Data trust and data privacy: A brake on the data and digital dividend?

ANU Centre for Social Research and Methods

Professor Nicholas Biddle[1,2]; Professor Matthew Gray[1]; and Associate Professor Steven McEachern[1,3]

1 ANU Centre for Social Research and Methods Australian National University

2 National Data Advisory Committee

3 Australian Data Archive

19th October, 2022

Abstract

Data is increasingly available at scale and many of the fastest growing companies are built on data and data analytics. Governments are also increasingly using data for service delivery and to a lesser extent policy development and evaluation. Regulating and managing the increasing availability and use of data by the public, community and private sectors requires new approaches and laws. In April 2022 the Australian Parliament passed the *Data Availability and Transparency Act 2022* which allows, Australian Commonwealth bodies to share data. While the legislation and associated regulation is important, so are the levels of community data trust and attitudes to data privacy.

This paper reports data on Australian's attitudes to data trust and data privacy and how these have changed since October 2018 using data from the ANUpoll series of surveys collected in October 2018, October 2019, May 2020, August 2021 and August 2022. This provides information on how attitudes have changed during the COVID-19 period and during a period of rapid digitisation and increasing availability and use of data.

The data shows that trust in key institutions with regards to data privacy increased during the early stages of COVID-19 period, and has stayed high through to mid-2022. Australians also for the most part think governments should be sharing data with researchers (particularly in universities) and making use of data internally. However, support for such uses of data is slipping. Part of the response to these trends is to make sure that when data is used, it is done so in a way that maximises benefits to society. Collectively, the Australian research and policy community also needs to better understand who is reluctant for their data to be used, why they are reluctant, and what the possible responses and safeguards might be to make better use of such resources whilst still maintaining a social licence.

Acknowledgements

The ANU Centre for Social Research and Methods COVID-19 Impact Monitoring series has received funding from the Australian Institute of Health and Welfare. The authors would like to particularly thank Matthew James, Deputy CEO Australian Institute of Health and Welfare, for comments on reports and survey instruments in this series. The opinions and conclusions in this paper should, however, be attributed to the authors only. The survey data is available for download through the Australian Data Archive (http://dx.doi.org/10.26193/FCZGOK)

1 Introduction

Data has been variously described as the new gold,[1] the new oil[2], or the lifeblood of capitalism[3]. This has been said so many times, that it has become a truism. Many of the fastest growing companies either are built on harvesting large datasets or are heavily reliant on data analytics. The use of data by governments is also growing rapidly which has the potential to increase the effectiveness of evaluation (Crato and Paruolo 2019) performance measurement and the delivery of services.

Although new barriers to access are often put in place or existing or existing barriers put in place, data is increasingly available at scale and in ways where the data has been or can be linked with other data sources. However, barriers to accessing data still exist and the quality of the data that is available is not always adequately interrogated (Zaveri et al. 2016; Reid et al. 2017; Sakshaug and Antoni 2017). Despite these limitations, there is significant scope to substantially increase the data-driven insights that can help guide the decision making of businesses, community organisations, governments at all levels, and ultimately citizens.

At the same time, it is essential that limits to be placed on the use of data by researchers, governments, and businesses. There are many examples both in Australia and other countries where the misuse of data by government or by businesses has had negative consequences for citizens. Data used inappropriately can exacerbate biases within society by gender, ethnicity, age, location, or other characteristics (Haijian et al. 2016; and Hoffman 2019). Those who hold data can exploit monopoly power, at the expense of new entrants into markets (Newman 2014), and data shared without privacy protections can expose individuals or businesses to adverse effects (Jain et al. 2016).

In the Australian context, one high profile example of the problematic use of data has been the Australian government's use of automated data matching across social security and taxation records to issue debt notices in a process that has become known as "robodebt". This process has been highly contentious, deeply flawed and ultimately found by a court in 2019 to be unlawful, resulting in the Commonwealth government needing to refund $751 million to over 370,000 individuals (Robert, 2020).

Over the last three years in there have been two developments that make it particularly timely to make far greater use of existing datasets, particularly those that had not initially been created for research purposes. First, the COVID-19 pandemic, 2019/20 Black Summer bushfires, and La Niña induced floods have highlighted the need to make rapid, informed decision making, and to do so in a way that is transparent, and data driven. At the same time, the passage of the Data Availability and Transparency Act[4] has accelerated the development of infrastructure that can help make government data available to both government agencies and university researchers in a safe, privacy preserving way.

Introduced by the former Coalition Government and passed by Parliament in April 2022, the *Data Availability and Transparency Act 2022* established the 'DATA Scheme' which allows Commonwealth bodies to share data with Accredited Users. These can be public servants using the data for policy development, research and service provisions, or researchers based at Australian universities accredited under the DATA Scheme. The Act also established the role of the National Data Commissioner, who is the regulator for the DATA Scheme, but also (in part through their office) provides education and support for best practice data handling and sharing.

It is therefore timely to reflect on the role of government and other data in supporting businesses, not for profit organisations, public policy development, and the delivery of services. In February 2022 the then Commonwealth Treasurer Josh Frydenberg in setting the terms of reference for the second five-year inquiry into the productivity performance of Australia to be undertaken by the Productivity Commission required the Commission to amongst other things 'Identify priority sectors for reform (including but not limited to data and digital innovation and workforce skills) and benchmark Australian priority sectors against international comparators to quantify the required improvement.'[5] The Productivity Commission's Second Interim Report as part of this inquiry focuses in particular on 'how digital technology and data can be used to improve Australia's productivity.'[6] The Productivity Commission report argues that technologies such as artificial intelligence (AI), the internet of things (IoT), robotic automation and big data analytics could further revolutionise how businesses operate across the economy.

