Australian Government
Department of the Prime Minister and Cabinet

BETA

BETA's Pragmatic and Better Practice Survey Guide

December 2024

Other uses

Enquiries regarding this license and any other use of this document are welcome at:

Managing Director
Behavioural Economics Team of the Australian Government
Department of the Prime Minister and Cabinet
1 National Circuit
Barton ACT 2600
Email: beta@pmc.gov.au

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Department of the Prime Minister and Cabinet or the Australian Government.

Research team

Su Mon Kyaw-Myint and Bethany Jones were the principal authors of this guide. Many current and former staff of BETA also contributed by reviewing or drafting internal survey guidance.

Acknowledgments

Thank you to Bradley Carron-Arthur and Phoebe Martin of the Australian Public Service Commission for the review of this guide and for your helpful comments.

Contents

Introduction

The Australian Public Service (APS), is increasingly using survey research to inform the design of programs, to understand a policy issue or as part of a consultation process. The type of survey we conduct varies, largely depending on the time and resources we can dedicate to survey development and pre-testing. While there are many textbooks available on 'best practice' surveys, little information exists on how to develop surveys in the middle of the survey quality continuum (see Figure 1). This guide was developed by the Behavioural Economics Team of the Australian Government (BETA) to aid APS officers when developing these types of surveys.

It is rare to have access to the time and resources required to develop a 'best practice' survey. These are typically conducted by institutions like the Australian Bureau of Statistics (ABS). They can take years to develop and can cost many millions of dollars. While these types of surveys are usually out of reach for the rest of the APS and many researchers, it is possible to design surveys to meet a minimum quality standard ('Pragmatic' surveys) and where feasible, aim for a higher quality ('Better Practice' surveys). This guide contains advice and steps to achieve 'Pragmatic' and 'Better Practice' surveys while also giving you some understanding of what developing a best practice survey may involve (e.g. pilot testing, random sampling with low survey error and high response rates).

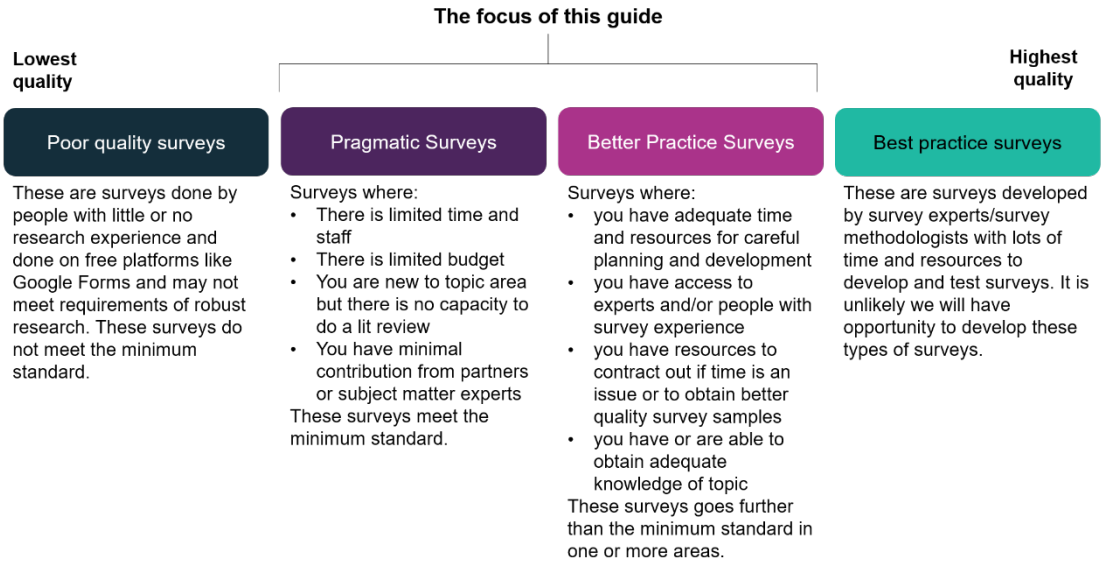

Figure 1 **- Survey quality continuum**

Aims of this guide

This survey guide aims to:

- increase your understanding of key concepts in survey research
- assist you in developing better surveys; and,
- create a shared approach to survey development across teams and agencies.

What's in this guide

The guide provides:

- protocols and templates for survey development, pretesting, implementation and analysis;
- advice to avoid common issues and errors in surveys; and,
- guidance on developing 'Better Practice' and 'Pragmatic' surveys.

While survey methods articles are cited in this guide, this guide is not intended to be a comprehensive review of the literature on this topic.

How to use this guide

This guide is best used as an electronic format as cross-references between sections are provided throughout the guide.

The guide starts with an essential question – Is a survey the right research method for your project? It explains how to answer this question for your project, and alternatives for projects where a survey is not suitable.

If you find that a survey is appropriate, you can go on to read sections on the basics of survey design, sampling approaches, survey modes, what you can do to survey response rates and reduce survey errors.

When developing a 'Pragmatic' survey we recommend that you at least read the section on Step-by-step guide to survey design and navigate to key parts of the guide as needed (see Figure 2). The step-by-step guide contains notes for developing a 'Pragmatic' survey and briefly highlights the limitations of which you need to be aware.

If you are unfamiliar with survey development, we suggest you read the full section of Survey Design 101. Of particular importance are:

- Planning and Conceptualising;
- Writing quality survey questions;
- Pre-testing the survey and
- Ethics and privacy (Step 5 & Step 11).

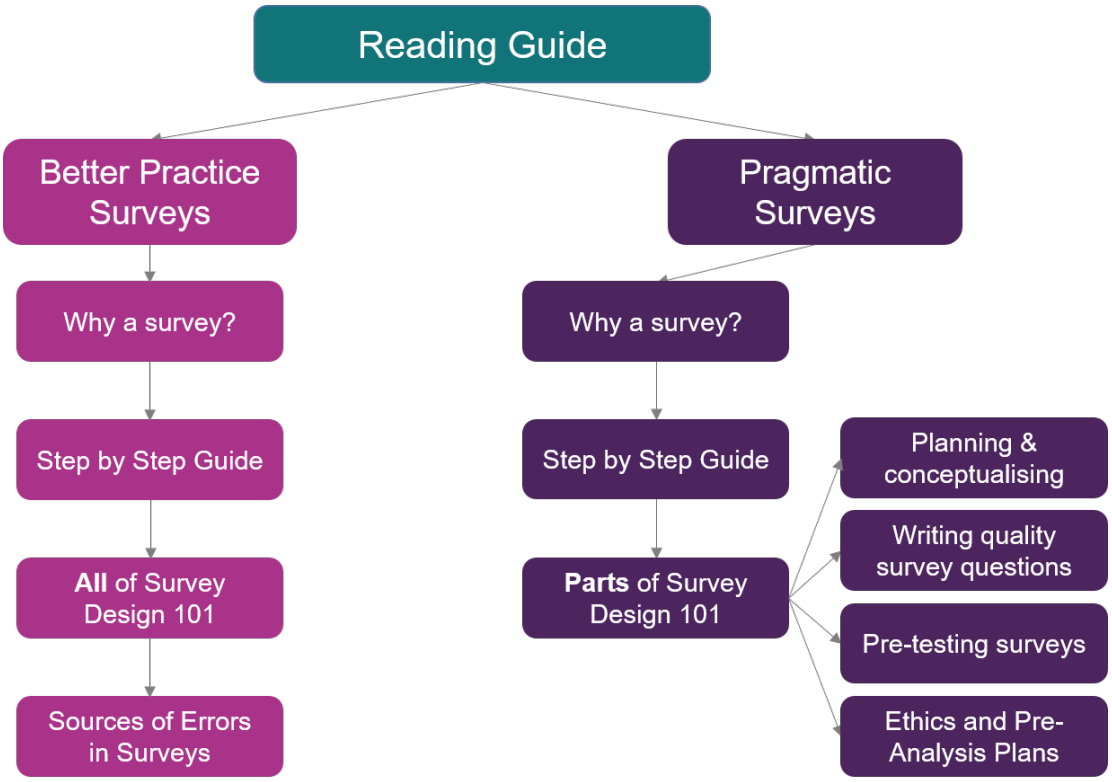

Figure 2 **- Reading Guide**

Why a survey?

Before commencing any work on a survey, there are two key questions to ask:

1 "Are there any existing data sources?"

2 "Is a survey the right method for the research question?"

Are there existing data sources?

Compared to even a decade ago, it's now quite easy and inexpensive to conduct surveys. However, before starting a new survey, you should explore existing datasets that you could use to see if the information you are after is already available. This will not only save time but will also help to reduce respondent burden.

The Australian Bureau of Statistics website has many data sources that are available through downloadable data tables or through Table Builder. There are also datasets that require an application for access, especially if you need access to microdata (unit record data). The page on ABS data services may be helpful.

Other data sources that may be helpful include:

- Household Income and Labour Dynamics in Australia (HILDA) Survey
- Administrative data sets (e.g. Person Level Integrated Data Asset or PLIDA[1]), and
- Publicly available datasets on data.gov.au.

Is a survey the right research method?

Surveys, qualitative interviews and focus groups all collect information and they all involve asking questions of individuals. Surveys are administered to a group of respondents and the same questions are asked consistently across the group. Interview and focus groups provide more in-depth, qualitative data. They are better at capturing data from open-ended questions and allow for a more unstructured or flexible approach. Interviews and focus groups also provide information such as body language, facial expressions and other cues, on top of verbal answers to questions. However, interviews and focus groups usually require more time to administer and analyse compared to surveys. Interviews and focus groups involve much smaller sample sizes and often, sample size is tentative at the beginning, guided by when saturation point is reached. In contrast, surveys require good estimates of the sample size needed, to enable specific statistical analyses or to ensure generalisability.

Another key difference between surveys and interviews or focus groups is that they generally answer different types of questions. All research falls somewhere within an exploratory to confirmatory spectrum. In exploratory research, you have no specific hypothesis or theory. In confirmatory research, you are trying to confirm a theory or a hypothesis. Often, interviews

[1]PLIDA is very time and resource intensive to get up and use. It can take many months, especially if you are unfamiliar with it and analysis of big administrative data. There are other limitations such as limited information on attitudes and beliefs.

and focus groups are more exploratory and surveys are more confirmatory. Table 1 contains a checklist to help you decide your research method. Please note this is a simplified check list and there may be circumstances where you could still do the research even if it is no in this checklist. The table does not provide nuanced answers often needed in research.

Table 1 **- Check which research method you need**

Purpose and characteristics of your research	Surveys (Quantitative research)	Interviews/ Focus Groups (Qualitative research)
Testing hypotheses.	Yes	No
Exploring data. Capturing and discovering meanings from the data.	No	Yes
Confirm or test something.	Yes	No
Understand concepts, thoughts.	No	Yes
Understand experiences, attitudes and beliefs	Yes	Yes
Concepts are in the forms of variables or scales.	Yes	No
Concepts are in the forms of themes and generalisations.	No	Yes
Data in numbers.	Yes	No
Data in words	No	Yes
The research can be replicated and procedures are standard.	Yes	No
Flexibility to adapt and change questions.	No	Yes
Compare answers from different groups (e.g. cross-tabs)	Yes	No
Generalizable data that can be compared between studies using the same survey instruments.	Yes	No
Capture nuances.	No	Yes
Analysis using statistics, tables or charts and how these relate to your hypotheses.	Yes	No
Analysis by extracting themes from the data and categorising and interpreting. Data is presented in a story or narrative rather than in tables or statistics.	No	Yes
Sensitive topics that participants may want to answer anonymously.	Yes	No

When developing major surveys it is common to conduct interviews and focus groups to help develop survey questions. The reverse may also be true. Sometimes it is helpful to gather quantitative data with surveys first then probe and explore further by following the survey with interviews or other qualitative research. This is especially useful in Randomised Controlled Trials (RCTs) or survey experiments.

