	(examples) 01 02 03 04	a = parent 1 b = parent 2 c = study child p = PLE m = mother f = father t = teacher/carer i = between waves respondent	As required for grouping of like items. See examples below.
<b>OR</b>				
			Education/childcare items – see Data User Guide Section 5.1.3 for values	

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

a	s	m
Child age	(optional) Subject/informant	Up to 6-character mnemonic where possible
As above	As above	e.g. vocab = MCDI vocabulary measure score

Example:

- aaemp = P1 employment status when child aged 0–1 years
- bbemp = P2 employment status when child aged 2–3 years



## Household composition variables – aFnnxmmm

a	F	nn	x	mmm
Child age	Same letter in all variables ('f' for family)	Arbitrary number within topic	Sub-question indicator (optional)	Person
<b>Values</b>				
a = 0–1 years b = 2–3 years c = 4–5 years d = 6–7 years (e = 8–9 years) etc.  z = does not change with age of child	f	E.g. 01 02 03 04 etc.	E.g. a b c d etc.	<b>mnn</b> m1 = study child m2 = W1 P1 m3 = W1 P2 m4-15 = other hh members t1-6 = temporary hh members

OR

	<b>cpp</b> c = child's age
	<b>and pp is</b> m = mother f = father p1 = P1 p2 = P2

Examples:

- 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)

## Parent identifier variables – apMN

These take values of 1-15 or missing.

a	p	MN
Child age	Parent	Same letters in all variables
As above	m = mother f = father p1 = P1 p2 = P2	mn

Examples:

- ammn = member number (mn) for mother (m) when child aged 0–1 years (a)
- bp1mn = member number for P1 when child aged 2–3 years

# Appendix D: Weighting variables

## B cohort

Variable name	Cohort	Type	Waves or surveys cases responded to	Used for
aweight	B	Population	1	Wave 1 cross-sectional analyses
aweights	B	Sample	1	Wave 1 cross-sectional analyses
aweighted	B	Day	1	Wave 1 cross-sectional analyses
bweight	B	Population	1 & 2	Wave 2 cross-sectional analyses and longitudinal analyses involving Waves 1 & 2
bweights	B	Sample	1 & 2	Wave 2 cross-sectional analyses and longitudinal analyses involving Waves 1 & 2
bweighted	B	Day	1 & 2	Wave 2 cross-sectional analyses and longitudinal analyses involving Waves 1 & 2
cweight	B	Population	1 & 3	Wave 3 cross-sectional analyses and longitudinal analyses involving Waves 1 & 3
cweights	B	Sample	1 & 3	Wave 3 cross-sectional analyses and longitudinal analyses involving Waves 1 & 3
cweighted	B	Day	1 & 3	Wave 3 cross-sectional analyses and longitudinal analyses involving Waves 1 & 3
bcwt	B	Population	1, 2 & 3	Longitudinal analyses involving all waves up to Wave 3
bcwts	B	Sample	1, 2 & 3	Longitudinal analyses involving all waves up to Wave 3
bcwtd	B	Day	1, 2 & 3	Longitudinal analyses involving all waves up to Wave 3
dweight	B	Population	1 & 4	Wave 4 cross-sectional analyses and longitudinal analyses involving Waves 1 & 4
dweights	B	Sample	1 & 4	Wave 4 cross-sectional analyses and longitudinal analyses involving Waves 1 & 4
eweight	B	Population	1 & 5	Wave 5 cross-sectional analyses and longitudinal analyses involving Waves 1 & 5
eweights	B	Sample	1 & 5	Wave 5 cross-sectional analyses and longitudinal analyses involving Waves 1 & 5
bdwt	B	Population	1, 2 & 4	Longitudinal analyses involving Waves 2 & 4, or Waves 1, 2 & 4
bdwts	B	Sample	1, 2 & 4	Longitudinal analyses involving Waves 2 & 4, or Waves 1, 2 & 4
cdwt	B	Population	1, 3 & 4	Longitudinal analyses involving Waves 3 & 4, or Waves 1, 3 & 4
cdwts	B	Sample	1, 3 & 4	Longitudinal analyses involving Waves 3 & 4, or Waves 1, 3 & 4
bcdwt	B	Population	1, 2, 3 & 4	Longitudinal analyses involving all Waves up to Wave 4
bcdwts	B	Sample	1, 2, 3 & 4	Longitudinal analyses involving all waves up to Wave 4
bcdewt	B	Population	1, 2, 3, 4 & 5	Longitudinal analyses involving all waves up to Wave 5