It is widely recognised that, from a public sector perspective, the better use of data and appropriate technologies would improve public policy and service delivery (Productivity Commission 2022; Thodey 2019) and that governments can make better-informed decisions about policy design and implementation, both at the system level and to address local community needs. However, the interim report from the Productivity Commission argues that there are several factors that could limit further adoption of data and digital technology among Australian businesses. This includes inadequate internet, lack of skills, low awareness and uncertainty about benefits, security concerns, cost, and legacy systems. These limitations are arguably also present for the government and community sectors.

At the time of writing this report, Australia has been reminded of these genuine concerns by the "Optus data breach", where the records of what might turn out to be over 10 million Australians were illegally downloaded, including some records with very sensitive information. The breach has re-raised the question of why so much data needed to be retained by the company, with some arguing that companies like Optus 'are collecting – and keeping – much more personal information than they need without a truly legitimate commercial or legal purpose.'[7]

The Productivity Commission in its report identifies three enablers where government investments and policies can provide a foundation for adopting productivity-enhancing digital and data tools. One of these is increased data sharing and integration. It is certainly the case that many of the data assets held by governments could be integrated in ways that they haven't been before, including across jurisdictions or between different levels of government. For example, during the COVID-19 period, whether or not someone has received a COVID-19 vaccine was recorded through the Australian Immunisation Registry. Whether or not someone has tested positive for COVID-19 is collected by individual states and territories. These two pieces of information have not been systematically linked at the individual level, despite the obvious benefit in terms of understanding the ongoing distribution of vaccine uptake in Australia and the effectiveness of the current set of vaccines in the Australian context.

Any new data that is integrated, as well as much of the existing data that is available, will only be put to maximum use though if it is shared with those who are best able to add value to it. This could be sharing within government or across levels of government, with universities or other research institutions, or with the commercial sector. However, sharing of this data can create real risk, especially if not done safely and securely. Here a distinction needs to be made between data sharing where there is an agreement between the data custodian and the data

user in terms of how the data can be used, and data release where data is essentially openly available with minimal restrictions on who can use the data and how it can be used.

Whether data is shared or released more openly, it should be done in a way that takes into account risk to the individuals (or businesses/organisations) that the data is about, and that ensures maximum benefit from that data. The Productivity Commission makes the point that 'community trust in new applications of technology is critical for future uptake, as businesses and governments need to maintain their social licence to deliver digital and data-enabled services.' Without that social licence, and without a careful weighing of risks and benefits, there is a real chance that low public acceptance of data integration and sharing will inhibit Australia's ability to take advantage of the data and digital dividend.

Careful monitoring of public attitudes and behaviours related to data integration and data sharing is therefore an important component of the data and digital ecosystem. Since October 2018, the ANU Centre for Social Research and Methods has been monitoring attitudes to data trust and data privacy, through the ANUpoll series of surveys. A number of questions in the most recent wave of data collection in August 2022 focus on data trust and data privacy and questions were also included in the May 2020 and August 2021 waves.

From April 2020 the ANUpoll series have been used to collect data for the ANU Centre for Social Research and Methods COVID-19 Impact Monitoring Survey series.[8] While the primary focus of these surveys has been on tracking the impact of COVID-19 on Australians and related attitudes, they have also included questions on a range of other topics. While the ANUpoll series of surveys is longitudinal, in this paper the data is treated as a series of repeated cross-sections.

The August 2022 survey collected data from 3,510 Australians aged 18 years and over.[9] For this survey, the Social Research Centre (on behalf of the ANU) collected data online and through Computer Assisted Telephone Interviewing (CATI) from respondents on the Life in Australia™ panel. Around 3.5 per cent of interviews were collected via CATI.[10] A total of 4,294 panel members were invited to take part in the August 2022 survey, leading to a wave-specific completion rate of 81.7 per cent.[11]

The aim of this paper is to summarise the survey results, including with a comparison of previous waves of data collection on related topics. Section 2 looks at trust in institutions, and how that has changed through time. Section 3 looks at concerns amongst the general public regarding their own data, and Section 4 presents results on how people think government data should be used. Section 5 provides concluding comments.

2 Trust in institutions

The October 2018, May 2020, August 2021, and August 2022 ANUpoll surveys included questions about attitudes toward data trust and data privacy. Respondents were asked: 'On a scale of 1 to 10, where 1 is no trust at all and 10 is trust completely, how much would you trust the following types of organisations to maintain the privacy of your data?'. Respondents were asked about eight types of organisations as listed below, and the order in which the organisations were presented to the respondent was randomised:

a) The Commonwealth Government in general
b) The State / Territory Government where you live
c) Banks and other financial institutions
d) Social media companies (for example Facebook, Twitter, Google)

