

# 2021 Australian Digital Inclusion Index Index Technical Report

July 2021



Social  
Research  
Centre

### Report prepared for:

Abigail Brydon

Telstra

Strategy and Inclusion

Melbourne, VIC, 3000

[Abigail.Brydon@team.telstra.com](mailto:Abigail.Brydon@team.telstra.com)

Indigo Holcombe-James

RMIT University

School of Media and Communication

Melbourne, VIC, 3000

[Indigo.holcombe-james@rmit.edu.au](mailto:Indigo.holcombe-james@rmit.edu.au)

Dr Sharon Parkinson

Swinburne University of Technology

Centre for Social Impact

Hawthorn, VIC, 3122

[sparkinson@swin.edu.au](mailto:sparkinson@swin.edu.au)

### Report prepared by:

Matilda Page, Research Associate

Paul Myers, Executive Director Quantitative Research

The Social Research Centre

Level 9, 277 William Street

MELBOURNE VIC. 3000

Tel: (613) 9236 8500

Fax: (613) 9602 5422

Email: [paul.myers@srcentre.com.au](mailto:paul.myers@srcentre.com.au)

[www.srcentre.com.au](http://www.srcentre.com.au)

Version: 1 (16/07/2021)



# Contents

- 1. Introduction .....1**
  - 1.1. About this report.....1
  - 1.2. Project background .....1
  - 1.3. Project overview.....2
- 2. Methodology .....3**
  - 2.1. Overview .....3
  - 2.2. Sampling approach .....3
    - 2.2.1. G-NAF sample frame.....3
    - 2.2.2. Sample design and selection .....3
    - 2.2.3. In-scope sample .....4
  - 2.3. Data collection .....4
    - 2.3.1. Approach and reminder details .....5
    - 2.3.2. Additional response maximisation techniques .....6
- 3. Questionnaire development.....7**
  - 3.1. Overview .....7
  - 3.2. Online survey instrument .....7
  - 3.3. Hard copy questionnaire .....8
- 4. Response analysis .....9**
  - 4.1. Overview .....9
  - 4.2. Response by completion mode ..... 10
- 5. Data processing ..... 12**
  - 5.1. Hard copy returns processing ..... 12
    - 5.1.1. Returns management and reporting.....12
    - 5.1.2. Data scanning and capture ..... 12
  - 5.2. Data cleaning rules ..... 12
  - 5.3. Coding..... 13
  - 5.4. Weighting ..... 13
    - 5.4.1. Overview.....13
    - 5.4.2. Design weights ..... 14
    - 5.4.3. Calibrating to population benchmarks .....14
    - 5.4.4. Benchmarks.....15
    - 5.4.5. Treatment of missing values ..... 15
- 6. References ..... 17**
- Appendix 1 Invitation letter and brochure ..... 18**
- Appendix 2 Reminder postcard ..... 19**
- Appendix 3 Reminder letter..... 20**
- Appendix 4 Final online questionnaire ..... 21**
- Appendix 5 Final hard copy questionnaire..... 22**
- Appendix 6 Data cleaning rules ..... 23**

# List of figures

Figure 1 Approach and reminder schedule ..... 3  
Figure 2 Completes by date and mode ..... 9

# List of tables

Table 1 Key project statistics ..... 2  
Table 2 G-NAF sample design ..... 4  
Table 3 Respondent location by mode (unweighted) ..... 10  
Table 4 Respondent characteristics by mode (unweighted)..... 11  
Table 5 Population benchmarks used for calibration..... 14  
Table 6 Population benchmarks used for calibration..... 15  
Table 7 Extent of missing values among weighting characteristics..... 16

# 1. Introduction

## 1.1. About this report

This report summarises the data collection and methodological aspects of the 2021 Australian Digital Inclusion Index (ADII) Survey (also known as the Australian Internet Usage Survey (AIUS)).

The survey was conducted by the Social Research Centre on behalf of Swinburne University and RMIT University (the 'ADII Research Team') between April and June 2021. The survey was funded by Telstra, who worked closely with the Social Research Centre and the ADII Research Team.

This report seeks to:

- consolidate and summarise project information and assorted reports generated throughout the survey period
- provide analysis relating to sample characteristics and utilisation
- summarise data processing and weighting processes
- consolidate issues for consideration related to the refinement of the methodology for future surveys.