Table continued on next page →

Variable name	Cohort	Type	Waves or surveys cases responded to	Used for
bcdewts	B	Sample	1, 2, 3, 4 & 5	Longitudinal analyses involving all waves up to Wave 5
fweight	B	Population	1 & 6	Wave 6 cross-sectional analyses and longitudinal analyses involving Waves 1 & 6
fweights	B	Sample	1 & 6	Wave 6 cross-sectional analyses and longitudinal analyses involving Waves 1 & 6
bcdefwt	B	Population	1, 2, 3, 4, 5 & 6	Longitudinal analyses involving all waves up to Wave 6
bcdefwts	B	Sample	1, 2, 3, 4, 5 & 6	Longitudinal analyses involving all waves up to Wave 6
gweight	B	Population	1 & 7	Wave 7 cross-sectional analyses and longitudinal analyses involving Waves 1 & 7
gweights	B	Sample	1 & 7	Wave 7 cross-sectional analyses and longitudinal analyses involving Waves 1 & 7
bcdefgwt	B	Population	1, 2, 3, 4, 5, 6 & 7	Longitudinal analyses involving all waves up to Wave 7
bcdefgwts	B	Sample	1, 2, 3, 4, 5, 6 & 7	Longitudinal analyses involving all waves up to Wave 7
hweight	B	Population	1 & 8	Wave 8 cross-sectional analyses and longitudinal analyses involving Waves 1 & 8
hweights	B	Sample	1 & 8	Wave 8 cross-sectional analyses and longitudinal analyses involving Waves 1 & 8
bcdefghwt	B	Population	1, 2, 3, 4, 5, 6, 7 & 8	Longitudinal analyses involving all waves up to Wave 8
bcdefghwts	B	Sample	1, 2, 3, 4, 5, 6, 7 & 8	Longitudinal analyses involving all waves up to Wave 8
i1weight	B	Population	1 & 9C1	9C1 cross-sectional analyses and longitudinal analyses involving Wave 1 & 9C1
i1weights	B	Sample	1 & 9C1	9C1 cross-sectional analyses and longitudinal analyses involving Wave 1 & 9C1
bcdefghi1wt	B	Population	1, 2, 3, 4, 5, 6, 7, 8 & 9C1	Longitudinal analyses involving all waves up to 9C1
bcdefghi1wts	B	Sample	1, 2, 3, 4, 5, 6, 7, 8 & 9C1	Longitudinal analyses involving all waves up to 9C1
i2weight	B	Population	1 & 9C2	9C2 cross-sectional analyses and longitudinal analyses involving Waves 1 & 9C2
i2weights	B	Sample	1 & 9C2	9C2 cross-sectional analyses and longitudinal analyses involving Waves 1 & 9C2
bcdefghi1i2wt	B	Population	1, 2, 3, 4, 5, 6, 7, 8, 9C1 & 9C2	Longitudinal analyses involving all waves up to 9C2
bcdefghi1i2wts	B	Sample	1, 2, 3, 4, 5, 6, 7, 8, 9C1 & 9C2	Longitudinal analyses involving all waves up to 9C2

## K cohort

Variable name	Cohort	Type	Waves or surveys cases responded to	Used for
cweight	K	Population	1	Wave 1 cross-sectional analyses
cweights	K	Sample	1	Wave 1 cross-sectional analyses
cweightd	K	Day	1	Wave 1 cross-sectional analyses
dweight	K	Population	1 & 2	Wave 2 cross-sectional analyses and longitudinal analyses involving Waves 1 & 2
dweights	K	Sample	1 & 2	Wave 2 cross-sectional analyses and longitudinal analyses involving Waves 1 & 2
dweightd	K	Day	1 & 2	Wave 2 cross-sectional analyses and longitudinal analyses involving Waves 1 & 2
eweight	K	Population	1 & 3	Wave 3 cross-sectional analyses and longitudinal analyses involving Waves 1 & 3