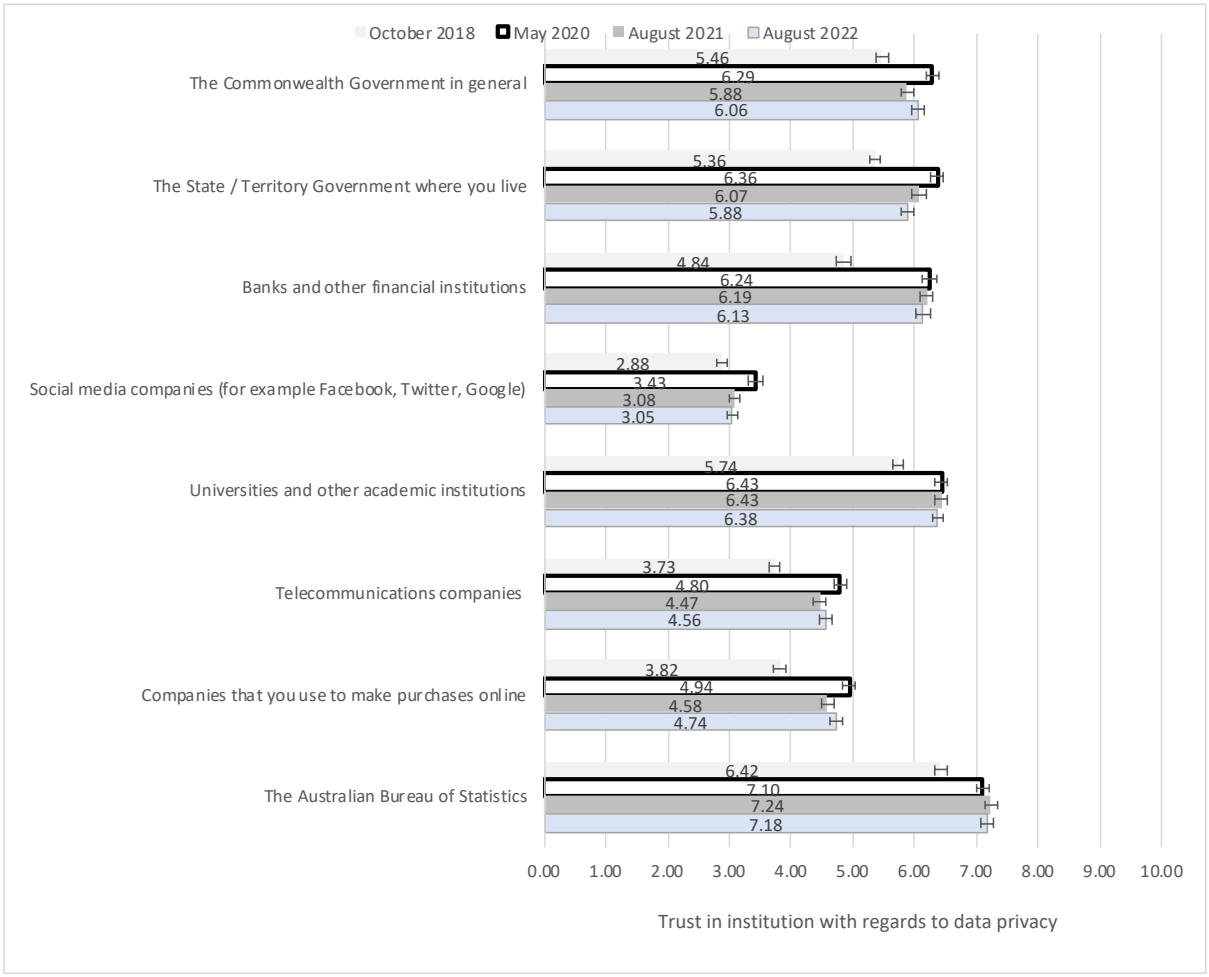
e) Universities and other academic institutions
f) Telecommunications companies
g) Companies that you use to make purchases online
h) The Australian Bureau of Statistics

For some analyses, it is useful to combine trust across multiple types of institutions. A principal components analysis suggests that a measure of trust in institutions through a single index with equal weights is appropriate.[12] Therefore the index of trust in data privacy is simply the average value of trust in the eight institutions asked about. A higher value of the index indicating that the individual has a higher overall level of trust in the ability of the different types of organisations to maintain their data privacy.

Following an increase in the overall level of trust in a range of types of organisations to maintain data privacy between October 2018 and May 2020 from 4.78 to 5.70, there was a decline between May 2020 and August 2021 to 5.49. There has been virtually no change between August 2021 and August 2022 when the average value was 5.50. Trust in August 2022 is still, however, well above that in October 2018.

There were some fluctuations in the levels of trust in the specific organisations asked about across the period (Figure 1). Between August 2021 and August 2022 there were small increases in trust in the Commonwealth Government and online shopping companies, but a small decline in trust in state/territory governments.[13] The institution that continues to have the highest level of trust to maintain data privacy is the Australian Bureau of Statistics and the type of organisation with the lowest level of trust to maintain data privacy is social media companies. Furthermore, trust in all organisations/institutions to maintain data privacy was higher in August 2022 than it was pre-COVID and the difference is statistically significant for all the types of institutions/organisations asked about.

Figure 1 Average level of trust in institutions to maintain data privacy – October 2018 to August 2022



Note: The "whiskers" on the lines indicate the 95 per cent confidence intervals for the estimate.

Source: ANUpoll October 2018, May 2020, August 2021, and August 2022.

Trust in institutions regarding data privacy varies substantially across the population. In order to understand the differences a regression model is used to estimate the association with various demographic, socio-economic and geographic characteristics are associated with the aggregate index of trust in institutions in maintaining data (described above). Because the index of trust in institutions is a continuous variable an ordinary least squares linear regression model is appropriate.

The results of the regression model are presented in Table 1. The table reports the regression coefficients which can be interpreted as the difference in the average level of trust holding constant other observed characteristics.

There are no differences between males and females in trust in institutions regarding data privacy, but there were significant differences by age, with a lower level of trust for young Australians (particularly those aged under the age of 35) and a higher level of trust for those aged 75 years and over. Those who speak a language other than English at home had a higher level of trust. There were large differences by education, with those who had not completed Year 12 and those with a Certificate III/IV having a lower level of trust. Income also mattered,

with those in the lowest household income quintile having much lower levels of trust than the rest of the population.

