



The Longitudinal Study of Indigenous Children

Footprints in Time



1800 106 235
www.fahcsia.gov.au/lsc



The Longitudinal Study of Indigenous Children: An Australian Government initiative



Data User Guide

Release 2.0 August 2011

Contents

Acknowledgements.....	3
Introduction.....	4
What is Footprints in Time?.....	5
Objective of the study	5
Who is involved?	6
Funding	6
Ethics	6
Survey Methodology.....	7
<i>Footprints in Time</i> sample selection	7
<i>Footprints in Time</i> study sites	8
Study development and testing	10
Study informants	11
Survey Instruments	12
Data collection.....	12
Consent process	12
Fieldwork periods	13
Table 2: Fieldwork periods	13
Fieldwork Response.....	13
New entrants	14
Interview length	14
Study Child K Cohort	14
Questionnaire content overview	15
Using the datasets.....	18
Locating variables	19
Variable Naming Convention	19
Identifiers	20
Cohort	20
Geographic variables.....	20
Household Form.....	20
Direct Assessments of child development	21
Qualitative data	22
Other - specify responses	23
Missing data coding	23
Merging datasets	24
Confidentialisation.....	24
Data access	25
Data linkage	25
Australian Early Development Index (AEDI).....	25
Getting more information	26
Appendix A	27
Data Dictionary	27
Appendix B.....	28
Qualitative free text questions\variable label	28
Appendix C.....	30
Examples of merging in stata.....	30

Abbreviations

ABS – Australian Bureau of Statistics

AEDI – Australian Early Development Index

AIATSIS – Australian Institute of Aboriginal and Torres Strait Islander Studies

AIHW – Australian Institute of Health and Welfare

DEC – Departmental Ethics Committee

FaHCSIA – Department of Families, Housing, Community Services and Indigenous Affairs

HILDA – Household, Income and Labour Dynamics in Australia

HREC – Human Research Ethics Committee

ICC – Indigenous Coordination Centre

LORI – Levels of Relative Isolation

LSAC – Longitudinal Study of Australian Children

LSIC – Longitudinal Study of Indigenous Children

NACOH – National Advisory Committee on Oral Health

NHMRC - National Health and Medical Research Council

P 1 – Parent 1

P2 – Parent 2

PEDS – Parent’s Evaluation of Developmental Status

RAO - Research Administration Officer

SCRGSP – Steering Committee for the Review of Government Service Provision

Acknowledgements

Footprints in Time—the Longitudinal Study of Indigenous Children (LSIC) was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA).

The study would never have been possible without the support and trust of the Aboriginal and Torres Strait Islander families who opened their doors to the researchers and generously gave their time to talk openly about their lives. Our gratitude goes to them, and to the leaders and Elders of their communities who are active guardians of their people’s wellbeing.

The study brings together people committed to making a positive difference in the lives of Aboriginal and Torres Strait Islander children.

Introduction

The purpose of this document is to provide a comprehensive reference for data users of *Footprints in Time*, the Longitudinal Study of Indigenous Children (LSIC). This document aims to provide what the data users need to know to use the LSIC datasets – such as the background to the Study, sample selection, study sites, research design, study development and testing, consent processes, questionnaire design and piloting, file structures, variable naming conventions and missing data coding.

Other useful documentation for data users are the mark up questionnaires and the Data Dictionary.

The Data User Guide, mark up questionnaires and Data Dictionary are available on the Study website, <<http://www.fahcsia.gov.au/lxic>>

We welcome any feedback you have about this Data User Guide. If there is something that you expected to find in this manual and didn't, or if you had difficulty understanding any section, please let us know by emailing lsicdata@fahcsia.gov.au .

What is Footprints in Time?

Footprints in Time is the name given to the Longitudinal Study of Indigenous Children (LSIC). *Footprints in Time* aims to improve the understanding of, and policy response to, the diverse circumstances faced by Aboriginal and Torres Strait Islander children, their families and communities. The study provides a data resource that can be drawn on by all levels of government, researchers, service providers, parents and communities.

The study collects important information about the lives of Aboriginal and Torres Strait Islander children, covering areas including:

- **the children**—their physical and mental health, how they develop socially and cognitively, their place in their family and community, and significant events in their life
- **the children’s families**—their health, work, lifestyle, and family and community connectedness
- **the children’s communities**—facilities, services, and social and community issues
- **services**—child care, education, health and other services used by the child’s family.

Objective of the study

The main objective of the study is to provide high quality data that can be used to provide a better insight into how a child’s early years affect their development. It is hoped that this information can be drawn upon to help close the gap in life circumstances between Indigenous and non-Indigenous Australians.

Footprints in Time has four key research questions, formulated under the guidance of the Steering Committee, which were designed to achieve this objective. These are:

- *What do Aboriginal and Torres Strait Islander children need to have the best start in life to grow up strong?*
- *What helps Aboriginal and Torres Strait Islander children to stay on track or get them to become healthier, more positive and strong?*
- *How are Aboriginal and Torres Strait Islander children raised?*
- *What is the importance of family, extended family and community in the early years of life and when growing up?*

Also of interest is the role that service use and support plays in the lives of Aboriginal and Torres Strait Islander children:

- *How can services and other types of support make a difference to the lives of Aboriginal and Torres Strait Islander children?*

Who is involved?

Footprints in Time is funded by the Australian Government and managed by FaHCSIA.

The LSIC Steering Committee has been chaired by Professor Mick Dodson since 2003. The Steering Committee oversees the design, development and implementation of the study. Its members are drawn from academic and community backgrounds, covering a wide range of disciplines such as health, early learning and child care. Subcommittees of the Steering Committee are formed to deal with particular issues as required.

Strategic guidance and leadership on future content, operation and analysis of *Footprints in Time* is also provided by the Longitudinal Studies Advisory Group (LSAG). The primary objective of the LSAG is to provide advice to the Longitudinal Study for Australian Children (LSAC), *Footprint in Time* (LSIC) and the Household, Income and Labour Dynamics in Australia (HILDA) studies and thereby assist in maximising their strategic importance to the Australian Government.

