# How can we ensure high quality data from online panels?

## Introduction – the problem

Online panel research – traditionally completed on PCs or laptops, but now increasingly on tablets and phones – has shown extraordinary growth in popularity over the past two decades.

The appeal of online surveys is easy to explain: it is quicker and cheaper than other forms of research, while still allowing clients to obtain relatively large datasets.

Online research, like face-to-face research, also allows for visual stimuli (pictures, video, etc.) which is not possible in traditional telephone research.

On the other side of the fence, the problems with online research are well known:

• Despite some online panels having impressive sizes, their make-up and recruitment methods mean they are unlikely to be representative of community views or behaviours.

• The same people tend to get surveyed over and over.

• There are often issues with data accuracy and integrity.

While the first two of these are widely recognised as inherent limitations of online methodologies, the last is causing increasing concern for research suppliers and clients alike.

**This paper addresses ways to maximise the data quality from online surveys using commercial panels.**

## What are the key concerns?

Online panelists complete surveys for points which can be redeemed for rewards such as cash and, prizes. While there is nothing wrong with this, clients need to understand that respondents' primary motivation is *to complete a survey as quickly as possible without being excluded or thrown out for "racing" through it*.

It's important for clients to understand that online panelists are different to those who may "opt-in" online or choose to complete a survey on a specific topic.

Those choosing to opt-in may be motivated either by interest in the subject matter or a sense of public spirit, whereas the online panelist is primarily driven by the chance to gain financial or other rewards.

In itself, that's not a problem. But there may be data quality issues caused by:

• Respondents who rush through surveys as quickly as possible (i.e., without necessarily even reading the questions fully or stopping to consider their response).

• Those who seek to cheat or "game" panels by setting up fake identities or creating automated "bots".

• Providing inaccurate responses to questions to progress through a survey.

• Respondents becoming conditioned to gaming screeners and quality control checks.

Taverner
RESEARCH GROUP

# How are online panels recruited?

There are a number of online panel providers in Australia, some general and others specific (e.g., B2B, regionally based, or industry-specific). The largest panels boast hundreds of thousands of members across the country, whom they recruit by:

- **Soliciting new panelists**. This can be done online, by phone, SMS, social media, etc.

- **Targeting members or clients of other organisations.** Some panelists are derived from other membership lists or customer loyalty programs, for example.

Airlines, supermarkets and ticketing companies (among others) have all set up customer research panels from their existing membership.

- **Tapping into non-profit or community organisation member lists.** Here the reward may be for the benefit of the organisation, respondent or both

There are also providers who form panels by amalgamating small panels into larger online panel 'marketplaces.

Panels may reach out to their peers/competitors to achieve larger sample sizes or difficult quotas, e.g. for geo-specific projects, or where low incidence rates mean that few people will qualify for a survey.

## How do we spot *speeding* or fake respondents?

Both panel providers and researchers can take steps to identify and weed out poor quality respondents. These include, but are not limited to:

- Identifying multiple surveys from the same IP address.

- Seeking out sequences of identical responses.

- Excluding responses if they are completed too quickly (generally defined as being >1 standard deviation less than the mean or median response time).

- Excluding respondents with one or more incidents of patterned responses or "straight-lining" (e.g., all responses in grid-style questions are the same).

- Excluding responses with poor quality open-ended comments – either those that make no sense within the question's context or are just gibberish.

- Including a CAPTCHA question at survey commencement to ensure a human is completing.

Responses that a client and panel provider agree are of poor quality are removed from the sample and replaced with other, more reputable responses. However this can sometimes be a fiddly, time-consuming and difficult process.

The bigger concern is that both respondents and bot-creators are aware of these fraud prevention measures and are becoming increasingly sophisticated in their efforts to circumvent them.

# Tips for clients to improve data quality

1. **Keep questions short.** There is substantial evidence that long questions (or preambles) encourage poor responses – and that a proportion of respondents will click any response to progress to the next question rapidly.

2. **Avoid long, matrix-style questions.** These are easy for respondents to "game" (or answer poorly). And many panelists are savvy enough not to "straight-line" their responses. Long, matrix-styled questions can also be lazy research practice and show little respect for the respondent.

3. **Mix up question styles.** This is more about good research practice than avoiding fraud. While mixing up question types creates a more enjoyable respondent experience; it also leads to less chance of fraud or thoughtless responses due to boredom.

4. **Add 'trick' (i.e., quality control) questions**. While helping weed out speeders, these questions should also fool the bots, and you can be explicit in explaining to respondents why these questions are being asked. Examples include:

a) "To continue the survey, please type a number between 30 and 40 in the following box."

b) "This question is designed to identify and exclude automated responses. Below are six photos: please click any that include a XXXXX."

c) "To ensure you are not a robot, ignore the question below, and click 'Not applicable' to continue the survey."

d) (Hidden within a matrix-style question) "Please tick 'XXXXX' (i.e., one of the answer options) on this line."

5. **Set clear guidelines on what is and isn't acceptable in relation to open-ended questions**. Given the nature of panel surveys, no researcher should expect long or detailed responses to open-enders. But they should at least relate to the question and make sense.

6. **Build the question into the answer frame** (as respondents often just read the answer frame and try to guess the question from there).

So, for example, instead of asking "How many coffees do you drink a day?", ask "Please select which is true for you?", with options such as "I drink coffee once a day", "I drink coffee twice a day", etc.