Table 1 Demographic, geographic, and socioeconomic factors associated with trust in institutions regarding data privacy, August 2022

Explanatory variables	Coeff.	Statistical significance
Female	0.054	
Aged 18 to 24 years	-0.201	
Aged 25 to 34 years	-0.301	***
Aged 45 to 54 years	-0.057	
Aged 55 to 64 years	0.013	
Aged 65 to 74 years	0.136	
Aged 75 years plus	0.823	***
Indigenous	0.102	
Born overseas in a main English-speaking country	-0.170	
Born overseas in a non-English speaking country	0.026	
Speaks a language other than English at home	0.311	**
Has not completed Year 12 or post-school qualification	-0.328	**
Has a post graduate degree	-0.142	
Has an undergraduate degree	-0.159	
Has a Certificate III/IV, Diploma or Associate Degree	-0.228	**
Lives in the most disadvantaged areas (1st quintile)	0.075	
Lives in next most disadvantaged areas (2nd quintile)	0.082	
Lives in next most advantaged areas (4th quintile)	-0.042	
Lives in the most advantaged areas (5th quintile)	0.064	
Lives outside of a capital city	-0.087	
Lives in lowest income household (1st quintile)	-0.367	***
Lives in next lowest income household (2nd quintile)	-0.101	
Lives in next highest income household (4th quintile)	0.057	
Lives in highest income household (5th quintile)	0.163	
Constant	5.661	***
Sample size	3,155	

Notes: Ordinary Least Squares regression model. The base case individual is male; aged 35 to 44 years; non-Indigenous; born in Australia; does not speak a language other than English at home; has completed Year 12 but does not have a post-graduate degree; lives in neither an advantaged or disadvantaged suburb (third quintile); lives in a capital city; lives in neither a high-income or low-income household (third quintile).

Coefficients that are statistically significant at the 1 per cent level of significance are labelled ***; those significant at the 5 per cent level of significance are labelled **, and those significant at the 10 per cent level of significance are labelled *

Source: ANUpoll August 2022

3 Level of concern regarding data and personal information

Respondents to the August 2022 survey were asked to indicate the extent of agreement or disagreement with the following statements:

a) You are concerned that your online personal information is not kept secure by websites
b) You are concerned that your online personal information is not kept secure by public authorities
c) You avoid disclosing personal information online
d) You believe the risk of becoming a victim of cybercrime is increasing
e) You are able to protect yourself sufficiently against cybercrime, e.g. by using antivirus software on [reverse coded when constructing the index measure].

The statements were randomised, with response options of "totally agree; tend to agree; tend to disagree; and totally disagree".

Respondents were also asked: 'Cybercrimes can include many different types of criminal activity. How concerned are you personally about experiencing or being a victim of the following situations:

a) Identity theft (somebody stealing your personal data and impersonating you)
b) Receiving fraudulent emails or phone calls asking for your personal details (including access to your computer, logins, banking or payment information).

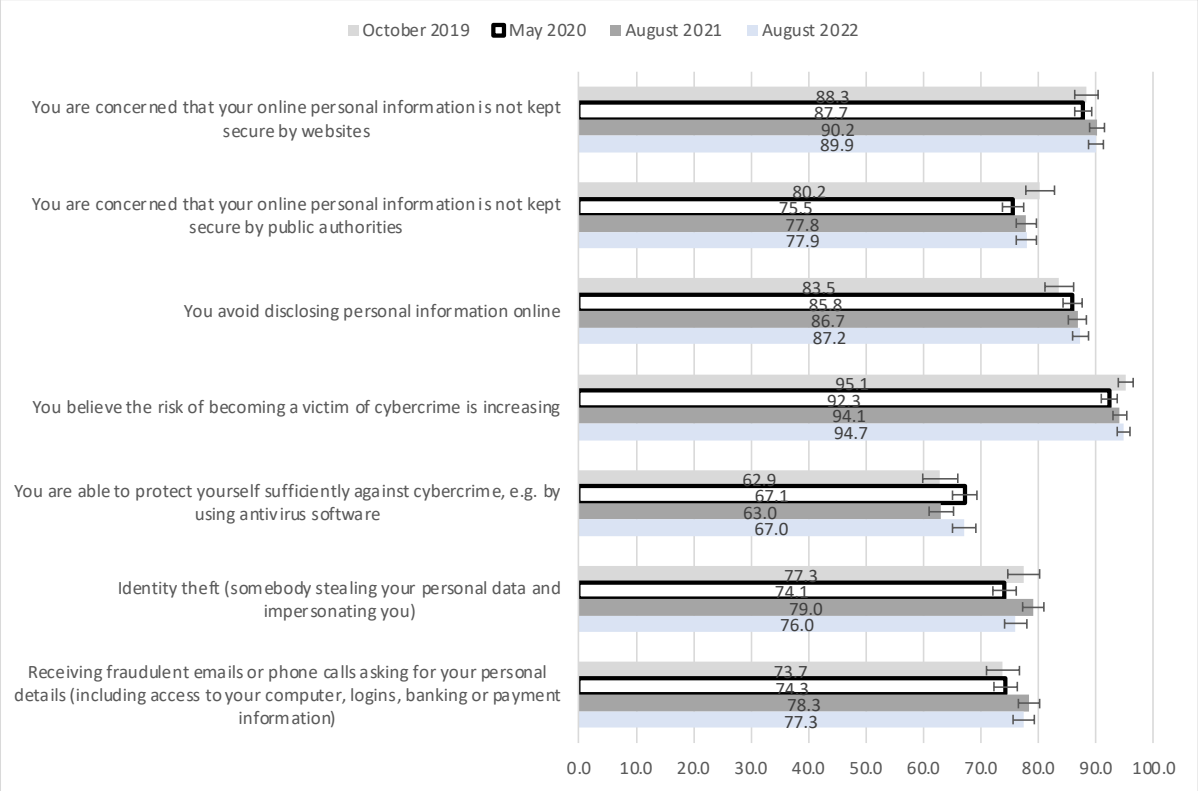
Response options for this question were "very concerned; fairly concerned; not very concerned; and not at all concerned". The order in which the scenarios were presented to respondent was randomised.

In order to measure the overall level of concern about the security of personal data, an index was produced which combined the seven questions on level of concern for data security and personal information (reverse coded where appropriate). The index varies from a value of 7 for those who were least concerned about personal information and data, to 28 for those who were most concerned, with an average of 21.3. These questions were also asked in October 2019, May 2020, and August 2021, with the index of concerns similar in August 2021 (21.4), which itself was an increase from May 2020 (from 20.82 to 21.40).