The Wave 1 and Wave 2 interviews were conducted primarily by FaHCSIA employed Aboriginal and Torres Strait Islander Research Administration Officers (RAOs). Roy Morgan Research has been contracted to produce the data collection instruments according to FaHCSIA design, assist in the management of pilot and live fieldwork, capture and compile survey data and report on fieldwork procedures, as well as response and non-response patterns.

FaHCSIA's LSIC section manages the entire project from the Canberra National Office.

Funding

The 2003–04 Federal Budget provided the initial resources for the *Footprints in Time* study. The first phase, from September 2003 to June 2004, involved extensive consultation with Indigenous peoples and communities about the study. The design and development of the study commenced in December 2005, with pilot testing continuing through 2006 and 2007.

LSIC received funding in the 2007-08 budgets for Waves 1-4 of the study. The study is classified as an ongoing measure and will continue as long as the sample retention enables the study to remain viable.

Ethics

Ethical clearance for the study has been obtained from the Australian Government Department of Health and Ageing Departmental Ethics Committee (DEC) which has been chosen as the primary Human Research Ethics Committee (HREC) for the study. In addition state/territory and/or regional ethics clearance and support was obtained for *Footprints in Time* sites through state and territory HRECs or their equivalents (in accordance with the National Health and Medical Research Council (NHMRC), 2003 and Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS)

guidelines). State and Territory departments of education and Catholic dioceses are also consulted to gain permission and support for pre-school and school teachers to complete questionnaires about the children involved in the study. State and territory departments managing out-of-home care were also consulted.

Survey Methodology

Footprints in Time employs an accelerated cross-sequential design, involving two cohorts of Indigenous children aged from 6 months to 2 years (Baby cohort) and from 3 years, 6 months to 5 years (Child cohort) in Wave 1. The design allows the data covering the first 9 or 10 years of Aboriginal and Torres Strait Islander children's lives to be collected in six years. The two-cohort design also facilitates the comparison of the cohorts when their ages overlap in order to detect any changes due to different social conditions and policy initiatives.

Aboriginal and Torres Strait Islander children are the sample units in the study. The majority of families in the study were recruited by the RAOs using lists of addresses provided by Centrelink and Medicare Australia of families who had at least one child who was identified by the agencies as being Aboriginal or a Torres Strait Islander and born between December 2003 and November 2004 (Child cohort) or between December 2006 and November 2007 (Baby cohort). Other informal means of contact such as word of mouth, local knowledge and study promotion were also used to supplement the number of children in the study. In practice, the Child cohort consists of children born in 2003, 2004 and 2005 and the Baby cohort consists of children born in 2006, 2007 and 2008.

Table 1: Ages of each cohort through the study.

	'04	'05	'06	'07	2008 W 1	2009 W 2	2010 W 3	2011 W 4	2012 W 5	2013 W 6	2014 W 7	2015 W 8	2016 Wave 9
Ages B cohort				B o r n	6-24 mths	1½-3	2½-4	3½-5	4½-6	5½-7	6½-8	7½-10	8½-11
Ages K cohort	B o r n				3½-5	4½-6	5½-7	6½-8	7½-9	8½-10	9½-11	10½-12	11½-13

Footprints in Time sample selection

Footprints in Time uses a non-random purposive sampling design from which eligible families were approached and voluntary consent obtained. The study focuses on eleven sites chosen, in part, to cover the range of socioeconomic and community environments where Aboriginal and Torres Strait Islander children live. Agreement and approval to participate in the study was sought from communities and Elders in these sites before research within the communities began.

The *Footprints in Time* sites were chosen to:

- ensure approximately equal representation of urban, regional and remote areas, thus enabling some geographical comparison
- represent the concentration of Aboriginal and Torres Strait Islander people around Australia
- contain a substantial Aboriginal and Torres Strait Islander population in the core and surrounding areas
- include locations engaged in the pilot of the study where existing relationships could be built upon
- be located near an Indigenous Coordination Centre (ICC), if possible, where Research Administration Officers (RAOs) could be based.

Footprints in Time was designed to select approximately 150 children in each of the each site, providing a sample of up to 1,650 children. This number represents 5-10% of Aboriginal and Torres Strait Islander children of the appropriate ages. As such, most interviewers were required to find a large proportion of the total number of children in each birth cohort and these children were spread across a wide geographical area, usually within a predominantly non-Indigenous population. An exception to this were remote areas where the proportion of Aboriginal and Torres Strait Islander people was high but the population numbers, particularly within the targeted age ranges, were low. Due to difficulties in sample recruitment related to small resident populations and geographic spread, for some sites it was not possible to find sufficient numbers of children to meet the study's targets. In other sites the number of eligible children was in excess of the required sample.

The same families who were interviewed in Wave 1 were approached again for an interview in Wave 2. However, a proportion of families could not be interviewed again because they could not be located, had moved substantial distances, refused interviews, or could not be interviewed for other reasons. However, the reduction in the number of study children was partially offset by recruiting 88 additional children from the eleven sites into the sample. These children were from families who had either missed out on or refused to participate in Wave 1 but were available and willing to participate in Wave 2.

***Footprints in Time* study sites**

The LSIC sample is not nationally representative; it reflects fairly well the distribution of Aboriginal and Torres Strait Islander children aged between 0 and 5 years in the states and territories and among urban, regional and remote areas. Following are the selected study sites:

New South Wales (NSW)

- Western Sydney (from Campbelltown to Riverston)
- NSW South Coast (from Kiama to Eden)
- Dubbo (including Gilgandra, Wellington and Narromine)

Victoria (Vic)

- Greater Shepparton (including Wangaratta, Seymour, Bendigo, Cobram and Barmah and areas in between)

Queensland (Qld)

- South East Queensland (including Brisbane, Ipswich, Logan, Inala, Gold Coast and Bundaberg)
- Mount Isa and remote Western Queensland (including Mornington Island, Doomadgee, Normanton and Cloncurry)
- Torres Strait Islands and Northern Peninsula Area (NPA)

Western Australia (WA)

- Kimberley region (including Derby, Fitzroy Crossing, Broome and One Arm Point)

South Australia (SA)

- Adelaide (including Port Augusta)

Northern Territory (NT)