## 1.2. Project background

The ADII Survey is a series of surveys beginning in 2014, which aims to provide a comprehensive picture of Australia's online participation. Until 2019, the ADII Survey utilised data from the Roy Morgan Single Source. In 2019, the ADII Research Team undertook a comprehensive review of the ADII to ensure it remains relevant and effective given:

1. rapid and ongoing changes in digital technologies and the growing significance of the online distribution and consumption of services
2. requests from stakeholders for richer data insights that include the release of more of the data that populates the Index
3. interest from stakeholders in having access to a customised digital inclusion survey and reporting instrument they could use to measure digital inclusion in their own communities.

Following this review, the ADII Research Team engaged the Social Research Centre to conduct a pilot study in 2020 using a new methodology, before a full launch of data collection in 2021.

The ADII aims to provide a long-term picture of digital inclusion across Australia by measuring three key dimensions: Access, Affordability and Digital Ability. The ADII investigates how these dimensions change over time according to people's social and economic circumstances, as well as across geographic locations. The data from the ADII Survey helps to provide a comprehensive picture of who uses the internet in Australia, what benefits Australians get from the internet and what barriers exist to accessing the internet. Findings from the ADII inform policy, community programs and business efforts to increase digital inclusion in Australia so that everyone can make full use of digital technologies.

### 1.3. Project overview

The in-scope population for the 2021 ADII Survey was adults (18 years of age or over) who are residents of private households in Australia. A sequential mixed-mode design was applied to data collection with participants self-completing online during the push-to-web phase or via hard copy during a second phase of collection.

The Social Research Centre managed the survey distribution and data collection in partnership with:

- Valiant Press – for hard copy questionnaire forms design, printing and distribution
- Datatime Services – for hard copy questionnaire scanning and data processing.

The data collection period was 6 April to 3 June 2021.

A total of 19,999 addresses were randomly selected from the Geocoded National Address File (G-NAF) for the 2021 ADII Survey. The selected respondent at the selected address was any household member aged 18 years or over. The final sample size was 2,287, equating to a response rate of 11.0% and sample yield of 11.4%.

Key project statistics are summarised in Table 1 below.

**Table 1** Key project statistics

Key project statistics	2021 Sample
Total sample	19,999
<b>Total completed surveys</b>	<b>2,287</b>
<i>Online</i>	1,170
<i>Hard copy</i>	1,117
AAPOR Response Rate 3 (%)	11.0
Sample yield (%)	11.4

Ethics approval for the 2021 ADII Survey was carried over from the previous survey in 2020. Ethics was granted by the Human Research Ethics Committee of Swinburne University (ref: 20203023-4585). All data collection activities were undertaken in accordance with the *Privacy Act 1988* (Cth) and the Australian Privacy Principles contained therein, the *Privacy (Market and Social Research) Code 2014*, the Research Society's Code of Professional Behaviour and ISO 20252 standards.

## 2. Methodology

### 2.1. Overview

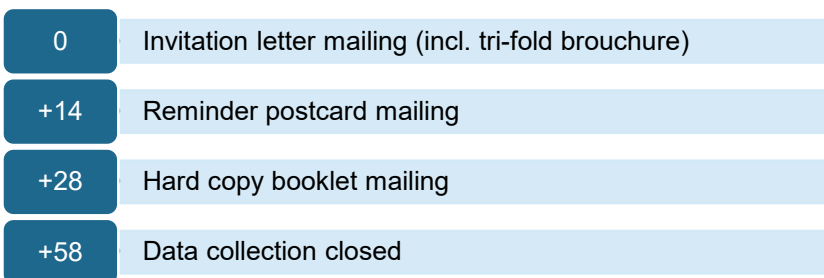
Designed by the Social Research Centre, the methodology for the ADII Survey involved a sequential mixed-mode approach to data collection. This approach involved an initial push-to-web phase whereby online self-completion was offered first, with hard copy completion initiated only after efforts to maximise online response were exhausted.

An overview of the contact approach implemented is shown in Figure 1.

Further details about this approach are outlined in Section 2.3.

**Figure 1 Approach and reminder schedule**

**Days' since commencement**



### 2.2. Sampling approach

An address-based sampling (A-BS) approach was used for the ADII Survey. A-BS is a sampling approach that selects addresses from a comprehensive listing of virtually all addresses. The sample frame was the Geocoded National Address File (G-NAF). Sample was selected using a stratified sample design in accordance with the distribution of the Australian residential population.

#### 2.2.1. G-NAF sample frame

The G-NAF sample frame is maintained by Geoscape Australia (a company jointly owned by the Commonwealth, State and Territory governments that was formerly the Public Sector Mapping Authority) and is the authoritative national address index for Australia.