Figure 2 shows the proportion of Australians who are concerned about the security of their personal data and information when asked in October 2019, May 2020, August 2021, and August 2022. Between October 2019 and May 2020 there was little change in the proportion of Australians who were concerned about the different potential threats to the security of their personal information and data. The main changes were a reduction in the proportion who were concerned that their online personal information is not kept secure by public authorities and an increase in being concerned about not being able to protect yourself sufficiently against cybercrime.

Keeping in mind that a higher value for the question on 'You are able to protect yourself…' indicates a lower level of concern, Figure 2 shows that for none of the aspects of information security asked about was there a decline in concern between May 2020 and August 2021. The biggest relative increase between May 2020 and August 2021 was for the proportion of people who were concerned or very concerned about 'Identity theft (somebody stealing your personal data and impersonating you)' or 'Receiving fraudulent emails or phone calls asking for your personal details …'. Between August 2021 and August 2021, the proportion of Australians who felt they were able to 'protect yourself sufficiently against cybercrime' increased and the proportion who said they were concerned about identity theft decreased.

Figure 2 Per cent of Australians who agreed or totally agreed that they are concerned about the security of their personal data and information, October 2019 to August 2022



Note: The "whiskers" on the lines indicate the 95 per cent confidence intervals for the estimate.

Source: ANUpoll October 2019, May 2020, August 2021, and August 2022.

4 Government use of data

Taken together, results presented in the previous two sections showed that compared to the pre-COVID period, there has been an increase in the level of trust in key institutions regarding data privacy. Although not as dramatic, there has also been a slight decrease with regards to the level of concern about key aspects of the security of personal data and information. In many ways this bodes well for taking advantage of the data and digital dividend. That does not mean, however, that the Australian public is necessarily supportive of data being used in research and policy making and as discussed in this section, there is some evidence that people are becoming less supportive of data being used in such a way.

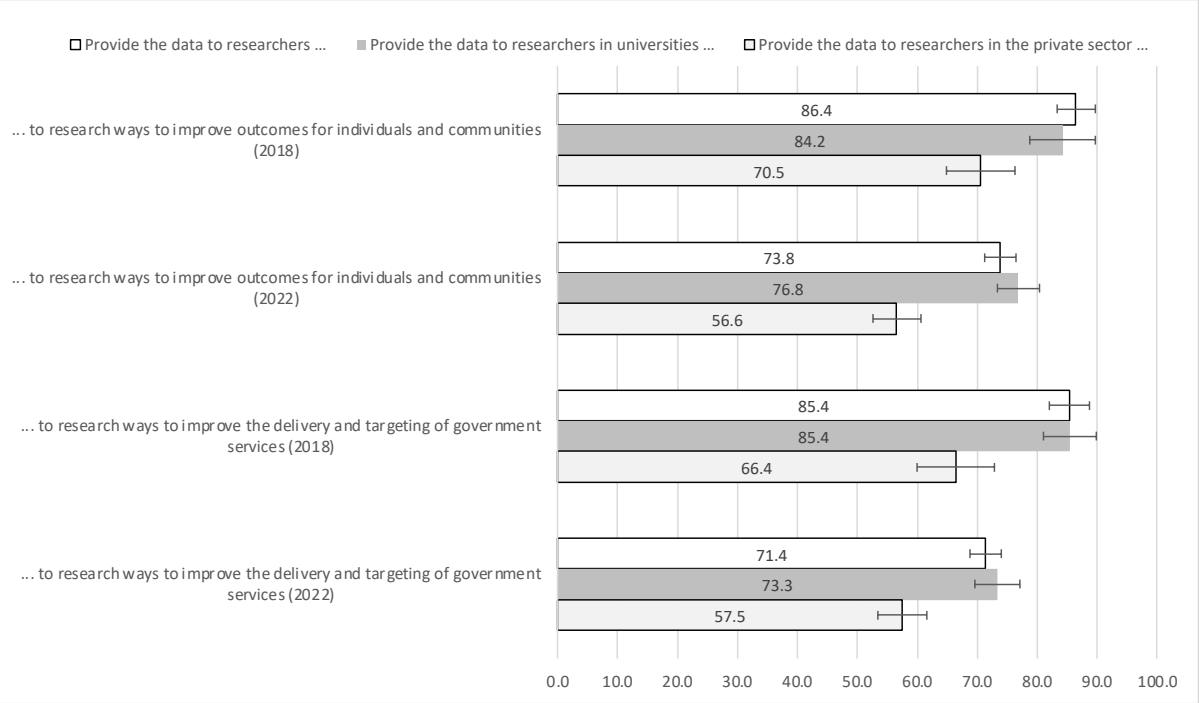
In the October 2018 and August 2022 survey respondents were first reminded that 'Governments across Australia collect a range of information on Australian residents.' Respondent were then asked, 'On the whole, do you think the Commonwealth Government should or should not be able to do the following?' with six potential uses of data. The first two of these relate to sharing of data with researchers:

a) Provide the data to researchers to research ways to improve outcomes for individuals and communities
b) Provide the data to researchers to research ways to improve the delivery and targeting of government services

A survey experiment was also included with these questions. For half of the sample (selected at random), it was not specified where the researchers were located, whereas for one-quarter of the sample it was specified that data is provided to researchers '*in universities'* with the remaining quarter specified that the data is provided to researchers *'in the private sector.*' For both questions (Figure 3), there was a slightly higher level of support when it is specified that the data is for "researchers in universities" rather than generically – from 73.8 per cent in the control group saying it definitely or probably should to 76.8 per cent when it is for researchers in universities for Question (a) and from 71.4 to 73.3 per cent for Question (b). However, neither of these two differences are statistically significant.