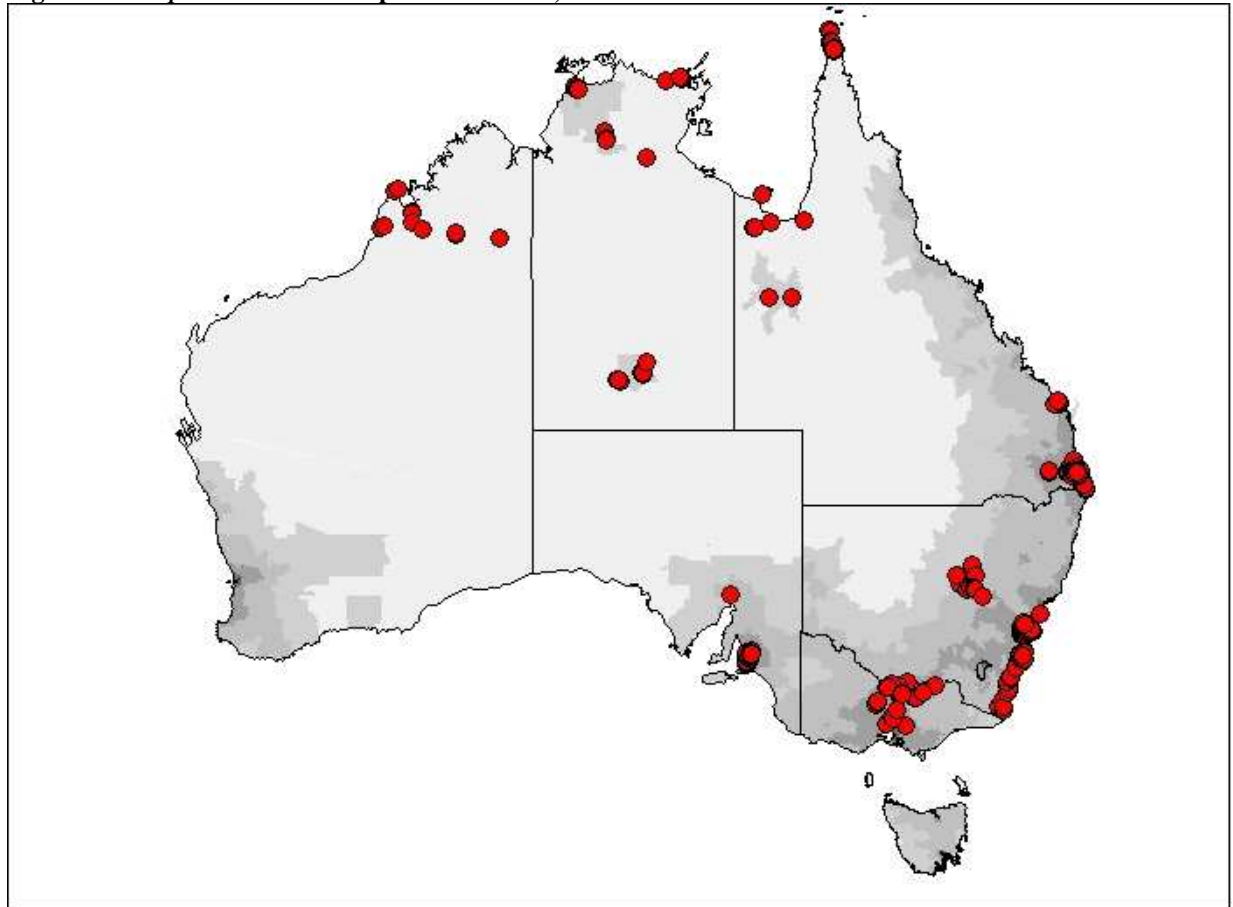
- Alice Springs (and some surrounding communities)
- NT Top End (including Darwin, Katherine, Minyerri and Galiwin'ku)

Apart from site names, *Footprints in Time* reports use the Level Of Relative Isolation (LORI) to describe geographical characteristics of families in the study. Site names are not released with the data for reasons of confidentiality but LORI is available on the datasets.

Box 1: Level Of Relative Isolation

Footprints in Time uses a classification system of remoteness known as the Level Of Relative Isolation (LORI). Previously used in the Western Australian Aboriginal Child Health Survey (WAACHS), LORI is based on an extension of the 18-point ARIA (Accessibility/Remoteness Index of Australia) called ARIA++. Five categories of isolation have been defined, ranging from None (such as the Brisbane metropolitan area) to Low (for example, Shepparton), Moderate (for example, Derby), High (for example, Doomadgee) and Extreme (for example, Moa Island).

Figure 1: *Footprints in Time* sample distribution, Parent 1 interviews



Study development and testing

The design of the study and the content was developed based on extensive consultations with urban, regional and remote Indigenous communities, organisations and service providers across Australia. The overriding goal of consultations was to ensure the design of the research reflected the interests of Aboriginal and Torres Strait Islander peoples, communities and service providers and that the data would genuinely benefit the children and their families.

Piloting of qualitative interviews and community engagement strategies was conducted from September 2004 to December 2005 in the Torres Strait and Northern Peninsula Area (NPA) and in the ACT/Queanbeyan region.

Content rationales were developed based on these consultations, as well as other research such as the Western Australian Aboriginal Child Health Survey (for example, Zubrick et al. 2003) and the National Aboriginal and Torres Strait Islander Health and Social Survey. These rationales were workshopped in November 2005 with members of the Steering Committee and other stakeholders and then used to develop draft questionnaires and computer assisted personal interview (CAPI) instruments.

Piloting of the design, sampling strategy and Wave 1 survey content was undertaken in partnership with the ABS. The ABS tested the questionnaire and field procedures in the first pilot sites in 2006. A number of content areas were adapted to be more

culturally appropriate and/or better understood by respondents before testing the questionnaires and field procedures again in 2007.

Initially six full-time Indigenous RAOs were employed and trained to manage the community engagement activities for the pilot research, including consent processes, data collection and dissemination of information in pilot communities.

Study informants

The study collects data from multiple informants as below:

- **Parent 1 (P1)** - is defined as the parent who knows the Study Child best. In most cases this is the child's biological mother

RAOs take an extensive interview with the primary carer of every study child, containing questions about the study child, parent 1 and the household. It is a face-to-face interview (All waves).