The G-NAF is compiled of existing addresses from three recognised sources: State and Territory Government land records, the Australian Electoral Commission's Electoral Roll data and validated addresses from Australia Post.

#### 2.2.2. Sample design and selection

The sample design involved establishing fifteen geographic strata based on state or territory and capital city/rest of state. Quotas were set for each strata in proportion to population. Addresses within strata were randomly selected. The final number of selections for each strata was also varied to account for the anticipated sample yield. This design aimed to yield a sample that closely approximates the distribution of the Australian population across states and territories and capital cities and other areas (see Table 2).

**Table 2** G-NAF sample design

Region	Available sample	Selected sample	Proportion of selected sample (%)	Australian adult population* (%)
Greater Sydney	2,854,105	4,377	201.9	20.6
Rest of New South Wales	1,804,253	2,177	10.9	11.3
Greater Melbourne	2,801,944	3,682	18.4	19.9
Rest of Victoria	1,014,800	1,155	5.8	6.3
Greater Brisbane	1,411,646	1,800	9.0	9.6
Rest of Queensland	1,726,413	2,224	11.1	10.3
Greater Adelaide	767,279	926	4.6	5.4
Rest of South Australia	346,897	371	1.9	1.6
Greater Perth	1,133,467	1,544	7.7	8.1
Rest of Western Australia	380,936	650	3.3	8.1
Greater Hobart	137,765	148	0.7	0.9
Rest of Tasmania	204,455	201	1.0	1.2
Greater Darwin	76,507	208	1.0	0.6
Rest of Northern Territory	34,090	300	1.5	2.2
Australian Capital Territory	230,157	236	1.2	1.7
<b>Total</b>	<b>14,924,714</b>	<b>19,999</b>	<b>100.0</b>	<b>100.0</b>

Address selections from the G-NAF are checked against Australia Post's Postal Address File (PAF) for deliverability and address type. Addresses without a delivery point identifier (DPID) or flagged as non-residential in the PAF were excluded.

A total of 19,999 sample addresses were randomly selected within 15 geographic strata (see Table 2) to ensure sufficient sample was utilised to achieve the desired number of responses for the ADII.

### 2.2.3. In-scope sample

A selected household was considered in-scope for the survey if it contained one or more Australian adults aged 18 years or over.

In the interest of response maximisation, a decision was made to allow any responsible adult within the household to complete the survey rather than apply a within-household selection procedure.

This decision was based on the knowledge that within-household selection methods have been found to add a layer of complexity that increases non-response (Battaglia et al., 2008). Thus, while a within-household selection method may be desired as a means of minimising coverage error, this is overshadowed by the potential to increase non-response error. Additionally, the accuracy of within-household selection procedures applied to address-based sampling studies has been questioned (Olson, Stange and Smyth, 2014).

## 2.3. Data collection

The 2021 ADII Survey used a sequential mixed-mode (online and hard copy) data collection methodology. Incorporating a push-to-web approach (not offering paper to begin with) was recommended as this has been shown to increase online response, as people are more inclined to complete via paper when offered the choice (Dillman, 2017; Holmberg, Lorenc and Werner, 2010; McMaster et al., 2017; Messer and Dillman, 2011; Millar and Dillman, 2011; Smyth et al., 2010; Tourangeau, 2017). Further, offering respondents mode choice has been shown to reduce overall response (Medway and Fulton, 2012).



### **2.3.1. Approach and reminder details**

Approach and reminder materials used in the 2021 ADII Survey were based on those used in 2020, with some minor modifications.

The sequential mixed-mode approach first involved a mailed invitation pack to invite recipients to complete the survey online, with the aim of maximising responses received via the online mode. Following the initial invitation pack, contact was made via postcard, again inviting participation solely via the online mode of collection (push-to-web).

Once these efforts were maximised, hard copy questionnaires were sent to non-respondents including those who were yet to complete or opt out of the survey. No further reminder activities were employed once the hard copy booklets were mailed, though the online completion option remained available.

#### **Invitation pack mailing**

The invitation pack aimed to provide information on the study and encourage online survey completion. The invitation pack was sent to all  $n=19,999$  sample members. Contents included an invitation letter and a brochure introducing the study, information on how to complete the survey online, referral to the website and Social Research Centre contact details.

The invitation letter and brochure are provided in Appendix 1.