There is, however, a very large difference between those who were asked about sharing data in general and those who were asked about sharing data with researchers in the private sector. Only 56.6 per cent of respondents think that governments definitely or probably should 'Provide the data to researchers in the private sector to research ways to improve outcomes for individuals and communities' and only 57.5 per cent think that governments should 'Provide the data to researchers in the private sector to research ways to improve the delivery and targeting of government services.'

Figure 3 Per cent of Australians who think that data definitely or probably should be shared with researchers for specific purposes, by researcher type, October 2018 and August 2022



Note: The "whiskers" on the lines indicate the 95 per cent confidence intervals for the estimate.

Source: ANUpoll October 2018 and August 2022.

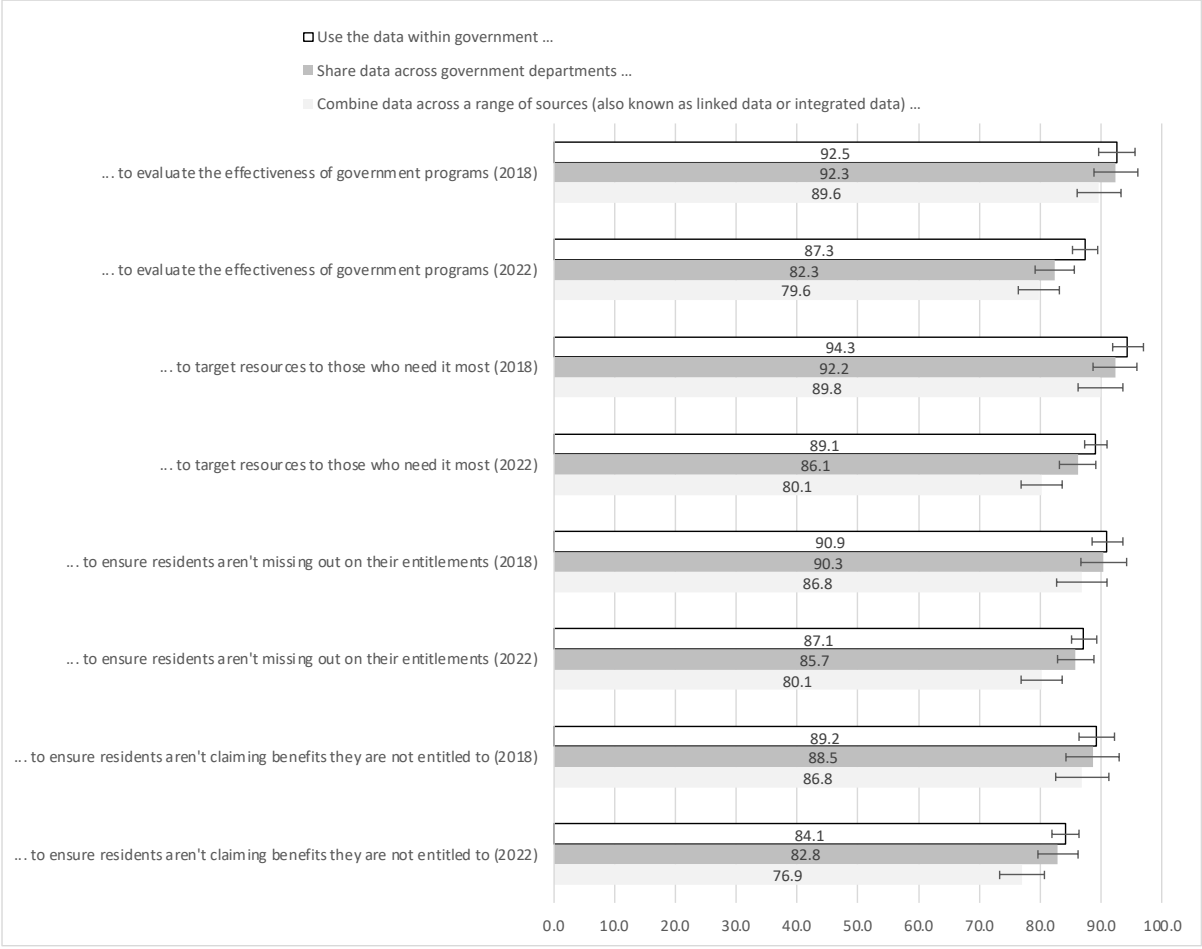
The remaining four questions about the use of data focus on how data should be used within government. There are four separate uses of data asked about, as follows:

c) evaluate the effectiveness of government programs
d) target resources to those who need it most
e) ensure residents aren't missing out on their entitlements
f) ensure residents aren't claiming benefits they are not entitled to

One half of the sample were simply asked whether respondents think governments should or should not '*Use the data within government*' for each of the four uses. However, for one-quarter of the sample, respondents were asked whether governments should '*Share data across government departments*' with the remaining quarter asked whether governments should '*Combine data across a range of sources (also known as linked data or integrated data)*' for each of the four uses. Figure 4 gives the per cent of Australians who think governments probably or definitely should be able to use data in each of the four ways, presented separately by the different types of data usage and by year.

For the most part, Australians are most supportive of governments using data within government generically, and least supportive of combining data across a range of sources. Sharing data across government departments has support that is somewhere in between.