- **Parent 2 (P2)** - is Parent 1's partner or another adult with a parent or carer relationship to the Study Child. In most cases this is the biological father, but step-fathers are also common. Although the surveys were designed to be answered by Grandmas or Aunties or other family who had a caring role there were few respondents who were not fathers. Sometimes this is a parent who does not live with the Study Child, most commonly the biological father after separating from the biological mother. (Waves 1 & 2)

RAOs take a face-to-face interview or a telephone interview (depending on preferences of Parent 2) after completing Parents 1's consent and if Parent 2 is willing to participate and able to be contacted. (Waves 1 & 2)

- **The Study Child themselves** – complete the Vocabulary assessments, practical exercises (such as “Who am I?”) and parent and child height and weight. It includes both face to face interview questions and direct assessments. (The child cohort only in waves 1 & 2)
- **Teachers and childcare workers** - Some data were collected from teachers and carers of study children but these records are relatively few in numbers and will not be released at this point in time. It is expected that more data from teachers and carers will be available in subsequent waves of the study. (Waves 1 & 2)
- **Dads (Wave 4)** - With the focus on Parent 1 and the Study Child, response rates for 'Parent 2' were quite low in Waves 1 and 2 and so Wave 3 data was not collected from Parent 2. A suggestion was put to the Steering Committee that information about Dads might be best collected as qualitative interviews. However the Steering Committee advised that quantitative data that could be linked to the Parent 1 and Study Child surveys was desirable. So the survey was reinstated for Wave 4 however, as the majority of respondents in Waves 1 and 2 were Dads and the Newcastle Family Action Centre advised that Dads should not be seen as secondary parents, the interviews were redesigned to focus only on Dads.

Survey Instruments

The main quantitative data collection for each wave is a structured computer assisted personal interview (CAPI) with the parent who knows the child best (Parent 1). Information collected in the interview includes household, family, pregnancy and birth details, health conditions, nutrition, child care, early education and activities undertaken. Information was also collected about parent health, social and emotional wellbeing, resilience, major life events in the previous year, language, culture, employment, income, and financial stress.

Information about the study child's development was also collected in the Study Child CAPI instrument. This included, for the Baby cohort, an Australian version of the MacArthur-Bates vocabulary checklist completed by Parent 1. For the Child cohort these were the Renfrew Word Finding Vocabulary Test, an assessment of expressive vocabulary, and the *Who Am I?* a school readiness instrument. Height and weight measurements were taken for both cohorts.

With parental or guardian permission, CAPI interviews were conducted with the study child's other parent or another carer (Parent 2) in waves 1 and 2. Questionnaires were also given to children's teachers or child care providers to complete either by hand or online.

Back-up hard copy questionnaires were created of all CAPI instruments so that interviews could still be conducted if there were technical difficulties with the laptops in the field. Copies of questionnaires are available from <http://www.fahcsia.gov.au/l sic> and are supplied with the data.

Data collection

Consent process

The process for gathering informed consent from the study child's parents or carers and their family, teachers and carers forms an integral part of the study.

Prior to being interviewed parents were provided with an introductory letter and a DVD describing the study and the consent process. At the interview RAOs went through each consent form with the participant/s and explained what permission was being sought. This enabled parents to make informed consent about their participation in the study. A plain language statement was also available for parents who preferred to read about the study. Parents gave consent on behalf of the Study child.

As well as seeking permission to take part in the study participants were separately asked for consent to:

- be voice recorded for the interview
- allow the other parent or another carer to be contacted
- allow the child's teacher or child care worker to be contacted
- allow the Study child to be photographed

At the conclusion of the consent process participants were given a summary sheet that recorded what they had agreed to. This sheet included contact details for the ethics committee and FaHCSIA. Participants were informed that they could change their consent and are able to withdraw from the study at any time.

Fieldwork periods

The pilot fieldwork for Wave 1 was conducted from the end of 2006 to early 2007. The main fieldwork for Wave 1 started on 21 April 2008 and finished 23 February 2009. The Wave 2 pilot was conducted between October and November 2008, and the main round of interviews between 03 March 2009 and 17 December 2009. Although it is the aim of the study to interview participants at 12 month intervals, this was not always possible in Wave 2, because of the availability of respondents and the logistics of interviewers' travel arrangements and scheduling. Nearly 84 per cent of Wave 2 interviews were conducted between six and 12 months after the Wave 1 interview. The average intervening time between Waves 1 and 2 interviews was 10 months. The Wave 3 pilot fieldwork was conducted from Sep to Nov 2009 and the main interviews from 03 March 2010 to 23 December 2010. Wave 4 pilot interviewing commenced in October 2010 and ended November 2010. The main interviewing for Wave 4 started on 7 March 2011 and is scheduled to end on 9 December 2011.

Table 2: Fieldwork periods

Wave	Pilot	Main
Wave 1	2006-2007 and Jan 2008	21 April 2008 to 23Feb 2009
Wave 2	Nov 2008	3 March to 17 Dec 2009
Wave 3	Oct 2009	3 March to 23 Dec 2010
Wave 4	Oct 2010	7 March and scheduled to end 9 Dec 2011

Fieldwork Response

Out of the 1,670 families who participated in Wave 1, *Footprints in Time* interviewers successfully interviewed 1,436 families in Wave 2, achieving an overall response rate of 86 per cent for the study. In Wave 3 a total of 1,315 study child families were interviewed which is 86 per cent of the previous Wave 2 numbers.

Table 3: Fieldwork response from Wave 1 to Wave 3

	Previous wave respondents interviewed	Additional interviews	Total interviews	% of retention from previous wave
Wave1	1,670	1,670

Interview				
Wave 2 responding	1,435	88*	1,523	85.6
Wave 3 responding	1,315	92**	1,407	86.3

*New entrant in Wave 2

** Interviewed in waves 1 & 3, not interviewed in wave 2

Note – New entrants were admitted into study in wave 2, but not in wave 3.

New entrants

In order to maintain the viability of the sample in remote regions and meet the requests of a small number of families who expressed a strong wish to be part of the study, *Footprints in Time* added 88 new entrant families to the study in Wave 2. With the addition of 88 new entrant families, the total number of responses achieved in Wave 2 was 1,524.