#### **Reminder postcard mailing**

Fourteen days after the mailing of the invitation pack, reminder postcards were sent to remind recipients to go online and complete the survey. Though there was limited time between the invitation mailing and the sample draw for the postcard mailing, some exclusions were made, reducing the sample size for the reminder postcard mailing to  $n=19,827$ . Addresses were excluded from the sample if they had been marked as having already completed, had opted out of the survey, or if the initial mailing had resulted in a 'return to sender'.

The reminder postcard is provided in Appendix 2.

#### **Hard copy questionnaire mailing**

A core feature of the push-to-web design (maximising completion via the online mode prior to introducing a second mode) involves allowing considerable time between sending the invitation mailing and the hard copy questionnaire booklet mailing (in this case, 28 days).

The sample preparation for the questionnaire mailing was conducted as late as possible to maximise exclusions while allowing time for printing. The questionnaire booklet was accompanied by a cover letter and reply-paid envelope and was sent to  $n=18,511$  households.

The cover letter was based on that used in 2020, with minor wording changes and removal of online login details to encourage hardcopy completion, particularly among non-users.

The cover letter is provided in Appendix 3.

## **2.3.2. Additional response maximisation techniques**

### **Providing a range of support services and materials**

Information and support were provided to respondents through a range of mechanisms to alleviate any concerns about the survey bona fides, address queries from sample members and encourage response. In all contact materials, respondents were encouraged to review further information available on the website or in the brochure provided as part of the invitation package. In addition to written materials, the Social Research Centre operated an 1800 helpdesk number and a project-specific email address to respond to individual queries or concerns.

### **Offering contingent incentives**

A \$10 contingent incentive in the form of a gift card was offered for survey completion. For respondents who completed the survey online, these were sent in batches via email at regular intervals following survey returns. For online respondents who did not provide an email address or completed the survey via hard copy, incentives were mailed upon the completion of data collection.

## 3. Questionnaire development

### 3.1. Overview

The 2021 ADII questionnaire was based on that used in 2020, with some sections presented in a different order, some questions included for the first time and some questions deleted.

The questionnaire included the following sections:

- Internet use
- Mobile phones
- Home internet
- Mobile broadband (excluding mobile phone services)
- Mobile phones
- Electronic devices
- Internet skills
- Internet activities
- Demographics.

The median online completion time for the survey was 14.9 minutes.

### 3.2. Online survey instrument

The online survey was programmed and tested in-house by the Social Research Centre. Our online survey software is specifically designed for survey research. It has the functionality to allow for ease of completion on a range of devices, including the ability to identify and tailor design for mobile devices.

Prior to launching the online survey, standard operational testing procedures were applied to ensure that the script truly reflected the agreed final version of the questionnaire. These included:

- programming the skips and sequencing instructions as per the final questionnaire
- rigorous checking of the questionnaire in 'practice mode', including checks of the on-screen presentation of questions and response frames on a range of devices
- randomly allocating dummy data to each field in the questionnaire and examining the resultant frequency counts to check the structural integrity of the script.

The 'live' survey was launched on the 6 April 2021 and was available at [www.srcentre.com.au/aius](http://www.srcentre.com.au/aius). Households were provided with a unique login code to access the survey. For security reasons, unique details were provided only in sealed mailings (not on postcards).

Refer to Appendix 4 for the final online questionnaire.

### **3.3. Hard copy questionnaire**

The hard copy questionnaire was typeset by a professional mail house.

The final booklet was 16 pages in length including a designed cover page, one page of completion instructions and contact details on the inside cover and a half page of return details on the back cover.

Prior to printing, careful testing was undertaken to ensure the contents accurately reflected the source questionnaire and the online programmed version. Additional checks were carried out by the data capture provider to ensure all scanning requirements were met.