Figure 4 Per cent of Australians who think that data definitely or probably should be used within government for specific purposes, by researcher type, October 2018 and August 2022



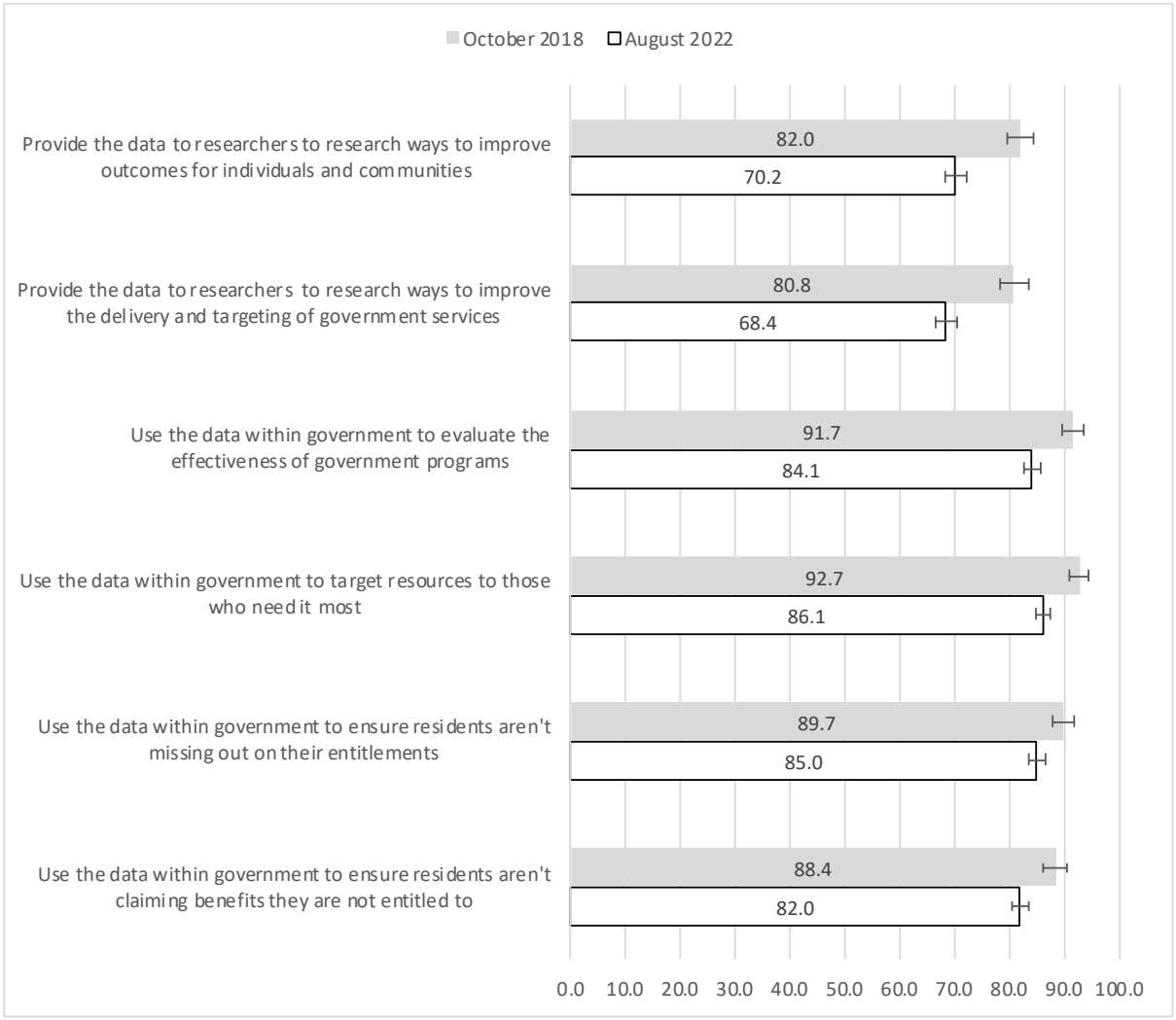
Note: The "whiskers" on the lines indicate the 95 per cent confidence intervals for the estimate.

Source: ANUpoll October 2018 and August 2022.

The previous two figures showed that, although the randomly allocated question wording had a significant effect on people's responses, that the differences were quite similar in 2018 and 2022. The three groups (control and two treatments) can therefore be combined into a single figure in order to make simpler conclusions about change through time (Figure 5). Across all six uses of data, support amongst the general public has declined between October 2018 and

August 2022. In particular, people are much less supportive of data being provided to researchers, with declines by more than ten percentage points over the approximately four years between data collections.

Figure 5 Per cent of Australians who think that data definitely or probably should be shared with researchers or used within government for specific purposes, October 2018 and August 2022



Note: The "whiskers" on the lines indicate the 95 per cent confidence intervals for the estimate.

Source: ANUpoll October 2018 and August 2022.

One of the reasons for this reluctance among the Australian public towards the Australian Government using and sharing data is a small decline in belief in the government's competence regarding data privacy. The last question in the data privacy module began with 'Following are a number of statements about the Australian Government and the data it holds about Australian residents. To what extent do you agree or disagree that the Australian Government…?' Respondents were then about four aspects of competency with data, as summarised in Figure 6.