Interview length

Roy Morgan Research has estimated the time taken for Wave 2 and Wave 3 interviews based on a combination of anecdotal evidence and on the computer-captured data excluding those cases that appeared implausible.

Wave 4 questionnaires have timestamps programmed into the start and finish of each questionnaire section in order to accurately capture the length of each interview section and the interview as a whole.

Table 4: Length of Interviews by Cohort and Wave

Cohorts	Range W2	Average length W2	Range W3	Average length W3
Study Child B Cohort	5-50 minutes	10 minutes	5-47 minutes	10 minutes
Study Child K Cohort	5-50 minutes	17 minutes	5-49 minutes	19 minutes
Parent 1 B Cohort	Half an hour to 3 hours	1 hour	Half an hour to 3 hours	52 minutes
Parent 1 K Cohort	Half an hour to 3 hours	1 hour	Half an hour to 3 hours	57 minutes
Parent 2 B Cohort	10-60 minutes	Half an hour	n/a	n/a
Parent 2 K Cohort	10-60 minutes	Half an hour	n/a	n/a

Source – Roy Morgan Research Reports

Questionnaire content overview

Wave 1 - 4 data includes a range of information which will be longitudinal (usually collected annually) as well as developmentally age-specific information. The following list shows briefly which material is included in each wave.

Sign √ means topic was asked, NE means asked only of new entrants and NPC means asked only of a new primary carer.

Parent 1

	W1	W2	W3	W4
Household				
Dwelling Type & Street traffic	√	√	√	√
Household demographics: sex, age, Indigenous status, relationship to Parent 1.	√	√	√	√
Child Health				
Maternal Health and Care, Alcohol; tobacco & substance use in pregnancy; Birth	√	NE		
Early Diet and Feeding	√	√		
Nutrition	√	√	√	√
Dental Health	√	√	√	√
Health Conditions	√	√	√	√
Injury		√		√
Hospitalisation	√	√	√	√
Child's sleeping patterns	√	√	√	√
Parental Health				
Ongoing health conditions	√	√	√	√
Strong Souls	√	NPC	NPC	√
Social and Emotional Wellbeing	√	√	√	√
Smoking habits and exposure (& alcohol Wave 2)	√	√	√	√
Gambling			√	

Parent relationship			√	
Stolen generation		√		
Parent living elsewhere	√	√	√	√
Child and Family Functioning				
Child social, emotional development	√			√
Strengths and Difficulties © Robert Goodman			√	√
Physical ability		√	√	√
Child temperament		√		
Brief Infant Toddler Social and Emotional Assessment		√		
Parent Concerns about language and development	√	√	√	√
Parental warmth, monitoring, consistency	√	√	√	√
Major life events	√	√	√	√
Socio-demographics				
Participant language, culture and religion	√	√	√	√
Parental Education		√	NPC	NPC
Work	√	√	√	√
Financial stress and income	√	√	√	√
Child Support and Maintenance		√		√
Housing and Mobility	√	√	√	√
Child Care and Early Education	√	√	√	√
School		√	√	√
Activities	√	√	√	√
Interviewer questions		√	√	√
Weight, height, language and school readiness				
Word Finding	√	√	√	√
Vocabulary checklist for babies	√	√	√	
Who am I?	√	√	√	√

Favourite Things		√		
School			√	√
Height and weight	√	√	√	√
Drawing task			√	
MATRIX reasoning (from WISC-IV)				√
Progressive Achievement Tests in Reading (PAT – R)				√

Study Child

Study child survey: Weight, height, language and school readiness				
Word Finding	√	√	√	√
Who am I?	√	√	√	√
Favourite Things		√		
School			√	√
Height and weight	√	√	√	√
Drawing task			√	
Matrix reasoning				√
New Module – PAT- R				√
Interviewer questions			√	√

Parent 2 (Dads in Wave 4)

	W1/P2	W2/P2	W4/Dad
Household			
Dwelling Type & Street traffic	√	√	√
Household Form	√	√	√
Parental Health	√	√	√
Ongoing health conditions	√	√	√
Strong Souls	√	√	√
Social and Emotional Wellbeing	√	√	√

Smoking habits and exposure	√	√	√
Parent relationship	√	√	√
Childhood and Parenting			
Stolen generations		√	
Parent living elsewhere			√
Child & Family Functioning			
Parent warmth, monitoring, consistency	√	√	√
Major life events	√	√	√
Socio-demographics			
Parental language, culture and religion	√	√	√
Parental Education	√	√	√
Work	√	√	√
Financial stress and income	√	√	√
Child Support and Maintenance		√	√
Housing and Mobility	√	√	√
School			√
Involvement with study child			√
Activities P2 does with study child	√	√	√
Does P2 ever take study child (preschool/kindergarten/preprep) or childcare?			√

Using the datasets

The table below shows the number of records in each file released to date.

	Wave1	Wave 2
Parent 1	1676	1523
Parent 2	257	269
Study Child	1485	1472
Teacher	n\a	n\a
Centre Based Carer (CBC)	n\a	n\a

Note – numbers in the datasets may vary from previously released version as participants has right to drop them from the study. On rare occasions they ask that all their data be removed.

Data was collected from some teachers and carers of study children in waves 1, 2 and 3. As the number of these responses was small in Waves 1 and 2, it is planned to release all three waves at once (probably with Release 3.0).

Locating variables

To locate variables of interest look through the mark up questionnaires and/or the Data Dictionary.

The mark up questionnaires give the full wording and sequencing of all questions, the variable names and answer categories for all variables. They also specify the new answer categories which were made after analysis of the free text responses and provide the names of derived variables.

The Data Dictionary is an excel workbook providing details of all variables in the LSIC datasets. There is a separate sheet for each survey instrument: P1, P2 and SC. Each spreadsheet contains the variables for all released waves of data. The Data Dictionary can be searched using filters to find variables of interest.

A description of each of the columns in the Data Dictionary can be found at **Appendix A**.

Variable Naming Convention

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. LSIC variables are a maximum of eight characters in length. The variable name is divided into four parts and provides information on the content of the variable.