When undertaking a mixed-method approach to research, it is important to consider how a survey is complementing other research and whether different research components need to occur in a specific sequence or whether they can be done concurrently. For example, let's say you are developing a survey about career change intentions in those over 45 years old. You could run a survey first to get a lot of data from a sample of 2000 people about what may influence career change intentions and decisions in this age group. However, while the survey has free text responses, many respondents tend to skip them or provide minimal input. So, you may want to ask people completing the survey whether they would be available for an in-depth interview for additional questions raised following the analysis of survey data. For more information on types of mixed methods research, see (Bryman 2006) or (Tashakkori et al. 2021).

What research questions and context are appropriate for survey research?

Surveys are a good tool for obtaining information about people's beliefs, opinions, intentions, self-reported behaviours and characteristics (Neuman 2007). When constructed appropriately, they can also provide information on identity, awareness and knowledge. Examples of questions suitable for surveys are provided below.

Self-reported behaviours: How often do you practice violin? Did you vote in the last election? In the last 12 months, how many times did you consult a GP (in person/telehealth)?

Attitudes/beliefs/opinions/preferences/awareness: To what extent do you agree or disagree with this statement: I feel secure in my job. What is most important to you in a career? What are the reasons for undertaking this field of study? How satisfied are you with your current job? How would you like to receive information from the Government to prepare for ageing? In general, how would you rate your experience with this service?

Characteristics/attributes: Are you married, never married, single, divorced? What is the highest level of education you have completed? What age are your children?

Expectations: Are you planning to have a child in the next 3 years? Will you be voting in the upcoming NSW election? Do you think the size of your industry will stay the same, grow or shrink in the next five years? In the next three months, how likely are you to take action to prepare for later life?

Self-classification: Do you consider yourself to be middle class, upper class or working class? Would you say you are religious or not religious? Do you identify as Aboriginal and/or Torres Strait Islander?

Knowledge/skills: What is the population of Canberra? What proportion of the Australian population are first generation migrants? Is it legal to carry fruits into Victoria from another state/territory? How would you rate your computer coding skills? Note that in many cases, these items measure *perceived* knowledge/skills.

What questions are not ideal for surveys?

While we can ask respondents to choose from a list of potential reasons for a belief or behaviour, it may not be possible to elicit in-depth or nuanced answers in a survey. Moreover, cross-sectional surveys are not usually suitable for causal analysis such as why a person holds a particular attitude, or what led to a specific behaviour. It is difficult to capture 'why' in a few survey questions and interviews and focus groups may be better suited to add in-depth information.

Survey questions are self-reported so they may also not be appropriate if you need to determine actual behaviours (rather than intention or self-reported behaviours). And while perception of knowledge can easily be obtained using a survey, it may not reflect actual knowledge, especially when respondents are asked questions such as 'what is your understanding of privacy legislation?' A person might rate themselves as having a good understanding of privacy legislation when in reality they do not. Survey questions must be carefully designed to test the respondents' actual knowledge of a topic, as in a multiple choice quiz in exams.

Advantages and limitations of survey research

Advantages

Surveys offer many advantages over other methods of research.

- They can capture a large amount of data from many people efficiently and at a low cost.
- The information is captured in a format that is can be analysed using a range of tools including Excel, R, SPSS and Stata, particularly for online or computer assisted surveys (CATI, CAPI).
- Surveys (if sampled and/or weighted appropriately) can allow you to generalise information to a larger population and are a useful tool for obtaining representative estimates (e.g. health conditions prevalence or prevalence of mental health stigma). This is not possible with qualitative research.
- Surveys allow comparison of different subgroups which are often of interest (e.g. gender, age, SES etc.).
- Data can also be pooled across different surveys. This is possible when identical questions are used across different surveys. This allows comparisons with other national or international data.
- Surveys can be longitudinal and these longitudinal surveys provide rich data on target cohorts over time. Examples are the Household Income and Labour Dynamics in Australia (HILDA) survey and the Longitudinal Study of Australian Children (LSAC) both funded by the Department of Social Services.

Limitations

However, there are also limitations of survey research.

- Only captures data for which respondents can provide accurate answers.
- Surveys are of limited use when asking about others' behaviours.

- Surveys do not allow researchers to probe for and obtain additional information or clarification. This can lead to information that is unclear, especially if questions are not designed appropriately.
- While open ended questions can be asked in surveys, most respondents skip these questions or rarely provide a long and detailed answer. Free-text responses are also time and resource intensive in terms of analysis (Note: the use of AI is making this less intensive but may still need human oversight).
- Surveys may miss specific groups such as those with limited English, certain socio-economic groups who are less likely to respond to online surveys or hard-to-reach groups like older and younger people, those living in residential care or other facilities (See the Section on Sampling approaches and Sources of errors in surveys).
- Similar to above, survey samples can be biased (different to estimator bias) and are often not representative of the target population. This is especially true for surveys using non-probability samples. For survey sample bias and survey errors, see sections on Sampling approaches and Sources of errors in surveys.

Step-by-step guide to survey design

The process of designing a survey can be broken down into a series of steps (Figure 3). Steps with an asterisk may be able to be simplified, modified or skipped.

Figure 3 **- Diagram of steps in survey development for 'Better Practice' and 'Pragmatic' surveys**

Each step in Figure 3 is expanded below with minimum standards and links to relevant sections.

Step 1. Develop purpose/research aim, hypotheses, and research questions (See Planning and Conceptualising section)

Recording these aims and questions will be important when developing an ethics application, drafting the survey, and for business continuity.

Step 2. Develop a conceptual framework

Take time to think through what you want to measure and how it will be used in analysis. If there is a relevant theoretical model in the existing literature, it could be a helpful starting point when planning a survey.

For 'Pragmatic' surveys this could be as simple as predicting relationships between variables, rather than a formal theoretical model. This step may be skipped if the research is exploratory (e.g. diagnostic phase).

Step 3. Understand the target respondents

When thinking about who to survey, it can be helpful to note what is known about the target audience, and what is assumed. This can include things like literacy, interests, values, and familiarity with survey topic. This may help define the survey sample. For further information on the sampling approach and mode(s) of survey, see section on Sampling approaches and Survey modes.

'Pragmatic' surveys typically use online non-probability samples sourced through online panels. Generalisability of the findings can depend on the quality of the survey sample and survey errors including the response rate. Some online panels are better than others. For more information see the Non-probability sampling section.

Step 4. Procure supplier for survey sample and/or cognitive interviews

This step is only relevant when using a third party to recruit participants. When sourcing your survey sample via a third party, combining recruitment for participants for cognitive interviews with the recruitment for survey participants can be an efficient use of time.

Step 5. Start thinking about ethics issues

Most surveys need some level of ethics review. Reviewers need a lot of detail about the design and materials used in research to review and provide advice. Thinking about what ethical issues might arise early in the process can help with a smooth ethics process. At this stage we recommend considering some key questions:

- Considering what will be included in the survey. Will any items be potentially sensitive for the planned participant group?
- What sampling method and population will you target for recruitment?
- How will you approach participants? Email? Online? In person? Will you use reminders?

- Who will have access to the data? Will you share with other organisations or even make it public?
- How will you get consent from participants? How can you make sure they understand the risks and benefits?

Check point 1. Check with an ethics adviser for advice on ethical aspects of the research.

Step 6. Sourcing and/or drafting questions and response options (Also see What is a quality survey question?)

This could include:

- Using and adapting existing measures. This is the preferred approach.
- Developing new questions.
    - Brainstorm, include examples of good and bad questions and think about what makes them good or bad.
    - Think about constructs. Do you want open ended or closed questions? Go back to the literature on similar theoretical concepts.
    - Seek initial feedback/review from colleagues and subject matter experts.
- Building the first draft of the survey. Consider ordering, formatting and skips.

Considering forced answers vs voluntary answers (Sischka et al. 2022)

It may be necessary to develop questions very quickly when conducting a 'Pragmatic' survey. Be aware that this can result in a higher degree of measurement error. Sourcing existing questions where possible can reduce this risk because these questions have already been tested and validated by researchers. However, it is important to retain the existing questions as they appeared in the validated survey. Changes in words that appear minor can have a large impact on responses in ways that are hard to predict.

Step 7. Prepare survey approach materials

This will include:

- first invitation and reminders;
- privacy notice and consent forms;
- advertisements (e.g. online/ flyers/ posters); and,
- frequently asked questions and similar staff or participant resources.

At this stage it is also helpful to specify the mode and timing for distribution.

Check point 2: This is a good point to consult with leaders and/or survey experts. This is particularly helpful for new survey developers or those writing a survey on an unfamiliar topic.

Step 8. Conduct cognitive interviews (Also see Pre-testing the survey)

Ideally cognitive interviews involve the same target population as the planned respondents for the survey. It may be possible to combine recruitment of participants for cognitive interview with recruitment of survey participants when using a third-party recruiter. It can also

be helpful to include expert reviews of possible questions if there are subject matter experts available.

When undertaking a 'Pragmatic' survey, it may not be possible to conduct formal cognitive interviews. Instead, ask some colleagues who have not been involved in survey development to volunteer for interviews or ask for volunteers from friends and family or from your professional networks.

Step 9. Revise the survey based on cognitive interview/expert review

This may be an iterative process, depending on the amount of time available for survey development.

Step 10. Prepare a Survey analysis plan (see Pre-analysis Plan (PAP))

For 'Pragmatic' surveys this could be as simple as a few dot points. A survey PAP is recommended but not mandatory. It can help avoid p-hacking and encourages you think deeply about analyses, which allows you to write better survey questions.

Check point 3: Seek input from technical specialists on the draft survey analysis plan

Step 11. Seek ethics approval (see Ethics and privacy).

Most aspects of the research must be finalised prior to ethics review. However, some survey questions that pose no risk to participants can be indicative. For example, if you wanted to change the age bins from 10 or 5 years after review, that would not need to go back to the ethics review body. However, if you wanted to include new questions, or change wording of questions that could affect the risk profile then that would need to be reviewed.

This step can also be completed earlier if the materials have been developed.

Step 12. Program the survey

This step is not necessary when the survey is being hosted by a third party.

Step 13. Test the survey before soft launch

Whether the survey is self-hosted, or hosted by a third-party organisation, it is important to test it prior to launch. Ask colleagues to preview and test the survey. They should especially focus on the flow of the survey (e.g. skips and logic), typographical or other errors, and presentation. This is an important step and should not be skipped.

This can also be a good opportunity to plan the analyses in the soft launch phase.

Step 14. Soft launch the survey and review the first few hundred responses

The sample size for the soft launch will depend on the total survey sample and what is manageable. Generally, BETA recommends at least 100 responses to ensure that problems with survey flow or the interpretation of questions may become apparent.

Typically, the recruiter will pause after the agreed soft launch responses have been completed and analysts will have a day or so to review the responses. The timeframe for analysis is typically extremely compressed, so identifying priorities for testing is essential.

Occasionally major changes may be needed after review of soft launch data. For example, some logic may not work as expected meaning a subset of respondents do not see the correct questions. Under these circumstances it may be necessary to discard soft launch data and extend the planned recruitment to replace the lost data.

Step 15. Distribute the finalised survey as planned

Step 16. Monitor response inflows

While often the recruitment plan specifies quotas similar to the Australian population demographics, some groups can be hard to reach. Under these circumstances it may be necessary to relax some quotas (if there are quotas) or extend the time the survey is in field.

Step 17. Analyse and store data (see Data storage)

Ensure that data storage and access are consistent with the commitments in the ethics application. Ensure analyses are conducted consistent with the PAP or conducted thoughtfully rather than many crosstabs and significance testing. Even best practice surveys have limitations – take time to understand the limitations of the survey and how to communicate them.

Survey Design 101

Answering questions: What's involved?

This section is useful for Step 3 of the Step-by-step guide.

Understanding how respondents think about questions and answer survey questions is essential to developing survey questions that capture good quality data. While there are several theories exploring how people answer questions, Tourangeau and co-authors summarised that there are essentially four components (Tourangeau et al. 2000):

- Comprehension
- Retrieval
- Judgement and
- Response.

Issues in any of these four components of survey response can lead to measurement error (further discussed in the *Survey errors* Section). These are illustrated below using this example question:

"How many times have you used International Money Transfer services in the last two years?"