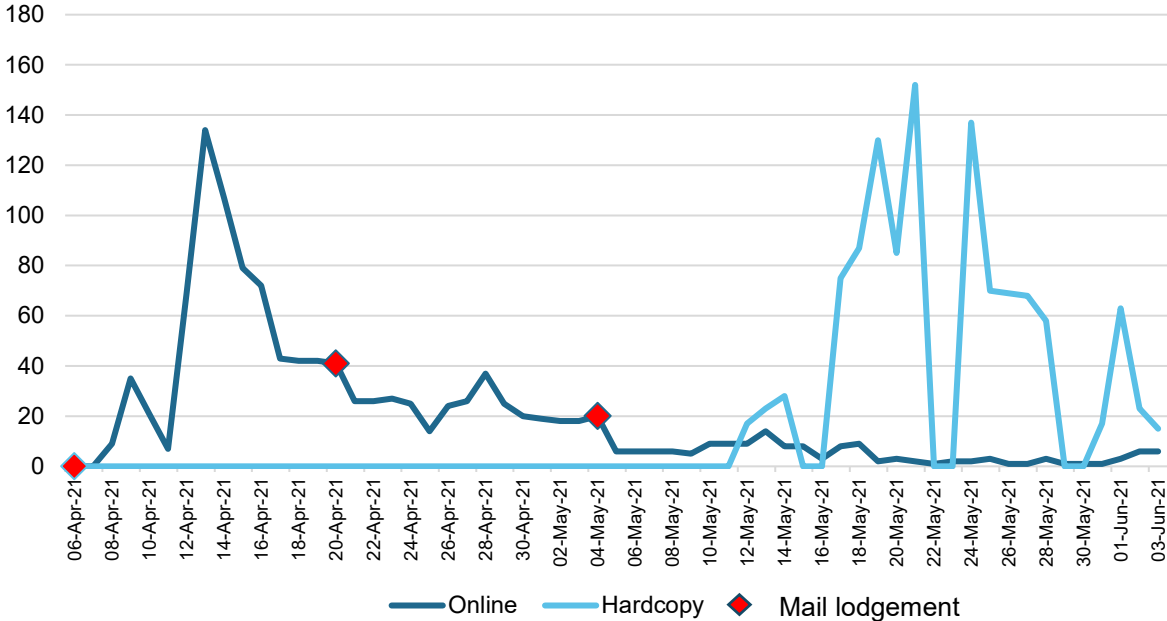
Refer to Appendix 5 for the final hard copy questionnaire.

# 4. Response analysis

## 4.1. Overview

A total of 2,287 completed surveys were received prior to the closing date of 3 June 2021. Of these, 1,170 were completed online, while 1,117 hard copy returns were received. Figure 2 below shows the number of completed online and hardcopy surveys by date.

**Figure 2** Completes by date and mode



## 4.2. Response by completion mode

Table 3 shows the geographical distribution of all respondents by mode of completion. It shows the final sample was distributed roughly in line with the sample drawn, although Rest of QLD, Rest of WA and Rest of NT were slightly underrepresented. There were limited variations by mode.

**Table 3 Respondent location by mode (unweighted)**

Region	Online		Hard copy		Total		Sample drawn	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Greater Sydney	267	22.8	208	18.6	475	20.8	4,377	21.9
Rest of NSW	118	10.1	132	11.8	250	10.9	2,177	10.9
Greater Melbourne	240	20.5	200	17.9	440	19.2	3,682	18.4
Rest of VIC	51	4.4	73	6.5	124	5.4	1,155	5.8
Greater Brisbane	117	10.0	97	8.7	214	9.4	1,800	9.0
Rest of QLD	118	10.1	127	11.4	245	10.7	2,224	11.1
Greater Adelaide	74	6.3	90	8.1	164	7.2	926	4.6
Rest of SA	13	1.1	23	2.1	36	1.6	371	1.9
Greater Perth	86	7.4	96	8.6	182	8.0	1,544	7.7
Rest of WA	23	2.0	24	2.1	47	2.1	650	3.3
Greater Hobart	11	0.9	15	1.3	26	1.1	148	0.7
Rest of TAS	12	1.0	12	1.1	24	1.0	201	1.0
Greater Darwin	11	0.9	4	0.4	15	0.7	208	1.0
Rest of NT	5	0.4	4	0.4	9	0.4	300	1.5
ACT	24	2.1	12	1.1	36	1.6	236	1.2
<b>Total</b>	<b>1,170</b>	<b>100.0</b>	<b>1,117</b>	<b>100.0</b>	<b>2,287</b>	<b>100.0</b>	<b>19,999</b>	<b>100.0</b>

Table 4 overleaf summarises respondent characteristics by completion mode. It shows respondents who completed online were considerably more likely to hold a university degree than those who completed via hard copy. In addition, respondents who completed via hard copy were more likely to be over the age of 55 and use the internet less often (in the last 6 months or more than 6 months ago) or never use the internet compared to those who completed online.