- First character – Wave identifier, with ‘a’ being used for Wave 1 and ‘b’ being used for Wave 2
- Second character – Subject/informant, a=Parent 1, b=Parent 2, c=Study Child,

- Third and fourth character – Topic name, such as HF for household form, SS for strong souls, etc
- Fifth to eighth character – Arbitrary number within topic – this mainly relates to question numbering and sub-numbering within the topic on the paper questionnaire. An underscore is used, where possible, for variable items that are a categorical answer to a question where more than one category can be chosen. For questions where only one category of answer is allowed, the underscore will not be used. Examples of these are:

 aamc2_1 (Wave 1, Parent 1, MC, Question 2, Category 1 – Mother and/or aunts)

An underscore is not used in variables where there is no room for it. For example:

 aac1baa (Wave 1, Parent 1, AC, Question 1b – Who did this with (him/her)? Sub-question A – play music, etc (Answer = Mother))

Identifiers

Each Study Child has a unique cross wave identifier (called xwaveid) which is constant for all interviews relating to that child (whether P1, P2, SC or teacher/carer) and remains stable across waves. It is composed of 6 numbers- the first two indicate the wave when the Child first entered the study. Records for children who were part of the Study from Wave 1 start with 01. Wave 2 new entrants start with 02. These numbers are followed by four randomly assigned numbers.

Cohort

At the beginning of each instrument, interviewers select whether the study child belongs with the B (baby) or K (child) cohort. This selection determines the sequencing of future questions as not all questions are asked of both cohorts. The variable for cohort is aachtype (wave1, parent1), bachtype (wave 2, parent1), abchtype (wave1, parent2) etc.

Geographic variables

LSIC was conducted in 11 sites from around Australia but for confidentiality reasons the site variable is not released. Researchers can use the variable alori which provides an indicator of the level geographical of remoteness. See p9 for further information about LORI.

Household Form

The household form in the P1 instrument collects basic demographic information (age, sex, indigenous status and relationship to P1) of all members of the household.

In Waves 1, 2 and 3 the information for P1, P2 and Study Child were entered into the first three places and other household members could be enumerated in any order. This meant that a particular individual could be member 4 in Wave 1 and member 6 in Wave 2. As researchers are not given access to the names of household members, it would be impossible to analyse movements of individuals in and out of households.

To overcome this problem, the **household data has been reorganised** so that each individual has a permanent household member number/position (from Release 1.2 onwards). All data is missing if the member is not present in that wave. The study child is always member 1. The P1 in wave 1 is always member 2, even if they are no longer remaining the P1. The P2 in wave 1 is member 3 (if there was a P2 in the household). Other household members take positions 4 onwards. If a new member joins the household they are given the next free position. Separate variables identify the member numbers of P1 and P2 (if present) in each wave.

During the reorganisation process, the household data was also cleaned to remove obvious inconsistencies. Three waves of data were available during this process, so logical decisions could be made on the basis of the combined data. For example, where a member was described as “Aboriginal” in two waves and “neither” in the other, this was changed to “Aboriginal”. Where sex varied across waves it was corrected in line with the person’s name (where obvious). Sometimes paper files were referred to for confirmation. It became evident that many people were not aging over the three waves. While it is possible for a person to remain the same age in two consecutive waves due to timing of interview and birth date, it is not possible to be the same age in three consecutive waves. A decision was made to reinterpret the age variable as “age this year” rather than “age at interview”, and to adjust age values so that all members age by one year each wave. Because the household data has been cleaned as described, it cannot be used to analyse variability in answers.

The household form collects date of birth rather than age for the Study Child (and P1 and P2). As child development occurs rapidly over the early years, a variable for the study child’s age in months at the time of interview (ascagem) has been derived to enable relevant analysis.

Direct Assessments of child development

Direct measures include the Who Am I? developmental assessment and the Renfrew Word Finding Vocabulary Test, which were undertaken by the Child cohort in Waves 1 and 2. These verbal and non-verbal measures assess processes that underlie the learning of early literacy and numeracy skills. The measures start at a point where the vast majority of children experience some success and most children enjoy them. Although the measures are designed to progressively get more difficult, they are stopped when the child is unable to complete the harder items.

Who am I? (de Lemos & Doig 1999) is a developmental assessment that requires the child to write their name, copy shapes, write letters, numbers and words in a small booklet, with simple instructions and encouragement from the interviewer. Who am I?

is not language dependent and is suitable for children with limited English. The assessment takes about 10 minutes to complete and is suitable for preschool children and children in the first two years of school.

The Renfrew Word Finding Vocabulary Test (Renfrew 1998) assesses children's expressive vocabulary – compared, for instance, with the Peabody Picture Vocabulary Test (Dunn & Dunn 2007)¹, which is a test of receptive vocabulary. The Renfrew Word Finding Vocabulary Test assesses a child's ability to accurately describe images as portrayed in the 50 pictures contained in the assessment. This ability is one aspect of the general ability to communicate one's ideas clearly and to understand the communication of others, which are vital pre-requisite skills to learning in the classroom. A child's strength or weakness in expressive language can be identified when we ask the child to ask and answer questions, describe images, articulate thoughts and ideas, and respond appropriately to the communication of other people.

Both of the direct assessments can provide information about the extent to which a child is ready for the early years classroom tasks that are associated with subsequent literacy and numeracy development at school.

Qualitative data

A range of qualitative data items are collected as part of *Footprints in Time* in the form of free text responses to a small number of open-ended questions in the survey. Free text entries responses to open-ended questions are included in the data releases, with references to places, individuals, employers, clans, family names and languages suppressed. References to rare circumstances that may have been of sufficient noteworthiness to be known by the wider community are also suppressed. The risk of identification is expected to be low given the confidentialised status of these data, however data users need to be mindful at all times of their responsibility to not risk identification of respondents. The text responses incorporated in the datasets have been truncated to a fixed number of characters. The full responses can be viewed in the spreadsheets which come with the data. A lists of these free text variables are provided at **Appendix B**.