Comprehension

To answer a question, respondents must decide on the main idea behind the question and what response is required (Robinson and Leonard 2019). Comprehension problems occur when respondents do not read the question and instructions, misinterpret them or when unfamiliar words are used. This can result in respondents interpreting the question differently to what was intended or find the question too difficult or complex to answer. Often this leads to respondents skipping the question or providing an inaccurate answer.

In the example question '*How many times have you used International Money Transfer services in the last two years?'* there are potential issues with comprehension. For example, respondents may not know what is meant by 'international money transfer services'. Defining the term or providing some examples of 'international money transfer services' can help with comprehension.

Retrieval

The second component in answering questions is retrieval. This relates to how the respondent recalls appropriate information from long-term memory. There are individual differences in how memory is recalled, and respondents may use different recall strategies to remember. Recall is affected by:

- how the question is worded;
- how particular memories are encoded;

- retrieval cues provided in the question (e.g. think about the last time you were in a cinema);
- the significance of the event you are asking about; and,
- the length of time between the event and the survey.

In the example question, the time used was 2 years. This is too long for most people to accurately remember how many times they have used international money transfer services. Generally, the longer the time, the less likely a person is to be able to accurately retrieve and recall (Tourangeau et al. 2000).

Judgement

Judgment is about formulating an answer. When a respondent retrieves information to answer a survey question, they assess how accurate their retrieval is. They need to determine if they need to make inferences if they are unsure about the accuracy of their retrieval (Robinson and Leonard 2019). They also need to decide if some information is more important than others and decide whether some recalled information should be omitted (Dillman et al. 2014).

There are three types of judgement according to Tourangeau and co-authors (2000):

- Judgement about dates and durations (e.g. how much time off did you need from work due to your illness, on what date did the event happen, how many hours did you usually work per week);
- Judgement about frequencies (e.g. how many times did you visit your doctor); and,
- Judgement about attitude questions (e.g. should students be allowed to use ChatGPT for their assignments?).

In the example question *'How many times have you used International Money Transfer services in the last two years?'* a respondent may recall transferring money four times but they would have to assess how accurate that figure is. They also need to assess whether those four times actually fall within the time frame in the question.

Response

The final component, response, involves selecting an answer to the question based on the first three components, and providing an answer in the survey (Robinson and Leonard 2019). Again, it's not a simple process. It can be affected by the type of response categories presented or how the question stem (the core of the question) is worded. For example, presenting high frequency response options (e.g. for vegetable consumption, high frequency might be 40-50 serves a week) indicates to the respondent that most people do this particular action frequently. This could cause them to adjust their answer from their original retrieval and judgement. In this way, response categories act as social norms or priming when providing an answer. For more information about selecting response options see What is a quality survey question?

Consider the example question '*How many times have you used International Money Transfer services in the last two years?'* The respondent recalled transferring money four times but wasn't too sure. The response categories which are 0, 1-5 times, 6-10 times, 11-20 times, 21 to 30 times, 31 to 40 times. Based on the categories, they may revise their original estimate and choose the 6-10 times category instead.

The four components of answering questions do not necessarily happen in order. This is why cognitive testing of draft survey questions is extremely helpful for insights into the processes and cues involved in answering a question. Ideally, questions encourage respondents to undertake these four steps whenever answering a question. For further information on Cognitive interviewing and templates for cognitive interviews see Pre-testing the survey.

In some cases, respondents may take a shortcut and make a guess rather than going through all four steps. This is known as satisficing. Satisficing is a portmanteau of “satisfy” and “suffice”. It means that people use the first possible answer without the extra cognitive load involved in assessing whether another answer might be more accurate or appropriate. Alternatively, some respondents leave the question blank or make up answers (Dillman et al. 2014). We want to avoid these scenarios in our surveys to maximise data quality.

Writing quality survey questions

A quality survey starts with having a clear idea of the purpose of the survey and having a conceptual model, even a simple one. A quality survey also includes survey questions that reflect an understanding of how people answer survey questions, and careful ordering of questions.

Planning and Conceptualising

This section relates to Steps 1 and 2 of the Step-by-step guide to survey development.

One of the most common challenges in survey design is having an unclear or unarticulated purpose. This can give rise to several design flaws (Robinson and Leonard 2019) and therefore the planning and conceptualising step is critical to developing quality surveys.

The planning stage involves thinking about and writing down the following:

- What is the purpose of the survey?
    - Having a clear purpose will help answer questions about the length of the survey, mode of the survey and its format.
- What are the research questions?
    - Research questions are broad questions that help drive and focus a research project. They allow you to identify concepts to measure and relationships between variables.
    - These research questions should not be the actual questions in the survey. Typically, they are too broad for respondents to provide accurate responses.

The conceptualisation stage allows the translation of research questions into survey questions. A literature review can help discover relevant concepts and gaps as a starting point. Even a quick review of the literature and existing surveys will save time down the track and help ensure that the data is robust and measures the intended constructs.

A literature review can also reveal conceptual models or theories. These can help determine other key variables that should be included. In this way it is helpful to have a conceptual model, no matter how rough or simple.

Example

Let's say you want to measure people's attitudes on climate change and what factors contribute to these attitudes. First, you would need to define concepts of what you mean by attitudes? Are they attitudes about a particular policy? Is it about how much they are likely to give/do for reducing emissions? Is it about changes in particular industries? Here the attitudes would be the outcome measures. If your research questions also include how attitudes lead to behaviour, then you would also have behaviour as the distal outcome measure

Sometimes a visual conceptual model can help clarify the relationship between variables, such as constructs and outcome measures and how they relate to each other. It can be helpful to have this to hand (along with purpose and research questions) while drafting the survey. For each new question, ask where it fits into the model and why it is important. If the answer is that it is just 'interesting to know', the question can be dropped (Bradburn et al. 2004).

Sometimes a conceptual model is less helpful, such as exploratory work. Regardless, it is good practice to document why each question is included. This can aid analysis and ensure that the survey is as short as possible with only relevant questions. Things to document at this stage of survey development (in addition to purpose, background, conceptual model) are:

- How will the survey data help answer the research questions and how it links to other research efforts for the project? (if you are also doing other research)
- Who will use the information obtained from the survey?
- How the information will be used? (e.g. is it to inform qualitative research, it is for designing an intervention, is it to evaluate a program?)
- What format the data will need to be to maximise its usefulness? (to determine response categories; e.g. do you want a frequency scale, do you want a neutral option)
- Information on sampling approach & sample size, and
- Sources of questions if obtained from existing surveys.

The next section will focus on developing good quality questions.

What is a quality survey question?

> *"A good question is one that yields a truthful, accurate answer" (Bradman et al. 2004 p. 325)*

This section relates to Step 6 of the Step-by step guide.

A good quality survey consists of quality survey questions, conceptual foundations, and considered format and design. Survey questions that produce inaccurate responses increase measurement error (for more information, see the section on Survey errors).

Components of a survey question

Survey questions are made up of a question stem and/or response options. The words used in the question stem and the relationship between the question stem and response categories influence survey responses (Robinson and Leonard 2019). Further information about wording influences is available in this seminal paper on Experiments in Wording Questions II (Rugg 1941) and the paper on Words matter: the importance of issue framing and the case of affordable housing (Goetz 2008).

The section above Answering questions: What's involved? covered information about how people respond to survey questions. Key considerations include:

- whether the respondent can understand the question;
- whether words and phrases are likely to be interpreted the way you intend; and,
- reducing cognitive load as much as possible and providing reference periods that are appropriate to aid recall.

A **quality survey question** has the following characteristics:

- Asks for only one answer on only one dimension. This means avoiding double-barrelled questions (e.g. How often do you log in to My Aged Care and Medicare? This is asking about two separate services in one question)
- Response categories are exhaustive. Respondents may become confused if an appropriate response for their situation is not available. It can be difficult to know what all the possible answers may include. Therefore, including an 'Other – please specify' option may a great idea to allow respondents to answer accurately
- Response categories are mutually exclusive. Overlapping categories make the data difficult to use and interpret. It also makes it harder for the respondent to pick an answer.
- Produces variability in response. Questions in which everyone answers the same are not useful in analysis. For example questions with an obvious correct answer will show low variability such as "I am good at my job: Yes/No".
- Follows the previous question logically. This can include grouping similar topics together, going from generic to detail or detail to generic ordering, and labelling or signalling new section if the topic changes. For example, grouping all demographic questions together allows respondents to calibrate their thinking to their personal circumstances.
- Reliable and valid. Reliable items provide a consistent measures in a comparable situation so it's about 'repeatability'. Valid questions are those that measure the intended construct. See glossary for more information.
- The question should have the same meaning for all respondents. This is why pre-testing the survey with people who share similar characteristics as your survey sample is very useful.
- Avoids social desirability bias.

Box 1 provides examples of things to avoid when drafting a survey question.

Box 1: Common issues in survey questions

Avoid using qualifiers in questions

I find online banking very frustrating

Using the qualifier ‘very’ in this question means the response becomes extremely difficult to frame. What if the respondent finds it only slightly frustrating? Using neutral language in question stems allows the strength of the response to be encoded in the response frame.

Avoid ambiguity and confusion

What is your income? Do you exercise regularly?

This may sound simple but is it referring to weekly or annual income? Is it gross or net income? Does it mean personal income from all sources or from wage and salary only? Could it be household income?

What does exercise mean here? Formal exercise or informal? What does regularly mean? For one person this could mean daily and for others it could mean once every two weeks. What time frame?

Avoid double barrelled questions

Do you enjoy playing volleyball and basketball?

Make each question only about one idea or topic. In this example should people answer in the affirmative only they like both types of sports? What if they like volleyball but not basketball? How should they answer this?

Avoid leading questions

Leading or loaded questions can bias a respondent to answer in specific way. So phrase questions in a neutral format.

Should we spend more taxpayer money to ensure income recipients are not living in poverty?

This is a leading question and people are more likely to say yes or agree with this question. Or they respond with anger or irritation leading them to question the value of answering the survey.

Avoid questions that are difficult to answer/too cognitively taxing

How much did you spend on your healthcare in the past year?

This question is difficult to answer for people who don’t’ regularly track spending or people who may have periodic expenses. It would also be helpful to define “healthcare” and whether they should report out-of-pocket costs only or total cost. Often you can make a question less cognitively taxing by breaking it down into a couple of questions.

Avoid mismatched question stem and response categories

Have you heard of in-home aged care services?

Strongly agree; agree; disagree; strongly disagree

Here the responses that respondent have to choose from do not match the actual question rendering it difficult to answer.

Do not use poorly defined terms

Do you oppose or favour problem gambling legislation?

This question assumes knowledge of the legislation, or requires people to interpret what it might be. Is it banning ads for online bets during specific hours and programs? Is it reducing or removing pokies? Is it asking gambling companies to put in betting limits? If this has not been explained before in the survey further clarifying information could help people answer accurately.

Neutral questions and words should be the norm. Even when there is no desire to bias the question and responses, it can be difficult to construct a neutral question that does not lead the respondent. For example, the US General Social Survey conducted experiments to assess the effects of different question wording (Marsden et al. 2020). On priorities for national spending, using the words 'halting the rising crime rate' led to more support than using 'law enforcement' and using the words 'assistance to the poor' provided more support than using 'welfare' (65% vs. 19%) (Smith 1987).

Developing survey questions (Step 6)

When compiling survey questions measuring relevant concepts or models it can be helpful to use or adapt existing survey measures; brainstorm questions or speak to subject matter experts and/or potential respondents.

Caution about small wording changes

Very small word changes and wording differences can make a big difference to survey responses (Bradburn et al. 2004). The previous section described the processes involved in answering questions and how the wording can affect responses. This is why we encourage you to use well validated and tested questions where possible. Try not to make unnecessary changes to original questions based on individual preference and intuition. Any alterations to a validated question should be underpinned by pretesting (see Pre-testing the survey section for more information).

Even when you are writing questions from scratch, reviewing surveys on similar topics can help identify whether there are tested and validated questions available. Most researchers are willing to share a full copy of the surveys for use, when the source is acknowledged. However, not all 'validated' surveys are of equal quality. To determine quality of an existing questionnaire or construct, look for published articles on the development of the construct or questionnaire and check how vigorously the questions were developed. If you cannot find this information online, you could contact the survey author directly. Validity is explained in the section on Sources of errors in surveys.