**Table 4 Respondent characteristics by mode (unweighted)**

Respondent characteristics		Respondents					
		Online (n=1,170)		Hard copy (n=1,117)		Total surveys (n=2,287)	
<b>Total</b>		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<b>Gender</b>	Male	618	52.8	479	42.9	1,097	48.0
	Female	540	46.2	621	55.6	1,161	50.8
	Non-binary / Gender fluid	7	0.6	5	0.4	12	0.5
	Different identity	4	0.3	1	0.1	5	0.2
	No response	1	0.1	11	1.0	12	0.5
<b>Age</b>	18 to 24	67	5.7	18	1.6	85	3.7
	25 to 34	170	14.5	73	6.5	243	10.6
	35 to 44	198	16.9	111	9.9	309	13.5
	45 to 54	200	17.1	134	12.0	334	14.6
	55 to 64	207	17.7	219	19.6	426	18.6
	65 to 74	223	19.1	313	28.0	536	23.4
	75 and over	105	9.0	249	22.3	354	15.5
<b>Education</b>	Secondary school or below	216	18.5	387	34.6	603	26.4
	Non-university	290	24.8	278	24.9	568	24.8
	University	630	53.8	393	35.2	1,023	44.7
	Other	29	2.5	43	3.8	72	3.1
	No response	5	0.4	16	1.4	21	0.9
<b>State</b>	NSW	385	32.9	340	30.4	725	31.7
	VIC	291	24.9	273	24.4	564	24.7
	QLD	235	20.1	224	20.1	459	20.1
	SA	87	7.4	113	10.1	200	8.7
	WA	109	9.3	120	10.7	229	10.0
	TAS	23	2.0	27	2.4	50	2.2
	NT	16	1.4	8	0.7	24	1.0
	ACT	24	2.1	12	1.1	36	1.6
<b>Internet use</b>	Used internet in the last week	1,158	99.0	1,001	89.6	2,159	94.4
	Used internet in the last month	6	0.5	11	1.0	17	0.7
	Used internet in the last 3 months	1	0.1	4	0.4	5	0.2
	Used internet in the last 6 months	1	0.1	12	1.1	13	0.6
	Used internet more than 6 months ago	1	0.1	6	0.5	7	0.3
	Never used internet	3	0.3	83	7.4	86	3.8

## 5. Data processing

### 5.1. Hard copy returns processing

Hard copy questionnaire returns were sent to and processed by Datatime, a professional data capture provider using data scanning technology.

#### 5.1.1. Returns management and reporting

Hard copy survey returns were logged by on a daily basis and sorted into the following categories for mid-field status reporting:

- Accepted for processing – some effort had been made to complete the form
- Refusal – includes blank returned forms that were not ‘return to sender’
- Out of scope – includes ‘deceased’
- Sample loss / unusable sample – includes ‘return to sender’.

A log of hard copy and online returns was updated daily to consolidate scanned survey outcomes and online survey completions. This log was used to update the interactive dashboard. The dashboard included the total number of completes for both the online and hard copy surveys, with additional information available for online completes such as demographics and key outcome measures.

#### 5.1.2. Data scanning and capture

Once received by the data capture provider, hard copy returns were scanned and processed using a mixture of optical mark read and key from image technologies.

Fully trained data entry operators reviewed scanned images of the returned survey forms to:

- resolve multiple responses for questions requiring a single response
- verify that responses to multiple response questions were valid
- check ‘blanks’ where the survey sequencing suggested that the respondent should have answered the question
- a double-key and verify process was used to ensure the accuracy of data capture and a batch processing workflow was used to track returns from initial logging to the completion of data capture.

Hard copy forms were securely destroyed following delivery of data and a reference file of scanned images to the Social Research Centre.

### 5.2. Data cleaning rules

Rules used to clean the hard copy and online data to ensure data integrity and logic flow included:

- cleaning the hard copy data based on the online routing
- if multiple responses were given for a single response question, the response was coded in accordance with previously established cleaning rules
- if no answer was provided, the response was coded in accordance with previously established cleaning rules



- 'no response provided' variables were added to all multiple response questions
- if age was blank, records were excluded
- if there were duplicate IDs, the online response was taken over the hard copy response
- if there was a discrepancy in the affordability data (i.e., C6, C7, D6, D7), data was updated based on actual expenditure provided (C7, D7)
- derivation of system missing data at all relevant variables.

For consistency, cleaning conventions followed the rules established from previous iterations of the survey, including the 2020 ADII. Further details of the specific data cleaning rules applied to hard copy data can be found in Appendix 6.

## 5.3. Coding

Verbatim responses were back coded in line with agreed coding rules.

## 5.4. Weighting

### 5.4.1. Overview

Sample surveys are a commonly used tool for making inferences about a population using responses from just a subset of it. To be able to do so, however, requires a probability sample – one in which every element of the population has a known, non-zero chance of selection. Since some units in the population may not have a chance of selection (such as persons whose address is not covered by the GNAF) and there may be different rates of response across unit characteristics, many sample surveys yield subsets that imperfectly cover their target populations despite the best possible sample design and data collection practices (Valliant *et al.*, 2013). In such situations, weighting can reduce the extent of any biases introduced through non-response or non-coverage.