Data users are permitted to directly quote free text responses on the basis that such usage poses no risk of the respondent being rendered identifiable. Quotes can be accompanied by relevant unit record data such as age or occupation if these details are required for meaningful interpretation but the unit record data used should be the minimum required for the data user's purpose and should manifestly carry no risk of identifying the respondent. For example, it is acceptable to report that 'One mother who has a Bachelor degree commented "I want him to go to university and have a good career"' but it is not acceptable to use unconfidentialised data to report that 'One mother, who works as a Professor of Indigenous studies at a university, commented "I want him to go to university and have a good career"'.

For the purposes of keeping data files to a manageable size, free text entries in the data releases are truncated to a maximum character length, with any remaining characters discarded. Data users interested in accessing free text responses that have not been truncated to a maximum character length are encouraged to contact the *Footprints in Time* team.

Other - specify responses

The LSIC questionnaire is unusual in permitting interviewers to enter an “Other-specify” response for almost every question (in Wave 1 & Wave 2). This enabled interviewers to type in a response to the question, rather than requiring them to guess which category to select when there was no obviously appropriate category. This enabled us to limit the number of answer categories because we didn’t need to have a category for every possible response. It also meant that we were made aware when an important answer category had been missed and could amend the questionnaire if the question was repeated in a later wave to include this new category. For some variables, responses entered in “Other- specify” have been backcoded to existing categories where appropriate, but the majority have not.

The “Other- specify” category is of limited use to researchers without the accompanying text file. It was decided to code all "Other- specify” responses to “-1” for ease of use. They can be easily included or excluded from analysis, and do not cause confusion when variables are numeric quantities (eg. Age, number of weeks, etc) or Leichardt scales. The only exception to this is for multiple response questions, when a respondent can legitimately select both “Other- specify” as well as another response category. In these cases there is a separate variable indicating whether “Other- specify” was selected.

Missing data coding

The convention for dealing with missing data in LSIC is similar to, but not the same as, either of the conventions used by the LSAC and HILDA survey. Missing data is coded “-2” to “-8”, as per the table below.

- 1 Other (When explicitly available as an option in the questionnaire)
- 2 Don’t know (When explicitly available as an option in the questionnaire)
- 3 Refused (When explicitly available as an option in the questionnaire)
- 4 Refused section (When explicitly available as an option in the questionnaire. Used both for the screener question for a section that can be refused as well as all the variables within that section)
- 5 Not asked (Indicates a question that has been skipped due to normal sequencing or a free text or numeric answer category which has been intentionally left blank)
- 6 Not asked of this cohort (Some questions are only asked of B cohort or K cohort)
- 7 Implausible value (Indicates where value has been deleted during cleaning- eg 800kg person)

- 8 Missing data (Data not collected where it might be expected. Used where an answer is not provided, although based on sequencing and programming an answer should have been provided)

Merging datasets

Datasets can be merged across waves or within the wave (eg P1 and P2 for wave1) by one-to-one matching on the unique identifier (xwaveid). The code to do this will be specific to the analysis package used. Two examples of merging in stata are provided at **Appendix C**. If researchers are experiencing difficulty merging datasets, they should contact the LSIC Data Team at <LSICdata@fahcsia.gov.au>.

Confidentialisation

A number of variables have been removed from the data as these could easily compromise the identity of the respondent. These include:

- Names of household members
- Date of birth of Parent 1, Parent 2 and Study Child
- Site: Although we release site/community related information in the form of pamphlets and in publications, unit record data about site or state is not released. However, a variable containing geographic information about 'Level of Relative Isolation' (LORI) is included with the dataset.
- Respondent ID: This is the identifier that is known to the participants and contains site information, which is different to the anonymous "xwaveid"

The General Release dataset has been further confidentialised in a range of ways.

- Age – age person turns in year of interview were topcoded for all persons aged over 65. All P1s aged over 65 were given the average age of this group. All P2s aged over 65 were given the average age of their group. All other people over the age of 65 were given the average age of the non-P1/P2s over 65.
- Language – respondents could choose from 179 language names. As some of the languages are only spoken in specific geographical areas, this data has been grouped into six language categories.
- All names of people, names of companies and names of businesses have been removed from the free text responses in the datasets and in the accompanying spreadsheets.

A small number of free text responses with references to rare occupations, health conditions or major life events have been generalised in order to protect individuals from identification.

Data access

There are strict security and confidentiality protocols surrounding use of the data. Prospective users are required to complete a dataset application and read and sign a deed of licence. These can be found on the <<http://www.fahcsia.gov.au/l sic>>

Queries about the study or the data should be sent to <LSICdata@fahcsia.gov.au>. Queries about applying for the data or licensing arrangements should be sent to <longitudinalsurveys@fahcsia.gov.au>.

The arrangements for access are very similar to those already applying for access to the datasets for the Longitudinal Study of Australian Children (LSAC) and the Household, Income and Labour Dynamics in Australia (HILDA) survey. However applicants and licensed users will be asked to openly acknowledge their standpoint in their application and in the reporting of data outputs in reports or publications.

Those who are interested in applying for the *Footprints in Time* (LSIC) data should read Section 9 and Appendix A in the *Manual for the access to and use of FaHCSIA longitudinal survey datasets* ([link](#)) before completing their application.

Data linkage

Australian Early Development Index (AEDI)

The Australian Early Development Index (AEDI) is a nation-wide program that looked at the development of young children. Between 1 May and 31 July 2009, teachers completed the AEDI checklist for children in their first year of full time school. The AEDI measured five areas of early childhood development: physical health and wellbeing, social competence, emotional maturity, language and cognitive skills (school-based), and communications skills and general knowledge.

LSIC will be doing two types of data linkage with AEDI scores:

1. LSIC will include aggregated AEDI data for all LSIC children. This will be based on the suburb that the child lived in at their Wave 2 LSIC interview. The aggregated dataset will include the average AEDI scores across the five areas of early childhood development by suburb, the proportion of ‘developmentally vulnerable’ children in that suburb, as well as some demographic information. This data linkage will be available in a later release.
2. We also sought permission from parents of K cohort children to link specifically to their child’s AEDI data. Where parental permission was obtained and an AEDI was completed by a teacher, LSIC will link that child’s AEDI data to LSIC data, in a separate data set and will be available for the data users in a later release.