Sometimes several questions or items may need to combine to measure a single concept. For example, in the APS census, employee engagement is measured by 10 questions. These 10 questions form a concept or construct.

Each question should contribute to at least one key concept that is encapsulated in the research questions driving the project. If a question does not contribute to a research question it may not be needed in the survey.

Most surveys contain many closed questions and a few open ended questions. **Closed** questions are easier to answer, analyse and compare between respondents. They can also help prompt respondents to answer sensitive questions. However, poorly designed closed questions can discourage respondents, particularly if the response options do not contain the answer the respondent wants to give.

Open ended (free-text) questions allow respondent to write their response rather than chose from response options such as categories, numbers or ranges. These can also include "Other please specify' options within closed questions. Open ended questions increase the overall time taken to complete the survey. If there are many open ended questions, it may take respondents a long time to answer and a lot of time and resources to code. Asking respondents for free-text responses but not analysing their responses due to a lack of time, is doing a disservice to people who provided their time and effort. If there are a large number of open ended questions that are of interest in the research it may be worth considering using interviews or focus groups to help clarify the survey design, or as a complement.

Free text fields can often result in data that is not useable. Wherever possible, use drop-down boxes or other selection methods. For example, instead of APS respondents typing in the name of the department in which they work, provide drop down boxes with the names of agencies. If the list cannot be exhaustive, include at least the 10 most common responses and provide an option for "Other please specify".

If free-text responses cannot be avoided, for example where a respondent must supply an email address or phone number, please ensure data validation is in place. Validation works by checking a response against a rule. For example that the input follows the pattern xxxx@xxxx.xxxx or is a 10-digit number for phone numbers. While respondents can still include errors it reduces common errors due to typing or inattention.

Response categories

Most questions will be closed and therefore will need to appropriate response categories. How response categories are constructed might influence how a respondent answers the question. Hence careful framing of response categories is as important as constructing the question stem.

There are many forms of response categories such as:

- Frequency scales (never, rarely, often, sometimes, always) (2-3 times a week, once a month, once a year)
    - Think about the range of frequencies, the range and how you order them (Courneya et al. 2003)
- Likert scales (e.g. strongly agree, agree, neither disagree or agree, disagree, strongly disagree)
- Ranking scale (rank your top three reasons for leaving your job)
- Rating scale (on a scale of 1 to 10, how rate how well the Life Checks website performed in these areas: information is easy to understand, you know what areas you need to spend more time on, the links and resources provided are useful for your circumstances etc.)

The **most important** rule is that response categories must be exhaustive and mutually exclusive (Robinson and Leonard 2019). There are no hard and fast rules for how many response categories are appropriate, it's about balancing the level of detail required and providing enough choices for respondents to answer easily.

Other things you need to consider (some of these are highly debated) are:

- Whether to specify a single response or multiple responses e.g. "select all that apply" type questions.
- Scale length for Likert or Likert type scales
    - There are many studies in the literature and there is no consensus on the best length but usually they are 4-7 points (Debell et al. 2019)
- Odd or even number of response categories (e.g. 5 point or 6 point)
- Whether to offer middle point or neutral options such as 'neither agree nor disagree'
    - Include a neutral option if you think some respondents might be neutral and a neutral response is meaningful. Sometimes, participants use these categories as a *de facto* "Don't know" option, rendering it largely meaningless.
    - There is some suggestion that the middle point is a face saving way to say Don't know (Sturgis et al. 2014). A middle option improves data quality for bipolar questions (Wang and Krosnick 2020)
    - Some respondents may feel angry or irritated when no neutral is provided, particularly where questions are poorly worded and/or cognitively taxing.
- Whether to provide a balance of response categories or not
    - A bipolar scale has an equal number of response categories on the positive and negative ends of a scale. For example strongly disagree, disagree, agree, strongly agree.
    - A unipolar scale has an unbalanced distribution. For example unsatisfied, satisfied, somewhat satisfied, completely satisfied. This can be useful when grades of satisfaction are of interest, but you plan to group all unsatisfied respondents together.
- Whether to include a 'Don't Know' or 'Not Applicable' option
    - "Don't know' can indicate the question is hard to answer or if you are asking knowledge/awareness type questions and you want to know how many were not aware. If it is likely that most respondents will be able to answer, exclude 'don't know'
    - 'Not applicable' may be useful if there are questions that are not relevant to some respondents and you don't have the ability to hide questions.
- Whether to include 'Other – please specify' option. Use these *only if* necessary as the back coding and analysis can take substantial time.

- Ensuring question stem and response categories match (see Box 2 below)

Box 2: An example of mismatch question stem and response categories

Overall, how satisfied were you with your visit to the clinic today? (Question stem)

Very poor, poor, satisfactory, good, excellent (response categories)

Note: Adapted from (Robinson and Leonard 2019)

Ordering and context effects

Both the ordering of questions and the ordering of response categories can influence how a respondent answers a particular survey question. This is because previous questions establish a context in which following questions are interpreted. Respondent's previous answers are another mechanism through which order effects can occur. Responses can also be affected by visual presentation, whether general questions are asked first and cultural context with individualistic countries having a stronger response to order effect compared to non-individualistic countries (Stark et al. 2020; Stefkovics and Kmetty 2022). Ordering effects appear to be stronger for those who do not have strong views, have a lower level of education, older respondents or those with memory loss (Narayan and Krosnick 1996; Knauper 1999). Box 3 presents an example of an ordering effect. Some online survey platforms will allow you to randomise the order of groups of questions which can help with order effects.

Box 3: Question ordering

Question 1. Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children?

Question 2. Do you think it should be possible for a pregnant woman to obtain a legal abortion if there is a strong chance of a serious defect in the baby?

Source: (Stark et al. 2020)

Questioned First	Yes to Question 1 (%)	Yes to Question 2 (%)
Question 1	60.7	48.2
Question 2	84.0	83.0

Pre-testing the survey

When writing original questions it is hard to predict how respondents will interpret and answer the question. Pre-testing questions can uncover unexpected ways in which people interpret questions and therefore it can help us avoid delivering a survey only to discover that the responses look like nonsense. If the purpose of the survey is to compare responses by subgroups it is essential to test the questions with different respondents. Some groups may differ in how they interpret and respond to question, so this is especially important if the survey contains original questions or complex ideas. Even when using validated scales, or minor modifications to validated scales, pre-testing can help identify differences due to

country, cultural differences or other contextual differences. The pretesting phase can be iterative, or just one round if time is short. Remember, some pre-testing is better none.

There are a few ways to pre-test the survey, including:

1. Expert reviews (e.g. academics in the relevant field of research or experienced survey methodologists);
2. Cognitive interviews;
3. Focus groups;
4. Survey assessment tools; and,
5. Pilot testing.

There are also use the free online tools that offer easy feedback. However, they are not as good as seeking expert reviews or conducting your own pre-testing such as cognitive interviews and other types of pre-testing. These tools should not be the only pre-testing you do.

Cognitive interviews (Step 8 in Step-by-step guide to survey design)

Cognitive interviews are a type of a qualitative research method. Their main purpose is to understand how respondents answer survey questions (also see Answering questions: What's involved?). The two common approaches are 'Think aloud' and 'Probing' methods. Table 2 presents examples of probes in cognitive interviews.

If the aim is to test the whole survey, allow a couple of hours per cognitive interview. For shorter interviews, carefully select which survey questions to test. Try to test questions that are of most concern. If the participant consents, record the interview to provide the opportunity to review the interview at a later stage. This can help pick up things that may have been missed and to look at facial expressions and body language. It can be helpful to have a second person in the interview to act as a note taker. This means the interviewer can fully focus on the cognitive interview rather being distracted by taking notes and capturing key information.

For a 'Better Practice' survey, those participating in these cognitive interviews (usually n<10) should be similar to respondents in the survey. They should be recruited and screened for eligibility. This can be done cheaply and quickly (< $10k) by using a supplier for recruitment. Generally this level of cognitive testing can be included even when timelines and budgets are relatively tight.

For a 'Pragmatic' survey, colleagues and friends can act as a convenience sample for cognitive interviews. Doing at least two or three cognitive interviews with a convenience sample is a minimum for a 'Pragmatic' survey and we recommend five interviews. However, this approach may not pick up all the issues in the survey as the interviewees may not adequately reflect the diverse range of survey respondents (e.g. in terms of income, CALD, age, industry).

To learn more about conducting cognitive interviews, please see the cognitive interview text listed in the Further reading Section or talk to colleagues experienced in cognitive interviews.

Table 2 **- Examples of types of cognitive interview probes (adapted from (Robinson and Leonard 2019))**

Cognitive interview probes	Example
Comprehension/interpretation probe	What do 'suitable duties' mean to you?
Paraphrasing	Can you try rephrasing the question in your own words?
Recall probe	How do you remember that you have eaten 10-12 serves of vegetables this week?
Confidence judgement	How sure are you that you have planned for ageing in terms of finances?
Specific probe	Why do you think musculoskeletal disorders are the most serious issue in the workplace?
General probe	How hard or easy was this to answer? What made you arrive at that answer?

Checking the survey for errors (Step 13 in Step-by-step guide to survey design**)**

Once the survey is programmed on a survey platform, ask colleagues to test the survey. Ask them to make sure branching is correct and all figures are readable on both desktop and mobile and to look out for typos. Where possible, especially if the survey is complex, provide clear instructions for testing and specific things to look out for.

This step should be completed regardless of what type of survey you are conducting.

Soft Launching (Step 14 in Step-by-step guide to survey design**)**

A soft launch allows you to check for errors in the survey flow or obvious errors in questions by pausing early in the data collection process to analyse data. It also allows you check that survey is accessible to potential respondents (e.g. not blocked by firewalls on the APS IT environments). The sample size depends on a number of factors but 100 responses – or more if the survey has complex, branching logic – is generally sufficient to find errors. This soft launch data provides an additional check point and provides the opportunity to make minor changes and have a sense of how the quotas are falling (if you have quotas).

- A main differences between soft launch and pilot testing are the degree of testing done, the type of changes that can be made, the time taken and whether you use the data in your analysis. In pilot testing, typically the pilot sample is not used in final analyses as

there are often substantial differences made to the survey questions or the way the survey is implemented. Data from the soft launch is typically used for final survey analyses.

In addition, a pilot allows detailed analysis of the sample to enable substantial changes in the survey (e.g. if scales are not working). A pilot also allows the testing of timing, method and the effect of survey reminders. Pilot testing is like a more rigorous type of pre-testing compared to a soft launch. More information is provided in Box 4: pilot testing.

Box 4: Pilot testing (usually done in best practice surveys)

- Pilot testing is essentially field testing of a survey, starting from recruitment, to reminders to the actual mechanics of the survey itself. The sample size may be 50 to 500, depending on how complex your survey is. Pilot testing can:
- Check whether survey respondents will need additional instructions
- Identify questions that are problematic (by looking at survey responses or asking a question about difficult-to-answer questions, e.g. 20% skipping a question is a sign that question may need to be revised)
- Check scales (existing or original e.g. by using factor analysis
- Check whether there are issues with skips and branching, and
- Determine if changes are needed in terms of survey implementation.

Sampling approaches (Step 3)

There are many sampling approaches for surveys, depending on purpose and the need for generalisability. The sampling approach determines the survey error. High quality sampling methods and sampling frames mean you can generalise to the population from which the sample was drawn, and get accurate estimates from a smaller sample size. However, larger sample sizes do not automatically mean accurate estimates. Deficiencies in sampling may mean that even very large samples can result in inaccurate estimates. Hence a basic understanding of sampling methods for surveys is critical.

To learn more about probability samples and nonprobability online panels in Australia, the ANU The Online Panels Benchmarking Study is a good primer (Pennay et al. 2018).

Probability sampling (Box 5) is hard to achieve and can be very expensive. However, it is included it in this guide because an understanding of different sampling approaches enables you to understand the limitations of different survey methodologies.

Non-probability sampling

With the tremendous increase in online surveys and the substantial drop in landline use (coupled with an increase in mobile usage), probability sampling methods like random digit dialling (RDD) are now much less in use. In non-probability sampling, we don't know a respondent's probability of being selected to take part in the survey and sampling is often non-systematic.