Table continued on next page →

Variable name	Cohort	Type	Waves or surveys cases responded to	Used for
eweights	K	Sample	1 & 3	Wave 3 cross-sectional analyses and longitudinal analyses involving Waves 1 & 3
eweightd	K	Day	1 & 3	Wave 3 cross-sectional analyses and longitudinal analyses involving Waves 1 & 3
dewt	K	Population	1, 2 & 3	Longitudinal analyses involving all waves up to Wave 3
dewts	K	Sample	1, 2 & 3	Longitudinal analyses involving all waves up to Wave 3
dewtd	K	Day	1, 2 & 3	Longitudinal analyses involving all waves up to Wave 3
fweight	K	Population	1 & 4	Wave 4 cross-sectional analyses and longitudinal analyses involving Waves 1 & 4
fweights	K	Sample	1 & 4	Wave 4 cross-sectional analyses and longitudinal analyses involving Waves 1 & 4
dfwt	K	Population	1, 2 & 4	Longitudinal analyses involving Waves 2 & 4, or Waves 1, 2 & 4
dfwts	K	Sample	1, 2 & 4	Longitudinal analyses involving Waves 2 & 4, or Waves 1, 2 & 4
efwt	K	Population	1, 3 & 4	Longitudinal analyses involving Waves 3 & 4, or Waves 1, 3 & 4
efwts	K	Sample	1, 3 & 4	Longitudinal analyses involving Waves 3 & 4, or Waves 1, 3 & 4
defwt	K	Population	1, 2, 3 & 4	Longitudinal analyses involving all waves up to Wave 4
defwts	K	Sample	1, 2, 3 & 4	Longitudinal analyses involving all waves up to Wave 4
gweight	K	Population	1 & 5	Wave 5 cross-sectional analyses and longitudinal analyses involving Waves 1 & 5
gweights	K	Sample	1 & 5	Wave 5 cross-sectional analyses and longitudinal analyses involving Waves 1 & 5
defgwt	K	Population	1, 2, 3, 4 & 5	Longitudinal analyses involving all waves up to Wave 5
defgwts	K	Sample	1, 2, 3, 4 & 5	Longitudinal analyses involving all waves up to Wave 5
hweight	K	Population	1 & 6	Wave 6 cross-sectional analyses and longitudinal analyses involving Waves 1 & 6
hweights	K	Sample	1 & 6	Wave 6 cross-sectional analyses and longitudinal analyses involving Waves 1 & 6
defghwt	K	Population	1, 2, 3, 4, 5 & 6	Longitudinal analyses involving all waves up to Wave 6
defghwts	K	Sample	1, 2, 3, 4, 5 & 6	Longitudinal analyses involving all waves up to Wave 6
iweight	K	Population	1 & 7	Wave 7 cross-sectional analyses and longitudinal analyses involving Waves 1 & 7
iweights	K	Sample	1 & 7	Wave 7 cross-sectional analyses and longitudinal analyses involving Waves 1 & 7
defghiwt	K	Population	1, 2, 3, 4, 5, 6 & 7	Longitudinal analyses involving all waves up to Wave 7
defghiwts	K	Sample	1, 2, 3, 4, 5, 6 & 7	Longitudinal analyses involving all waves up to Wave 7
jweight	K	Population	1 & 8	Wave 8 cross-sectional analyses and longitudinal analyses involving Waves 1 & 8
jweights	K	Sample	1 & 8	Wave 8 cross-sectional analyses and longitudinal analyses involving Waves 1 & 8
defghijwt	K	Population	1, 2, 3, 4, 5, 6, 7 & 8	Longitudinal analyses involving all waves up to Wave 8
defghijwts	K	Sample	1, 2, 3, 4, 5, 6, 7 & 8	Longitudinal analyses involving all waves up to Wave 8
k1weight	K	Population	1 & 9C1	9C1 cross-sectional analyses and longitudinal analyses involving Wave 1 & 9C1
k1weights	K	Sample	1 & 9C1	9C1 cross-sectional analyses and longitudinal analyses involving Wave 1 & 9C1
defghijk1wt	K	Population	1, 2, 3, 4, 5, 6, 7, 8 & 9C1	Longitudinal analyses involving all waves up to 9C1

Table continued on next page →

Variable name	Cohort	Type	Waves or surveys cases responded to	Used for
defghijk1wts	K	Sample	1, 2, 3, 4, 5, 6, 7, 8 & 9C1	Longitudinal analyses involving all waves up to 9C1
k2weight	K	Population	1 & 9C2	9C2 cross-sectional analyses and longitudinal analyses involving Waves 1 & 9C2
k2weights	K	Sample	1 & 9C2	9C2 cross-sectional analyses and longitudinal analyses involving Waves 1 & 9C2
defghijk1k2wt	K	Population	1, 2, 3, 4, 5, 6, 7, 8, 9C1 & 9C2	Longitudinal analyses involving all waves up to 9C2
defghijk1k2wts	K	Sample	1, 2, 3, 4, 5, 6, 7, 8, 9C1 & 9C2	Longitudinal analyses involving all waves up to 9C2