Getting more information

More information on *Footprints in Time* and its progress can be found on the LSIC website: <<http://www.fahcsia.gov.au/l sic>>

Further enquiries can be directed to <LSICdata@fahcsia.gov.au> or by contacting:

Jason Brandrup

LSIC Data Manager

PO Box 7576, Canberra Business Centre

ACT - 2610

Tel: +61 2 6244 5384

Fax: + 61 2 6204 4726

Appendix A

Data Dictionary

Headings	Description
Variable	Indicates the variable name in the dataset (without the first letter which pertains to wave).
Variable label	Indicates the variable label as used in the dataset.
Question	Gives the exact question wording used in the questionnaire.
Value label	Indicates the answer categories available in the questionnaire together with the numeric value used in the dataset.
Wave	Indicates the year in which interviewing took place- Wave 1 is the first year of interviewing and takes place annually.
Person label	Indicates whether the respondent is P1 (parent one), P2 (parent two), SC (study child) or a teacher or carer.
Variable type	Indicates whether a variable is categorical, ordinal, text, continuous or multiple responses.
Derived variables	Indicates variables derived from information collected through the survey instruments.
Topic Code	Indicates the two initials which designate which section of the questionnaire the question comes from, also appears in the variable name, eg. HH or SS.
B Cohort	Question asked of the Baby cohort (mainly aged 6-24 months in wave 1)
K Cohort	Question asked of the Child cohort (mainly aged 3.5-.5 years in wave 1)
Population (see also 'Cohort')	Indicates whether sequencing affects the population of whom each question was asked. eg. if question is only asked of birth mothers.
Topic	Indicates the name of the section in which the question appears in the questionnaire, eg. Household Form or Strong Souls.
Sub-topic	Gives a specific description of the question. There is one sub-topic per question in the questionnaire. Multiple response variables relating to one question will share one sub-topic.
Theme	Shows some thematic links between questions in different parts of the questionnaire.
Notes	Provides extra information that might be useful in interpreting the data, eg. Variations in question wording for different cohorts or information on how some variables were derived.
File order	Indicates the original order of the variables in the dataset. Can be used to sort the spreadsheet back to the original order.

Appendix B

Qualitative free text questions\variable label

Question\variable label	Variable Name (excluding first letter indicating wave)	Wave1	Wave 2	Wave 3
Why do you feel [this community] is unsafe/dangerous?	ahm13_t/ bhm13_t	P1, P2	P1, P2	
Is there anything else you want to say about your community?	ahm14_t	P1, P2	P1	
Are there any new major things that need fixing - text	ahm7_t	P1, P2	P1	
Do you know where to get help?	ahm17_t		P1	
Whether at home or away what are a couple of things you enjoy doing with the Study Child?	aac8_t/bac8_t	P1, P2	P1, P2	
Whether at home or away what are a couple of things Study Child enjoys doing with you?	aac9_t/bac9_t	P1, P2	P1, P2	
Apart from health and happiness what do you want for your Study Child?	aac10_t	P1, P2		
What is it about Aboriginal or Torres Strait Islander culture that will help Study Child grow up strong?	aac11_t	P1, P2		
What happens before sleep – text	acs4_t		P1	
Reason for hospitalisation of SC – REASON 1 - text	aho1a1t	P1	P1	
Reason for hospitalisation of SC – REASON 2 - text	aho1a2t	P1	P1	
Reason for hospitalisation of SC – REASON 3 - text	aho1a3t	P1	P1	
Reason for hospitalisation of SC – REASON 4 - text	aho1a4t	P1	P1	
Reason for hospitalisation of SC – REASON 5 - text	aho1a5t	P1		
P1 main job - text	awo3_t	P1		
P1 main field of study - text	ape4_t	P1,P2		
Concerns about how Study child uses hands and fingers to do things.	ald8_t	P1	P1	
Concerns about how Study child uses arms or legs.	ald9_t	P1	P1	

Concerns about how Study child behaves.	ald10_t	P1	P1	
Concerns about how Study child gets along with others.	ald11_t	P1	P1	
Concerns about how Study child is learning pre-school and school skills.	ald12_t	P1	P1	
Concerns about how Study child is learning to do things for himself/herself.	ald13_t	P1	P1	
Concerns about Study child's learning or development.	ald14_t	P1	P1	
What other major events or stressful situations happened to you, your family or (STUDY CHILD) since this time last year?	ame16_t/ bme16_t	P1,P2	P1,P2	
What do you do to cope with stress - text	ame17_t/ bme17_t	P1	P1,P2	
SC attend playgroup or baby group – text	ace1_t	P1	P1	
SC has been bullied at preschool/school because they are Indigenous – text	ace23_t		P1	
How P1 deal with racist bullying of SC – text	ace24_t		P1	
What P1 hopes that SC will do or learn next year - text	aact19_t		P1	
Good education for SC – text	aac18_t		P1	
Anything else P1 wants to tell us – text	aac12_t/bac12_t	P1, P2	P1, P2	
Bush tucker eaten – text	anu4_1	P1		
Something that's happened for the SC since last year	brt1_t		P2	
What would a good education be for SC	bac32_t		P2	
Thinking about the next year, what are you hoping SC will do or learn?"	bac33_t		P2	

Appendix C

Examples of merging in stata

```
*Stata example of merging P1 wave 1 and P1 wave 2 data

version 11 /*merge syntax is slightly different for earlier versions
of ststa*/
use "[Substitute folder name here]\lsicplw1_20c.dta", clear
merge 1:1 xwaveid using ///
    "[Substitute folder name here]\lsicplw2_20c.dta.dta"
```

```
*Stata example of merging P1 wave 2 and P2 wave 2 data

version 11 /*merge syntax is slightly different for earlier versions
of ststa*/
use "[Substitute folder name here]\lsicplw2_20c.dta", clear
merge 1:1 xwaveid using ///
    "[Substitute folder name here]\lsicp2w2_20c.dta.dta"
```

¹ The Renfrew Word Finding Vocabulary Test allows interviewers to record words used in other languages, while the Peabody Picture Vocabulary Test uses complex line drawings of pictures not commonly seen in Australia, for example, toboggans. Therefore, Renfrew was considered more appropriate for this study.