Many online surveys use non-probability samples although the quality of panels can differ. Some online panels have detailed profiles on participants that allow targeted recruitment of particular groups and/or quota sampling.

Although there is estimated 96% internet penetration in Australia[2], there are some differences in social media platform and type of usage within the population, based on age and user preference. Those with poor English language skills are also less likely to take part in online surveys.

There are various approaches to conducting non-probability online surveys and some are more robust and reliable than others. These approaches have differing methodological implications (Lehdonvirta et al. 2021). A survey researcher needs to have some knowledge of different non-probability sampling approaches to make decisions on how to source respondents and to understand methodological limitations of each approach. While steps can be taken to reduce bias, non-probability samples contain a higher degree of bias than probability samples. Non-probability samples are less precise and/or that precision can be difficult to estimate. Generally non-probability sampling is used when it is not possible or practical to use probability sampling. It is good to be aware of limitations in these approaches.

Purposive sampling is when a sample is drawn by experts in the area of study. **Haphazard or convenience sampling** is often surveys done with whoever is available and willing. It could also involve recruiting via social media, newsletters or on a web page. **Snowball sampling** basically involves relying on initial respondents to spread the word and get additional respondents (note: sometimes we use this sampling method for our qualitative research). These methods are useful when surveying hard to reach groups, but can result in poor coverage where the intended sample is heterogeneous.

River sampling (or intercept sampling) is when participants are invited to take an online survey via webpage, email or other avenues where it may reach the target population, similar to convenience or haphazard sampling. Typically a pop up appears on a website for web evaluations (Note: if invitations/pop-ups randomly appear with known probability, it will be a random sample of website users and won't be classified as river sampling). It can also involve distribution of a survey through a third party (not a survey panel) where we expect the target population will see the invite and participate but we cannot be certain. The issue with this approach is coverage bias (Lehdonvirta et al. 2021) arising from unequal access to the survey (e.g. some people are not online, survey not being distributed or advertised through their particular organisations/ networks). Such samples could also be prone to deliberate bias, especially when we rely on a third party to distribute survey invitations.

The use of non-probability online survey panels (also called **panel sampling**) is one of the most common types of non-probability sampling. They are easy and inexpensive to use. However, these online panels over-represent certain groups and under-represent others in ways that cannot necessarily be measured. Moreover there are issues arising from 'professional survey respondents' as respondents usually receive financial incentives for participating. Issues include speeding through the survey, random responses (selecting random answers instead of real answers), straight lining (selecting the same response across all questions), frequent responders, bots and liars (people who lie to either qualify for the survey or provide responses in a certain way to obtain advantage). There have been many studies comparing non-probability online studies with probability samples and national benchmark data, all showing that even with post survey adjustments (e.g. with weights), the

[2] https://datareportal.com/reports/digital-2023-australia

non-probability online samples produce less accurate results (Brüggen et al. 2016; Yeager et al. 2011). However, there is some evidence that non-probability online panels offer better estimates than river sampling (Lehdonvirta et al. 2021).

Quota sampling occurs when the researcher specifies quotas for respondents with particular characteristics (e.g. 50% of the sample must be male). The actual selection of respondents is then up to interviewers or online panel programming. Researchers often specify quotas for online panel sampling. This is often implemented to obtain a more representative sample or ensure that we have enough respondents for our groups of interest (e.g. 30% of the sample must be the main financial decision maker in their household). However, quota sampling does not allow the findings to be generalised to the broader population. For example, using quotas to match the sample with the Australian population statistics on age, gender and state or territory of residence does not allow you to generalise to all Australians.

Box 5: Probability sampling

Probability sampling methods are the ideal sampling methodology in survey design. Using these methods, respondents are selected using some type of probabilistic method, we know the probability of selecting a member of the sampling frame (Dillman, Smyth et al. 2014) and the probabilities are all non-zero i.e. every unit in the sampling frame has a chance of being selected. Probability sampling based on simple random sampling or variations on the random sample are mostly likely to provide information that is representative of the population. Probability sampling also allows calculation of sampling error whereas other methods of sampling do not.

We can use probability sampling for online surveys in some circumstances. For example, where there is a complete list of all email addresses for an organisation we can randomly sample participants. For surveys embedded on a website, we can also randomise pop-up invitations with a known probability (e.g. 10% probability that visitors to the site will see the popup).

Probability sampling methods

In simple random sampling, random numbers are generated and units or elements from a sampling frame are selected based on the random number. This is a similar process to assigning people to arms during a trial. The difference is that selected sample is used for survey distribution rather than whether a person is assigned to the treatment or control groups. Note that just because it is a random sample, it is not guaranteed to perfectly represent the population. However, the larger the sample, the more likely it is to represent the population from which is was drawn.

Systematic sampling uses a sampling interval instead of using random numbers. This is often commonly used in clinical trials with recruitment of patients presenting at a participating medical centre. If you know the size of the population and the sample size you would like, you can come up with the sampling interval (e.g. every 5th person is selected).

Stratified sampling divides the sample into strata (or groups) first and then draw a random sample from with each stratum. Strata must be exhaustive and mutually exclusive and each unit must fall into only one stratum. The stratification variable should also be correlated with the outcome variable of interest. This is done if you want data for a number of groups and you want to ensure there is enough sample from each group. Examples of strata are

occupation, industry and business size and gender. It is possible to sample different strata at different rates (e.g. oversampling First Nations participants). However, we recommend involving a statistician/survey methodologist to ensure this is done well and accounted for in the estimates.

Cluster sampling is often used when there is a lack of a good sampling frame or if the cost to reach the unit of the sample is considered prohibitive. Cluster sampling is when a group or a collection of individual units are selected as the first part of random sampling. For example, you could randomly select suburbs in capital cities (first cluster), then randomly select streets (second cluster) and then randomly select households within selected streets. This is especially helpful for face-to-face surveys as it can reduce the cost of travel for interviewers. Cluster sampling increases sampling error and sampling variance compared to simple random sampling.

It is common to use these techniques – and others – in combination. For example many samples might be both stratified and clustered and stratified designs often use systematic sampling.

Probability sampling unit

Sampling isn't always applied to individuals, in fact it is common for the sampling frame to list units other than people.

Previously, random digit dialling (RDD) was a common approach for telephone or computer assisted telephone interviews (CATI). RDD overcomes the problem with not everyone having internet access and ensures wide coverage by randomly creating phone numbers and dialling them. If differences in coverage of certain populations and areas are known, they may be adjusted (weighted) to correct for differences in probability of selection (Keeter, Kennedy et al. 2006).

Address-based sampling is another approach. Some online survey panels may use this approach obtain their panel (e.g. if they are setting up probability panels) and then use online methods for subsequent communications. This can be helpful to first invite participants and ask them to take the survey online. This is the primary method used for the 2021 ABS Census

Other probability samples

Lastly, there are online probability panels available, such as the Life in Australia Panel through the Social Research Centre. It was established using landline and mobile random digit dialling. However, probability online panels, as with other probability methods, are more costly than non-probability panels.

Survey modes (useful for Step 3)

When we talk about survey modes, we mean both the mode of survey recruitment and reminders and/or the mode of the actual survey.

It is common to use online surveys with recruitment of respondents from an online panel. Online surveys provide a cost-effective method of collecting data compared to other survey modes. However there are some disadvantages of online surveys, particularly coverage error. This method may be missing older people or those without reliable internet at home, and even among those with readily accessible internet it only covers those who have signed

up to the panel. Panel members are likely to be different from people who are not panel members in ways we can't measure. Moreover, it is very difficult to obtain probability samples unless initial survey invitation is done by another method (e.g. paper) or online probability panels are used. Both of these options considerably increase the cost of the survey recruitment.

Occasionally there is need to use other modes, such as surveys targeting groups with poor internet access. These are mentioned briefly here.

Surveys used to be largely paper based. Paper surveys were posted out to respondents and required either intensive data entry or printing of machine readable forms that could be converted more easily into data. There are companies specialising is online paper forms, reading responses, and printing and mailing survey approach materials. The advantage is that it is relatively cheap compared to telephone or face to face surveys. Probability sampling is feasible with this mode.

Telephone surveys were popular when the majority of the population had landlines, especially with the availability of RDD and computerised interviewing to reduce data entry efforts and interviewer error. This method is more expensive than mail surveys but can be completed quickly and have similar benefits to face to face interviews but at a lower cost. However, the popularity of this approach has waned as households and younger people moved to mobile phones. Households with landline are now uncommon.

Mixed mode surveys use more than one method. For example, the 2016 census was mixed mode with invitation by mail and respondents were asked to do online surveys but could request a paper form if preferred. The Australian Bureau of Statistics also used SMS messages and emails as part of their approach methods as they found contacting respondents by multiple modes of communication was effective (personal communication, ABS).

Improving survey response rates

Good survey design considers what motivates people to complete a survey. A good survey:

- establishes trust with the participant
- shows the personal benefit of completing it (including offering incentives)
- creates a sense of urgency; and
- is quick and easy for people to complete.

These 4 criteria are based on BETA's experience and the literature on survey design (specifically we draw in social exchange theory).

Establish trust

Establishing trust is a key factor for whether people complete a survey or not. People are constantly bombarded with requests (including from potential scammers) to complete surveys. Trust can be built by increasing respondents' confidence in how their data will be handled, and by allowing respondents to verify the authenticity of survey requests by providing contact emails. Tailoring the language to suit potential respondent shows a level of thought about who will complete the survey. Making sure all communications and content relating to the survey are holistic and that different communications are linked, is also likely to increase respondents' confidence that they survey is credible and trustworthy.

Another way of increasing trust is ensuring we have ethics oversight and providing ethics committee's contact details for issues and concerns.

Messengers and sponsors

Trusted messengers and sponsors can be used to leverage existing relationships. Who is sending out the survey makes a difference to whether people respond. This is related to trust as well as benefits of survey participation.

People are more likely to respond if they think the survey is from a legitimate organisation and the request is legitimate. In a Cochrane systematic review of RCTs that investigated survey design features to improve response rates, it found that surveys from universities were more likely to be filled out than those sent by government or commercial organisation (Edwards et al. 2009).

Increase benefits

One of the ways to encourage respondents is to increase the benefits of taking part in a survey. Ways of increasing the perceived benefits are;

- outlining how the survey data will be used,
- asking for assistance or advice on a policy problem (creating a sense of reciprocity),
- emphasising that opportunities to respond are limited,
- using social norms to say others like them have responded,
- providing incentives to show you appreciate the time and mental cost to the participant (see below for more on incentives).

Incentives

Using cash or gift cards is an effective way to increase survey response rates. Incentives encourage people to start a survey (including an online survey) and they are more likely to complete the survey if an incentive is offered compared to when there are no incentives (Göritz 2006).

Things to consider if offering an incentive include:

- whether the incentive is provided with survey invitation or on completion of the survey (Gajic et al. 2012). Providing an incentive with the survey invitation can be effective for mail-out surveys because it creates a sense of reciprocity (Birnholtz et al. 2004), but may not be practical or feasible for online surveys.
- Whether the incentive is provided as a lottery or for everyone. While incentive for every respondent are generally more effective than lotteries, lotteries re more effective than no incentive and can be more cost effective.

Create a sense of urgency

Creating a sense of urgency, and imposing a deadline by having the survey open for a relatively short time (for example 2 weeks) can increase participation compared with a longer deadline. If the deadline is too far in the future people delay responding because there is plenty of time, which then increases the risk that they will forget or run out of time.

Make it easy to complete

Response rates increase when it is easier to complete a survey. Strategies to make surveys easy to complete include:

- Reducing survey length.
- Reducing complexity as complex questions take more time and cognitive load.
- Using channels that make it easy to respond. For example, ensure surveys can be completed on a mobile device, and include survey links in SMS and emails using shortened links or QR codes.
- Minimise asking for sensitive or personal information and allow people to skip questions with which they are uncomfortable.
- Provide regular reminders. With every reminder, you can see a bump in the response rate. Four to five reminders are common.