The approach for deriving weights generally consists of the following steps:

1. compute a design weight for each respondent as the inverse of their chance of selection
2. adjust the design weights so they match population distributions across a range of respondent characteristics.

The first step is essential in providing the statistical framework necessary for making population inferences from a sample survey. The second step aims to reduce non-response and non-coverage bias and to ensure that survey estimates are consistent with other sources (such as published results from the ABS Census of Population and Housing).

### 5.4.2. Design weights

The design weights are calculated as the inverse of the probability of selection of each dwelling in the stratum as shown in Table 5:

$$d_h = \frac{1}{p_h} = \frac{N_h}{n_h}.$$

where:

- $N_h$  denotes the number of dwellings in stratum  $h$  according to the 2016 Census<sup>1</sup>,
- $n_h$  denotes the number of respondents (one respondent per dwelling was selected) from stratum  $h$ ,
- And  $h$  denotes the geographic location (or stratum) described in Table 5.

**Table 5 Population benchmarks used for calibration**

Location (Stratum $h$ )	Number of dwellings ( $N_h$ )	Number of respondents ( $n_h$ )	Design weight ( $d_h$ )
Greater Sydney	1,858,607	475	7,900
Rest of New South Wales	1,208,357	250	9,759
Greater Melbourne	1,834,359	440	8,417
Rest of Victoria	691,163	124	11,254
Greater Brisbane	902,891	214	8,518
Rest of Queensland	1,089,722	245	8,980
Greater Adelaide	562,821	164	6,929
Rest of South Australia	204,435	36	11,465
Greater Perth	818,947	182	9,085
Rest of Western Australia	254,616	47	10,938
Greater Hobart	99,247	26	7,707
Rest of Tasmania	143,255	24	12,051
Greater Darwin	55,466	15	7,466
Rest of NT	35,230	9	7,903
Australian Capital Territory	163,539	36	9,172

### 5.4.3. Calibrating to population benchmarks

To ensure estimates made from the dataset are representative of the target population, the design weights were adjusted to match external benchmarks of key demographic parameters likely to be correlated with the survey outcomes and the likelihood of response. For this study, these included gender, age, education and location.

The method for calibrating the design weights was generalised regression weighting which used non-linear optimisation to minimise the distance between the design and calibrated weights subject to the weights meeting the benchmarks. This method requires non-missing data, as such, values for the small number of respondents who did not provide answers to the weighting questions were estimated through statistical imputation.<sup>2</sup>

<sup>1</sup> <https://www.abs.gov.au/census>

<sup>2</sup> Refer to Lumley (2017) for more details on the implementation of regression calibration in R (R Core, 2018) and to Valliant et al. (2013) for a more general treatment of weighting and estimation for sample surveys.

#### 5.4.4. Benchmarks

The external benchmarks that were used for calibrating the design weights are shown in Table 3. These were obtained from 2016 Census data, updated for the December 2019 estimated resident population, sourced from the Australian Bureau of Statistics (2017, 2018). Also included in the table is the proportion of respondents in each category along with the average weight.<sup>3</sup>

**Table 6 Population benchmarks used for calibration**

Benchmark Category	Respondents (%)	Population (%)	Average Weight
<b>Age by highest level of education attained</b>			
18-24 years	3.72	12.06	3.25
25-34 years by Bachelor degree or higher	7.00	7.43	1.06
25-34 years by below Bachelor degree	3.63	11.81	3.25
35-44 years by Bachelor degree or higher	8.83	6.19	0.70
35-44 years by below Bachelor degree	4.68	10.92	2.33
45-54 years by Bachelor degree or higher	8.05	4.24	0.53
45-54 years by below Bachelor degree	6.56	12.04	1.84
55-64 years by Bachelor degree or higher	7.78	3.3	0.42
55-64 years by below Bachelor degree	10.84	11.56	1.07
65+ years by Bachelor degree or higher	12.33	2.78	0.23
65+ years by below Bachelor degree	26.59	17.66	0.66
<b>Gender</b>			
Male	48.62	49.1	1.01
Female	51.38	50.9	0.99
<b>Part of State</b>			
Capital city	67.86	66.9	0.99
Rest of state	32.14	33.1	1.03
<b>State</b>			
New South Wales	31.7	31.97	1.01
Victoria	24.66	26.21	1.06
Queensland	20.07	19.83	0.99
South Australia	8.75	7.01	0.80
Western Australia	10.01	10.23	1.02
Tasmania	2.19	2.14	0.98
Northern Territory	1.05	0.93	0.89
Australian Capital Territory	1.57	1.68	1.07

#### 5.4.5. Treatment of missing values

The regression weighting approach used to adjust the base weights requires that there are no missing values across the adjustment variables or values other than those for which there are reliable benchmarks.