Keep surveys short and interesting

Stating up front that there are 'just five questions' (or similar) can assist in improving survey completion rates. It's worth critically considering whether a long detailed response from 5% of your sample or responses to your top five questions from 20-50% of your sample is more important.

Introducing a mix of question types can make the survey more engaging compared to long repetitive blocks of matrix questions.

Take care when designing a survey to avoid things that annoy and disengage participants. These can include:

- There isn't a response option that fits. Try offering an 'other specify' option unless you are confident you have covered all possible responses.
- Asking the same question in slightly different ways. Only do this if you have to (copyrighted scale for example).
- Forced response questions. An alternative is to let participants skip questions rather than require a response and prevent the participant moving to the next question.
- Repetitive rounds of questions – for example if you ask people which of the following companies they've ever heard of (out of a list of ten), and then ask them a batch of ten questions about each one they picked.

Consider mixed mode surveys

Mixed mode surveys reduce the cost of participating in the survey by making it easier to respond, increases the benefit and also helps establish trust. Mixed mode surveys can take many forms. The following approaches have proven effective in boosting response rates:

- Survey data collection could commence with the cheapest mode (e.g. online) and then follow non-responders with a paper survey or telephone interview
- A pre-survey notice, invitation or incentive could occur through one mode (e.g. mail) even if the survey is to be conducted online. Or using SMS and email.
- Reminders (by email and text) can follow the initial contact.

While there is good evidence that mixed mode at recruitment is effective in improving response rates, offering mixed modes for survey completion (e.g. online and paper) do not always lead to improvement (Dillman et al. 2014).

Ethics and privacy (Step 5 & Step 11)

There is a great deal of overlap between preparing and drafting materials for a survey and preparing for an ethics application. When preparing for ethics review we suggest writing down research questions, planning analyses, considering sampling and calculating required sample size. Preparing for an ethics application should help with survey development and implementation and vice versa.

To comply with ethics and privacy requirements, you need to draft **Participant Information and Consent Form** (PICF) which is a standard requirement for all surveys. This is usually hosted on the landing page for the survey. The PICF usually incorporates the Privacy Collection Notice. Consult with your organisation's privacy team for advice on the collection notice and other privacy issues such as a Privacy Impact Assessment.

Data storage and sharing (Step 17)

In the APS, there are key Australian government guidelines and legislation that we need to follow including the Information Security Manual, Australian Privacy Principles and your organisation's ICT Security policy.

Survey data is collected using survey programming platforms (e.g. Google Forms, Qualtrics, Survey Manager, Survey Monkey). Generally, survey data must not be retained in these platforms once the survey is completed. Data should be transferred to either your organisation's records management environment and/or analysis environment. When receiving data from a third party it should be stored in these locations. All data from surveys conducted by the APS must be stored onshore, even while the survey is in the field.

Survey data is often sensitive and rarely anonymous. Removing direct identifiers (such as names and addresses) does not mean that the data are anonymous. When the data contains a number of demographic factors it is probably potentially re-identifiable. Consider the following situation: you conduct a survey and do not record any direct identifiers or the IP addresses of respondents. You ask a number of demographic questions that might be of relevance to your policy area. The intersection of rare categories can make a person identifiable, for example very old people, gender diverse people, people in rural or remote areas, people born overseas, or Aboriginal and Torres Strait Islander people. For this reason data must be kept secure and access to the data must be limited to people who have an operational need to use it.

If you wish to share data with other agencies or with the public this should be communicated in the PICF so respondents can consent.

Pre-analysis Plan (PAP) (Step 10)

Pre-analysis plans are common for confirmatory research such as randomised controlled trials (RCTs). However, it is not common to draft analysis plans for surveys. There are a number of benefits to PAPs for surveys:

- It will allow you to decide on the format of your questions (e.g. do you need a continuous variable, how many response categories do you need, do you intend to dichotomise, do you intend to compare with other studies)
- It will improve measurement and data quality, and
- It will allow a more consistent approach between how surveys are developed and analysed.

Survey PAPs do not need to be as extensive as those for confirmatory research. Survey research is often exploratory and by definition a very restrictive PAP is not appropriate. PAPs for a survey could be as simple as a document with example graphs and tables and/or a few dot points on how you intend to analyse and report the data. If you are doing confirmatory research in your survey, that is research to test hypotheses, a more comprehensive PAP may be appropriate.

With a documented survey purpose, research questions and key concepts and models, the PAP is well underway. Focus on key variables and relationship between variables. It is often useful to also know how you might want to display the results beforehand. This allows you to work backwards from analysis plan to question design.

You can find a survey PAP template on page 52 of this guide.

Sources of errors in surveys

Survey errors

There are four types of survey errors. Understanding the source of these errors and how to reduce them helps improve survey design.

Coverage error occurs when parts of the target population cannot be included in the sample. Understanding who is included the sample frame and who is not will help you either address the issue or inform how you interpret and present the survey findings. We can address coverage error by obtaining as complete a sampling frame as possible, using the most robust sampling method as possible and doing post-survey adjustments to weigh the sample to population of inference (Fricker 2016). However, it has shown that post survey weights in non-probability online surveys could actually make estimates less accurate (Brüggen et al. 2016).

There is no list of everyone who uses the internet or everyone's email addresses which creates coverage error for many online surveys.

Sampling error occurs when the target population is sampled. So any survey that uses a sample will have sampling error. This does not include a census, which includes every unit in the population of interest. This arises from the fact that different samples would generate different survey estimates. However, sampling error can be reduced by using probability samples of adequate sample sizes. With probability sampling, you can estimate sampling error fairly accurately. Even with probability sampling, methods such as cluster sampling increases sampling error.

Sample size needed will depend on your research questions and the types of analyses you intend to undertake.

Nonresponse error occurs when people fail to respond to the survey or skip items. Often, the response rate (the number of respondent divided by number sampled) is used to judge degree of nonresponse error and whether findings may be generalised. Surveys with higher response rates are generally considered less biased (nonresponse bias) than those with lower response rates, although this alone cannot tell us whether the findings can be generalised or whether there is nonresponse error (Groves and Peytcheva 2008; Hendra and Hill 2019). We need other information such as sampling method and variations from the target population to determine if generalisation is appropriate and whether nonresponse error still exists despite having a high response rate. As a rule of thumb surveys with less than 25% response rate cannot be generalised. This does not mean that those with a response rate higher than 25% are necessarily generalisable.

Measurement error is when survey estimates are different from the true values (Dillman et al. 2014). For example, respondents may not answer questions truthfully or there are unintentional errors from issues with the survey itself.

Two common types of measurement errors are response bias and response variance (also known as reliability). Response bias occurs due to social norms or social desirability bias where estimates are either overestimates (e.g. volunteering, protective health behaviours) or underestimates (e.g. rates of drug use, criminal activity). Response variance or low reliability occurs when measurement at different times produce varying results.

We can reduce measurement error by increasing the overall quality of the survey including:

- drafting quality survey questions
- minimising recall difficulty
- avoiding order issues
- the mode of the survey
- the distribution method
- pre-testing the survey.

Often, researchers tend to focus on a specific type of survey error, often based on errors that are typically emphasised in their field of training or research or the types of errors they know best (Dillman et al. 2014). This means that is it easy to overlook other sources of survey error. The total survey error (TSE) framework was developed to acknowledge this issue in a multidisciplinary way. In this framework, surveys are carefully designed to the extent possible by giving attention to many aspects of survey design such as planning, sampling, question wording, questionnaire design (format), and implementation and data analysis.

Tips to avoid common problems in survey design

A **long** survey is a common flaw in survey design. Often we think a longer survey is more cost effective and we might as well ask things that are of interest but not crucial to the purpose of the survey. Most telephone surveys keep below 20 minutes. Online surveys should take 5-10 minute ideally and avoid being longer than 20 minutes (Revilla and Ochoa 2017). The length of the survey is associated with response rate and response quality (Galesic and Bosnjak 2009).

As mentioned before, **ordering** can give rise to survey issues. It is important to consider the ordering of the whole questionnaire, ordering of specific questions and ordering of response categories. Previous questions can influence respondent's answer to later questions (Neuman 2007). Depending on mode of questions (telephone or online), there can be primacy and recency order effects (Dillman et al. 2014). There are some key ordering issues that apply to most surveys.

- Ask important questions earlier so the data are captured even if the respondent drops out. This is especially important if the survey is long or requires a lot of cognitive effort.
- Put similar questions together and either go from general to specific or specific to general.
- Add introductory statements at the start of new topics. This is helpful especially when you switch between recall periods or question formats or there is no relationship between topics asked.

- Avoid asking boring and sensitive questions first. Do not end with sensitive or threatening questions.
- Randomise the ordering of response categories where there is no set order.
- Present categories that are less likely to be selected at the top or at left and those most likely to be selected at the bottom or at right.
- Decide whether demographic questions are asked at the beginning or at the end of the survey and whether placement earlier can bias responses. Key demographic questions especially those affecting quota and/or branching must be placed earlier. If demographic questions are not as important as other questions, they are usually placed at the end of the survey.
- If you are using a sliding scale, start with the least socially desirable end of the scale on the left. Ordering is further discussed in the Ordering and context effects Section.

Often, we **force** respondents to answer questions to minimise missing responses. It is recommended only essential questions or one that inform branching and skips should be made mandatory. Mandatory questions can lead to drop outs or random answers to continue to the survey (Sischka et al. 2022). Mandatory questions might also include an option for 'prefer not to say' to minimise the risk of poor quality data.

Measurement validity and reliability

Measurement validity refers to how closely align the conceptual and operational definitions are – are you measuring what you think you're measuring? It is harder to achieve than reliability which is discussed later in this section. There are different types of measurement validity (Neuman 2007). See glossary for definitions of different types of validity.

Reliability is whether an instrument produces similar results when comparing measurements from different samples or at different times. Many psychometrically tested scales and constructs have a reliability score provided. When including the same questions, it is possible to calculate reliability statistics and compare it to the original questions. Pilot testing is often a good way to measure reliability.

References

Birnholtz JP, Horn DB, Finholt TA and Bae SJ (2004) 'The effects of cash, electronic, and paper gift certificates as respondent incentives for a web-based survey of technologically sophisticated respondents', *Social science computer review*, 22(3):355-362,

Bradburn N, Sudman S and Wansick B (2004) 'The social context of question asking', in (eds) *Asking questions: The definitive guide to questionnaire design*, Jossy-Bass, San Franciscio.

Bradman N, Sudman S and Wansick B (2004) *Asking questions: The definitive guide to questionnaire design*, Jossy-Bass, San Francisco.

Brüggen E, Brakel JVD and Krosnick. J (2016) *Establishing the Accuracy of Online Panels for Survey Research*, Statistics Netherlands, accessed 1 March 2024.

Bryman A (2006) 'Integrating quantitative and qualitative research: how is it done?', *Qualitative research*, 6(1):97-113, doi:10.1177/1468794106058877.

Courneya KS, Jones LW, Rhodes RE and Blanchard CM (2003) 'Effect of response scales on self-reported exercise frequency', *American Journal of Health Behavior*, 27(6):613-622, doi:10.5993/AJHB.27.6.4.

Debell M, Wilson C, Jackman S and Figueroa L (2019) 'Optimal Response Formats for Online Surveys: Branch, Grid, or Single Item?', *Journal of Survey Statistics and Methodology*, 9(1):1-24, doi:10.1093/jssam/smz039.

Dillman DA, Smyth JD and Christian LM (2014) 'The fundamentals of writing questions', in (eds) *Internet, phone, mail and mixed-mode surveys: The tailored design method*, John Wiley & Sons Inc, Hoboken, New Jersey.

Dillman DA, Smyth JD and Christian LM (2014) 'Reducing people's reluctance to respond to surveys', in (eds) *Internet, phone, mail and mixed-mode surveys: The tailored design method*, John Wiley & Sons Inc, Hoboken, New Jersey.

Dillman DA, Smyth JD and Christian LM (2014) 'Sample surveys in our electronic world', in (eds) *Internet, phone, mail and mixed-mode surveys: The tailored design method*, John Wiley & Sons Inc, Hoboken, New Jersey.

Edwards PJ, Roberts I, Clarke MJ, Diguiseppi C, Wentz R, Kwan I, Cooper R, Felix LM and Pratap S (2009) 'Methods to increase response to postal and electronic questionnaires', *Cochrane Database of Systematic Reviews*, 3):

Fricker RD (2016) 'Sampling methods for online surveys', *The SAGE handbook of online research methods*, 162-183,

Gajic A, Cameron D and Hurley J (2012) 'The cost-effectiveness of cash versus lottery incentives for a web-based, stated-preference community survey', *The European Journal of Health Economics*, 13(6):789-799, doi:10.1007/s10198-011-0332-0.

Galesic M and Bosnjak M (2009) 'Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey', *Public Opinion Quarterly*, 73(2):349-360, doi:10.1093/poq/nfp031.

Goetz EG (2008) 'Words matter: The importance of issue framing and the case of affordable housing', *Journal of the American Planning Association*, 74(2):222-229, doi:10.1080/01944360802010251.

Göritz AS (2006) 'Incentives in web studies: Methodological issues and a review', *International Journal of Internet Science*, 1(1):58-70, doi:No DOI.

Groves RM and Peytcheva E (2008) 'The impact of nonresponse rates on nonresponse bias: a meta-analysis', *Public opinion quarterly*, 72(2):167-189, doi:10.1093/poq/nfn011.

Hendra R and Hill A (2019) 'Rethinking response rates: new evidence of little relationship between survey response rates and nonresponse bias', *Evaluation review*, 43(5):307-330, doi:10.1177/0193841X18807719.

Knauper B (1999) 'The Impact of Age and Education on Response Order Effects in Attitude Measurement', *Public Opinion Quarterly*, 63(3):347-370, doi:10.1086/297724.
Lehdonvirta V, Oksanen A, Räsänen P and Blank G (2021) 'Social media, web, and panel surveys: using non-probability samples in social and policy research', *Policy & internet*, 13(1):134-155, doi:10.1002/poi3.238.
Marsden PV, Smith TW and Hout M (2020) 'Tracking US social change over a half-century: The general social survey at fifty', *Annual Review of Sociology*, 46(109-134, doi:10.1146/annurev-soc-121919-054838.
Narayan S and Krosnick JA (1996) 'Education Moderates Some Response Effects in Attitude Measurement', *The Public Opinion Quarterly*, 60(1):58-88, doi:10.1086/297739.
Neuman WL (2007) 'Quantitative and qualitative measurement', in (eds) *Basics of social research*, Pearson Canada Toronto,
Neuman WL (2007) 'Survey research', in (eds) *Basics of social research*, Pearson Canada Toronto,
Pennay DW, Neiger D, Lavrakas PJ and Borg K (2018) *The online panels benchmarking study: A total survey error comparison of findings from probability-based surveys and non-probability online panel surveys in Australia*, accessed 22 March 2024.
Revilla M and Ochoa C (2017) 'Ideal and Maximum Length for a Web Survey', *International Journal of Market Research*, 59(5):557-565, doi:10.2501/ijmr-2017-039.
Robinson SB and Leonard KF (2019) 'Constructing response options', in (eds) *Designing quality survey questions*, Sage Publications, Los Angeles.
Robinson SB and Leonard KF (2019) *Designing quality survey questions*, Sage Publications, Los Angeles.
Robinson SB and Leonard KF (2019) 'Finalizing questions and using pretesting strategies', in (eds) *Designing quality survey questions*, Sage Publications, Lose Angeles.
Robinson SB and Leonard KF (2019) 'Understanding respondents', in (eds) *Designing quality survey questions*, Sage Publications, Los Angeles.
Rugg D (1941) 'Experiments in wording questions: II', *Public opinion quarterly*, 5(1):91, doi:10.1086/265467.
Sischka PE, Décieux JP, Mergener A, Neufang KM and Schmidt AF (2022) 'The Impact of Forced Answering and Reactance on Answering Behavior in Online Surveys', *Social Science Computer Review*, 40(2):405-425, doi:10.1177/0894439320907067.
Smith TW (1987) 'That which we call welfare by any other name would smell sweeter an analysis of the impact of question wording on response patterns', *Public opinion quarterly*, 51(1):75-83, doi:10.1086/269015.
Stark TH, Silber H, Krosnick JA, Blom AG, Aoyagi M, Belchior A, Bosnjak M, Clement SL, John M, Jónsdóttir GA, Lawson K, Lynn P, Martinsson J, Shamshiri-Petersen D, Tvinnereim E and Yu R-R (2020) 'Generalization of Classic Question Order Effects Across Cultures', *Sociological Methods & Research*, 49(3):567-602, doi:10.1177/0049124117747304.
Stefkovics Á and Kmetty Z (2022) 'A comparison of question order effects on item-by-item and grid formats: visual layout matters', *Measurement Instruments for the Social Sciences*, 4(1):8, doi:10.1186/s42409-022-00036-z.
Sturgis P, Roberts C and Smith P (2014) 'Middle Alternatives Revisited:How the neither/nor Response Acts as a Way of Saying “I Don’t Know”?', *Sociological Methods & Research*, 43(1):15-38, doi:10.1177/0049124112452527.
Tashakkori A, Johnson RB and Teddlie C (2021) *Foundations of Mixed Methods Research*, Sage Publications, Thousand Oaks.
Tourangeau R, Rips LJ and Rasinski K (2000) *The Psychology of Survey Response*, Cambridge University Press, Cambridge.
Tourangeau R, Rips LJ and Rasinski K (2000) 'The role of memory', in (eds) *The Psychology of Survey Response*, Cambridge University Press, Cambridge.
Wang R and Krosnick JA (2020) 'Middle alternatives and measurement validity: a recommendation for survey researchers', *International Journal of Social Research Methodology*, 23(2):169-184, doi:10.1080/13645579.2019.1645384.

Yeager DS, Krosnick JA, Chang L, Javitz HS, Levendusky MS, Simpser A and Wang R (2011) 'Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples', *Public opinion quarterly*, 75(4):709-747, doi:10.1093/poq/nfr020.

Glossary

Term	Definition
CAPI	Computer Assisted Personal Interviews
CATI	Computer Assisted Telephone Interviews
Coverage error	It is a type of non-sampling error. This happens when certain people or groups are not in the sampling frame. For example, if we are doing a telephone survey (landline) but many people do not have landline numbers so they are excluded from our sampling frame and this causes coverage error.
Coverage rate	The proportion (could be estimated) of the target population that is included in the sample frame (e.g. 40% of bank customers with mobile contact details)
Measurement error	The difference between estimates (e.g. proportion of people who are sales workers from survey data) produced from the survey and the true value due to survey respondents providing inaccurate answers (e.g. actual proportion of people who are sales workers). This error could be due to a lack of clarity of survey questions, having inappropriate response options, problems with recalls or human error.
Non-probability sampling	Survey sampling in which invitation to participate in the survey are selected using volunteers or other methods that are not systematic (e.g. convenience samples, some online survey panels). The problem with this method is that the people who are not invited or do not respond to the survey may be different from those who do.
Non-response error	This error occurs due to inability to collect responses from everyone in the sample. This could be at the survey level (e.g. not reaching young people) or at the item or question level (people skipping questions).
Pragmatic surveys	These are defined as surveys when we have limited resources and skills. In this scenario, we still need to have a clear purpose. But we may have to condense, compromise or skip some of the survey steps involved to develop best practice surveys. Even for pragmatic surveys, we want them to go through a few steps (e.g. cognitive interview with a convenience sample, asking colleagues to test for logic and skips once the survey is programmed etc).
Probability sampling	A type of survey sampling where everyone in the sampling frame has a known probability of being selected to participate in the survey. Probability sampling allows researchers to draw inferences on the target population.
Priming	A phenomenon in which how a person respond to a survey question influences how they respond to subsequent survey questions (priming occurs in many contexts but here we are only talking about priming in surveys)

Term	Definition
Best practice surveys	Surveys developed with a clear purpose using conceptual models. They contain quality survey questions and are pre-tested rigorously. They take into account respondents' needs, interests and abilities and are developed using a purposeful survey development process.
Question stem	The actual question or statement
Respondent burden	The effort and time required by respondents to answer surveys. The number of surveys people are asked to complete, the length of surveys and poorly worded questions and response categories add to respondent burden.
Response rate	The number of people who responded to a survey expressed as a proportion of the total number of people invited to participate (i.e. survey sample). Note: there are different formulas for calculating response rates so do not assume all response rates you see in the literature are comparable.
Sampling error	This error occurs when a sample of the population is sampled rather than the whole population and happens because different samples can be drawn from the population with differing estimates. It is essentially the difference on a variable between those who were sampled and those who were not. Sampling error occurs whenever we are choosing a sample rather than surveying everyone as in the Census.
Sampling frame	Sampling frame is essentially a list of units (e.g. people/households) in the population. It also contains features and characteristics of the target population. Population is an abstract concept and it can be continually changing. A researcher needs to define a list of characteristics that would best approximate the population of interest (e.g. All businesses in Australia with a turnover of $2million or more, all persons residing in NSW between 1 March and 31 March 2023 aged 18 to 64). This list containing specific characteristics of the population is the sampling frame. In the case of online panels, all members (or those meeting specific characteristics) could be the sampling frame.
Sampling process	How units of the survey (e.g. people, households) are chosen from the sampling frame
Social desirability bias	A type of response bias that occurs due to the tendency of people to answer survey questions in a way that will be viewed favourably by others rather than answer truthfully. This can lead to over-reporting of positive behaviours (e.g. exercise) and under-reporting of negative behaviours (e.g. gambling, smoking, cheating).
Survey approach materials	These include original survey invitation, reminders and incentives as well as the paper survey if it is being mailed out.
Survey population	It is a part of the target population that is able to be surveyed. It is also called the coverage population.
Survey respondents	These are people who responded or participated in the survey. They are a subset of the survey sample.

Term	Definition
Survey sample	Members of the sampling frame who are selected to be surveyed. Note this is different from survey respondents who are people within the survey sample who responded to the survey.
Target population	This is the group that we are interested in and excludes groups that are out of scope for the survey (e.g. those in residential aged care, in prison, under 15 years of age). It is basically the population that we want to generalise the survey results and is also known as the scope of the survey. The target population is a theoretical population as there would be people in this population who could not be surveyed (e.g. difficult to contact, missing from sampling frame)
Unit/element	Person/household/business from which information is sought for a survey
Validity, face	Face validity is whether judgement by a group of people (either experts or those with some knowledge) accept that the question or indicator measures the construct that it is intended to measure. It is the easiest form of validity to measure. The constructs need to be clear and well-defined to achieve face validity.
Validity, content	Content validity is a type of face validity and it is an assessment of whether the full definition of a measure/concept is captured by the question or indicator. For example, if your definition of worker engagement includes having meaningful work, having autonomy, organisation support and growth opportunities but your questions do not include anything on growth opportunities or autonomy, we can say it has low content validity.
Validity, criterion	Criterion validity is when a criterion or some sort of standard is used to indicate a construct is measured accurately. Concurrent validity and predictive validity are types of concurrent validity. Concurrent validity is measured using level of agreement with a pre-existing indicator that is assumed to be valid. For example, if you are creating a new measure of intelligence, you can measure concurrent validity by correlating it with IQ scores. This is often done when you are developing health scales such as depression and anxiety scales where a validated or gold-standard measures already exist.
Validity, predictive	Predictive validity is where an indicator/question predicts future outcomes. Predictive validity is important in epidemiological surveys where you are measuring risk factors or constructs that are associated with health outcomes (e.g. depression, cancer, obesity). Another example is NAPLAN scores. If these are predictive of success later in life (e.g. in university), then we can say the scores have predictive validity (Note I am making up this NAPLAN example).
Weighting error	Loss of precision due to post-survey adjustments. This error can be estimated and included in the sampling error.

Term	Definition
Weights	Weights are used to adjust survey estimates. They correct for differences in the probability of selection and differences in the response rates among different groups in the survey sample. For example, if you oversample First Nations people so that you have enough data points on this group, you would have weights when generalising it to the Australian population to account for this oversampling. Weights allow you to generalise to the target population.