<sup>3</sup> Where the responding and population proportions are similar, the average weight will be closer to one.

A statistical model (Stekhoven and Buehlmann, 2012) was applied to each item with missing values to impute the most likely value for a respondent, conditional upon their other responses. Given the very low prevalence of missing values overall (a maximum of 0.75% for any item), the imputation process is expected to have a negligible impact on weighted estimates made from the dataset.

**Table 7**      **Extent of missing values among weighting characteristics**

Questionnaire item	Not stated*		Related weighting characteristic(s)
	#	%	
bdage (Age)	0	0	Age by Education
bducat (Education)	21	0.92	Age by Education
bdgender	29	1.27	Gender
Sample Variables	0	0	State Part of State

\*Not stated consists of *Don't know*, *Refused* or *Other* responses.

## 6. References

- Battaglia, M., Link, MW, Frankel, MR, Osborn, L., and Mokdad, AH (2008), 'An evaluation of respondent selection methods for household mail surveys', *Public Opinion Quarterly*, vol. 72, no. 3, pp. 459-469.
- Dillman, DA, 2017, 'The promise and challenge of pushing respondents to the web in mixed-mode surveys', *Survey Methodology*, vol. 43, no. 1, pp. 3-30.
- Holmberg, A, Lorenc, B and Werner, P, 2010, 'Contact strategies to improve participation via the web in a mixed-mode mail and web survey', *Journal of Official Statistics*, vol. 26, no. 3, pp. 465-480.
- Lumley, T. (2017) survey: analysis of complex survey samples. R package version 3.32.  
<https://CRAN.R-project.org/package=survey>.
- McMaster, HS, LeardMann, CA, Speigle, S, and Dillman, DA, 2017, 'An experimental comparison of web-push vs. paper-only survey procedures for conducting an in-depth health survey of military spouses', *BMC Medical Research Methodology*, vol. 17, p. 73.
- Medway, RL and Fulton, J, 2012, 'When more gets you less: a meta-analysis of concurrent web options on mail survey response rates', *Public Opinion Quarterly*, vol. 76, no. 4, pp. 733-746.
- Messer, BL and Dillman, DA, 2011, 'Surveying the general public over the internet using address-based sampling and mail contact procedures', *Public Opinion Quarterly*, vol. 75, no. 3, pp. 429-457.
- Millar, MM and Dillman, DA, 2011, 'Improving response to web and mixed-mode surveys', *Public Opinion Quarterly*, vol. 75, no. 2, pp. 249-269.
- Olson, K, Stange, M and Smyth, J, 2014, 'Assessing within-household selection methods in household surveys', *Public Opinion Quarterly*, vol. 78, no. 3, pp. 656-678.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Smyth, JD, Dillman, DA, Christian, LM, O'Neill, AC, 2010, 'Using the internet to survey small towns and communities: limitations and possibilities in the early 21st century', *American Behavioral Scientist*, vol. 53, no. 9, pp. 1423-1448.
- Stekhoven, D.J. and Buehlmann, P. (2012), 'MissForest - nonparametric missing value imputation for mixed-type data', *Bioinformatics*, 28(1) 2012, 112-118, doi: 10.1093/bioinformatics/btr597
- Tourangeau, R, 2017, 'Mixing modes: tradeoffs among coverage, nonresponse, and measurement error', in PP Biemer, E de Leeuw, S Eckman, B Edwards, F Kreuter, LE Lyberg, N Clyde Tucker and BT West (eds.), *Total Survey Error in Practice*, Hoboken, NJ, USA: Wiley, pp. 115-132.
- Valliant, R., J. Dever, and F. Kreuter (2013). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.
- Vanderhoeft, C. (2001) Generalized Calibration at Statistics Belgium. Statistics Belgium Working Paper No 3.

# Appendix 1      Invitation letter and brochure

# Appendix 2      Reminder postcard

# Appendix 3      Reminder letter



## **Appendix 4      Final online questionnaire**

## **Appendix 5      Final hard copy questionnaire**

## Appendix 6 Data cleaning rules