Note: definitions are adapted from various survey resources cited in this guide plus the Australian Bureau Statistic's definitions.

Key demographics for sampling

Table 3. Age distribution and median age in Australia as of 2021 Census

Age group	*n*	%
Median age	38	
0-4 years	1,463,817	5.8
5-9 years	1,586,138	6.2
10-14 years	1,588,051	6.2
15-19 years	1,457,812	5.7
20-24 years	1,579,539	6.2
25-29 years	1,771,676	7.0
30-34 years	1,853,085	7.3
35-39 years	1,838,822	7.2
40-44 years	1,648,843	6.5
45-49 years	1,635,963	6.4
50-54 years	1,610,944	6.3
55-59 years	1,541,911	6.1
60-64 years	1,468,097	5.8
65-69 years	1,298,460	5.1
70-74 years	1,160,768	4.6
75-79 years	821,920	3.2
80-84 years	554,598	2.2
85 years and over	542,342	2.1
Total	**25,422,788**	**100.0**

Source: Census of Population and Housing: Population data summary, 2021

Table 4. Population distribution for sex, 2021 Census

	n	%	Median age
Male	12,545,154	49.3	37
Female	12,877,635	50.7	39
Total	25,422,788	100.0	38

Source: Australian Bureau of Statistics, Snapshot of Australia, Census 2021

Table 5. Count and percentage of persons(a) by state and territory by sex, 2021 Census

	Male n	Male %	Female n	Female %	Total n	Total %
New South Wales	3,984,166	31.8	4,087,995	31.7	8,072,163	31.8
Victoria	3,200,963	25.5	3,302,528	25.6	6,503,491	25.6
Queensland	2,540,404	20.3	2,615,736	20.3	5,156,138	20.3
South Australia	878,592	7.0	902,924	7.0	1,781,516	7.0
Western Australia	1,322,855	10.5	1,337,171	10.4	2,660,026	10.5
Tasmania	273,765	2.2	283,804	2.2	557,571	2.2
Northern Territory	117,526	0.9	115,075	0.9	232,605	0.9
Australian Capital Territory	224,361	1.8	230,140	1.8	454,499	1.8
Australia(b)	12,545,154	100.0	12,877,635	100.0	25,422,788	100.0

Note: a. Based on place of usual residence. Excludes overseas visitors; b. Includes Other Territories; Source: Census 2021

Table 6. Generational population distribution in Australia, 2021 Census

	n	%
Gen alpha (2011-2021)	3,050,735	12.0
Gen Z (1996-2010)	4,626,947	18.2
Millennials (Gen Y) (1981-1995)	5,465,899	21.5
Gen X (1966-1980)	4,906,598	19.3
Baby Boomers (1946-1965)	5,465,899	21.5
Interwar (1945 or earlier)	1,957,555	7.5

Source: Australian Bureau of Statistics, Snapshot of Australia, Census 2021, n is calculated by multiplying percentages provided in source by Census 2021 population of 25,422,788 as this information was not provided on the website.

Table 7. Median income and income share by State and Territory

	Median total income 2020-21	Income Share		
	2020-21	Top 1% earners	Top 5% earners	Top 10% earners
Australia	$54,890	9.8%	22.5%	33.4%
New South Wales	$55,854	11.2%	24.3%	35.2%
Victoria	$54,088	9.8%	22.5%	33.5%
Queensland	$53,643	8.4%	20.6%	31.6%
South Australia	$52,767	8.0%	20.1%	30.7%
Western Australia	$58,591	9.5%	21.9%	32.9%
Tasmania	$50,130	7.5%	19.5%	30.1%
Northern Territory	$65,906	6.0%	16.9%	27.1%
Australian Capital Territory	$71,093	6.9%	18.6%	28.9%

Source: Australian Bureau of Statistics, Personal Income in Australia 2020-21 financial year

Table 8. Home ownership in Australia

	%
Owned outright	29.5%
Owned with mortgage	36.8%
Rented	29.1%

Source: Australian Bureau of Statistics, Housing Occupancy and Costs, 2019-20 financial year

Table 9. 1-digit level country of birth, Census 2021

1-digit level BPLP Country of Birth of Person	n	%
North-West Europe	1409655	6.8
Southern and Eastern Europe	649537	3.1
North Africa and the Middle East	417328	2.0
South-East Asia	997190	4.8
North-East Asia	813987	3.9
Southern and Central Asia	1070840	5.2
Americas	317489	1.5
Sub-Saharan Africa	349124	1.7
Supplementary codes	13977	0.1
Oceania and Antarctica	13594185	65.4
Not stated	1151475	5.5
Total	20784787	100.0

Source: Census 2021 Table Builder

Table 10. Industry 1-digit level, ANZSIC, Census 2021

1-digit level INDP Industry of Employment	n	%
Agriculture, Forestry and Fishing	282227	1.1
Mining	214759	0.8
Manufacturing	714759	2.8
Electricity, Gas, Water and Waste Services	134368	0.5
Construction	1067645	4.2
Wholesale Trade	311863	1.2
Retail Trade	1099617	4.3
Accommodation and Food Services	783737	3.1
Transport, Postal and Warehousing	546364	2.1
Information Media and Telecommunications	166872	0.7
Financial and Insurance Services	445121	1.8
Rental, Hiring and Real Estate Services	187621	0.7
Professional, Scientific and Technical Services	945094	3.7
Administrative and Support Services	388432	1.5
Public Administration and Safety	797031	3.1
Education and Training	1061320	4.2
Health Care and Social Assistance	1751717	6.9
Arts and Recreation Services	188479	0.7
Other Services	435459	1.7
Inadequately described	369939	1.5
Not stated	156996	0.6
Not applicable	**13373380**	**52.6**
Total	25422788	100.0

Source: Census 2021 Table Builder

Table 11. Occupation 1-digit level, ANZSCO, Census 2021

1-digit level OCCP Occupation	n	%
Managers	1645769	6.5
Professionals	2886921	11.4
Technicians and Trades Workers	1554313	6.1
Community and Personal Service Workers	1382205	5.4
Clerical and Administrative Workers	1525311	6.0
Sales Workers	986433	3.9
Machinery Operators and Drivers	755863	3.0
Labourers	1086120	4.3
Inadequately described	133815	0.5
Not stated	92659	0.4
Not applicable	**13373380**	**52.6**
Total	25422788	100.0

Source: Census 2021 Table Builder

Table 12. Population distribution by remoteness area, Census 2021

Remoteness Areas (National) (UR)	Male n	%	Female n	%	Total n	%
Major Cities of Australia	9007652	71.8	9307399	72.3	18315048	72.0
Inner Regional Australia	2233557	17.8	2302179	17.9	4535738	17.8
Outer Regional Australia	1029757	8.2	1019253	7.9	2049015	8.1
Remote Australia	148508	1.2	138860	1.1	287367	1.1
Very Remote Australia	96068	0.8	86619	0.7	182689	0.7
Migratory - Offshore - Shipping	753	0.0	107	0.0	864	0.0
No usual address	28851	0.2	23214	0.2	52069	0.2
Total	12545146	100.0	12877631	100.0	25422777	100.0

Source: Census 2021 Table Builder

Table 13. Population distribution by Labour force status, ALL persons, Census 2021

All persons	Male n	%	Female n	%	Total n	%
Employed, worked full-time	4344050	34.6	2751050	21.4	7095103	27.9
Employed, worked part-time	1399494	11.2	2563058	19.9	3962550	15.6
Employed, away from work	456667	3.6	535097	4.2	991758	3.9
Unemployed, looking for full-time work	223033	1.8	127163	1.0	350190	1.4
Unemployed, looking for part-time work	128215	1.0	168036	1.3	296250	1.2
Not in the labour force	2992063	23.9	3896023	30.3	6888081	27.1
Not stated	617785	4.9	583062	4.5	1200851	4.7
Not applicable	2383856	19.0	2254152	17.5	4638004	18.2
Total	12545154	100.0	12877635	100.0	25422788	100.0

Source: Census 2021 Table Builder

Table 14. Population distribution by Labour force status, 15-64 years old, Census 2021

15-64 years	Male n	%	Female n	%	Total n	%
Employed, worked full-time	4178515	51.5	2680359	32.3	6858877	41.8
Employed, worked part-time	1249232	15.4	2417667	29.2	3666901	22.4
Employed, away from work	423852	5.2	507668	6.1	931528	5.7
Unemployed, looking for full-time work	217544	2.7	125428	1.5	342984	2.1
Unemployed, looking for part-time work	121479	1.5	163561	2.0	285038	1.7
Not in the labour force	1446054	17.8	1977871	23.9	3423925	20.9
Not stated	481874	5.9	415560	5.0	897442	5.5
Total	8118550	100.0	8288114	100.0	16406695	100.0

Source: Census 2021 Table Builder

Cultural diversity

- 3.2% of the population are Aboriginal and/or Torres Strait islander (Census 2021)
- 27.6 per cent of the population were born overseas.
- Top 5 languages used at home, other than English, were Mandarin (2.7 per cent), Arabic (1.4 per cent), Vietnamese (1.3 per cent), Cantonese (1.2 per cent) and Punjabi (0.9 per cent).

Survey pre-analysis plan template

Project name*

[If different from survey name]

Project team members

[Include all project team and also specific who wrote the plan and who is going to analyse data as this would be helpful for business continuity and corporate knowledge)

Survey purpose and research questions

[This section should be very brief. Dot points are ok.]

Hypotheses*

[This section is not mandatory. Only have this if you have hypotheses.]

Survey design

This section should address the following:

- Sample selection and exclusion criteria: participant pool (estimated size, recruitment method) and eligibility/exclusions.
- Type of sampling method: probability, non-probability, online panel, river sampling
- Description of survey mode (online/mixed) and any relevant information on survey implementation
- Existing scales/questions used and any changes made and why

Conceptual/theoretical model

[This contains any concepts or theoretical models that are underpinning the survey. It can be as simple as we expect variable A and B to influence C and there may also be differences between group X and group Y. either text or a diagram would do.]

Main variables or scales

[This section should describe key variables and if there are scales. If existing validated questions are used, you can mention them here too].

Include definitions and derivation (if any) of study measures including”

- Main exposure/exploratory variables
- Outcome variables
- Covariates, potential confounders or mediators
- Sub groups you might want to analyse

Planned analysis

[This section can be as light or detailed as you like but should at least describe the main analyses you plan to undertake.]

- Detecting and dealing with poor quality data (e.g. would you do a speeder check, are there validation questions, sense checks- what would you do with suspect data?)
- Missing data- how would you deal with it (e.g. complete case analysis, imputation or sensitivity analysis).
- What are the main comparison groups?
- What types of analyses would be conducted and their sequence (only primarily descriptive or 2X2 analysis? Or would that be followed by modelling?, how would you avoid p-hacking if you are doing multiple tests. We suggest using chi-square/cross-tabs as information only and not present p-values, whether you would have thresholds for practical significance)
- How you would decide on final model(s)
- Analytical software and version

Planned tables and figures*

[Often it is useful to have dummy tables and figures for key results. This would be useful in reviewing the survey to ensure questions and data collected are in the right format and the survey would provide what you need]

Note: * sections are suggestions and not mandatory.

Further reading

Bradman, N, Sudman, S and Wansick B (2004). *Asking questions: The definitive guide to questionnaire design*, Jossy-Bass, San Francisco.

Collins, D, Blake M, D'Ardenne J and Gray M (2015). *Cognitive interviewing practice,* Sage Publications Ltd, Los Angeles.

Dillman, DA., Smyth JD and Christian LM (2014). *Internet, phone, mail and mixed-mode surveys: The tailored design method*, John Wiley & Sons Inc, Hoboken, New Jersey.

Robinson, SB and Leonard KF (2019). *Designing quality survey questions*, Sage Publication Ltd, Los Angeles.

Tourangeau, R, Rips LJ and Rasinski K (2000). *The Psychology of survey response*, Cambridge University Press, Cambridge.

Biemer PP, de Leeuw E, Eckman S, Edwards B, Kreuter F, Lyberg LE, Tucker NC and West BT (2017). *Total survey error in practice*, John Wiley and Sons, New Jersey.