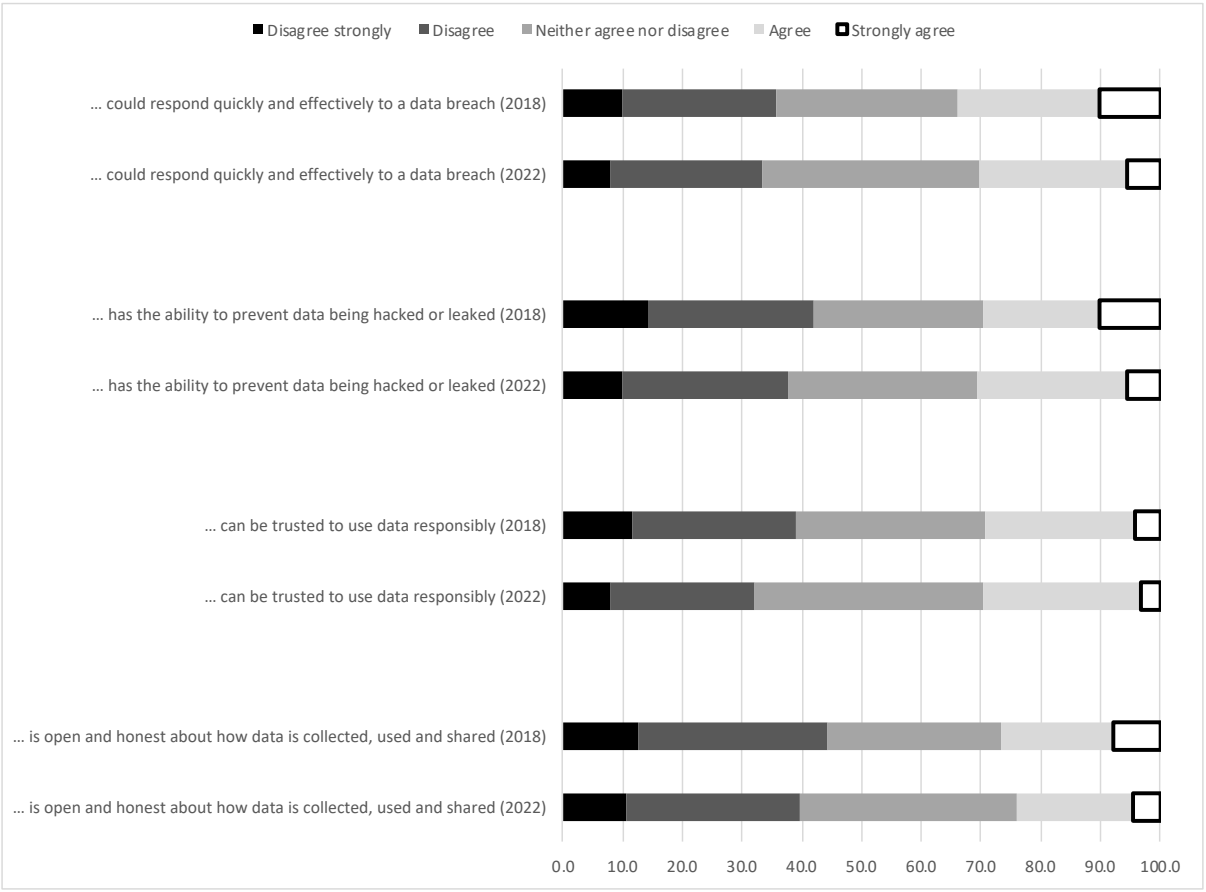
Focusing firstly on results from 2022, less than one-quarter of Australians (23.9 per cent) agree or strongly agree that the Australian Government 'is open and honest about how data is collected, used and shared.' In addition, less than one-third also agree or strongly agree that

the Australian Government 'can be trusted to use data responsibly' (29.6 per cent), 'could respond quickly and effectively to a data breach' (30.3 per cent), and 'has the ability to prevent data being hacked or leaked' (30.8 per cent).

This quite negative view of the "data competence" of the Australian Government is reasonably stable through time. While none of the views substantially improved, two stayed reasonably constant (ability to present hacks and can be trusted to use data responsibly). Views on the other two aspects of competency appear to have worsened. Fewer Australians agree that governments 'could respond quickly and effectively to a data breach' – down from 34.0 per cent in October 2018 to 30.3 per cent in August 2022, with a halving of those who strongly agree – from 10.0 to 5.5 per cent. There was also a decline in the per cent of Australians who think the Australia government is 'open and honest about how data is collected, used and shared', decreasing from 26.8 per cent in October 2018 to 23.9 per cent in August 2022.

An important point to note regarding the trends in these measures is that these declines are mainly due to an increase in the per cent of Australians who 'neither agree nor disagree.' The per cent of Australians who disagree strongly or disagree with these statements has also decreased.

Figure 6 Per cent of Australians by their views on abilities of Australian Government regarding data that it holds, October 2018 and August 2022



Note: The "whiskers" on the lines indicate the 95 per cent confidence intervals for the estimate.

Source: ANUpoll October 2018 and August 2022.

5 Concluding comments

On the 22nd of September 2022, the Australian telecommunications company Optus disclosed a security breach involving personal data of what may end up being around 10 million current or former customers.[14] While the breach did not involve government-held data, it did involve a significant amount of data generated for government administrative purposes and asked for by Optus, including passport, drivers licence, and Medicare numbers. Furthermore, the breach has called into focus the Commonwealth and state/territory legislation around data both specifically in terms of what Optus can, can't and has to do with data, but also more broadly.

About a month before this most recent data breach, the Productivity Commission released a second interim report as part of its 5 Year Productivity Inquiry, with this second report focusing on Australia's data and digital dividend. While acknowledging the risks related to collecting, using, and sharing customer/citizen data, the report by the Productivity Commission has a much stronger focus on the opportunities and benefits.

These two seemingly separate events highlight the real tension that exists with administrative data broadly, and government created data specifically. Data can help businesses make profits and enhance the customer experience. It can also make government decisions and the delivery of services more efficient and affective. However, misuse of data, or release of data with malicious intent can have real-world impacts on these same customers and decisions.

It is important, therefore, to keep trying through cultural change and legislation to get the balance right. That balance can better able to be achieved when public opinion on data use, access, and sharing is taken into account. This paper provides some positive news with regards to data and key institutions, but also some challenging trends for government and researchers.

On the one hand, trust in key institutions with regards to data privacy increased during the early stages of COVID-19 period, and has stayed high through to mid-2022. The Australian Bureau of Statistics is the most trusted of the institutions asked about, but the Commonwealth Government and universities are also reasonably well trusted. Much improvement is still possible, but there is a solid base to build upon.

Australians also for the most part think governments should be sharing data with researchers (particularly in universities) and making use of data internally. However, support for such uses of data is slipping. There has been a drop by half in the per cent of Australians who think governments definitely should provide 'data to researchers to research ways to improve outcomes for individuals and communities' since October 2018 (from 28.0 to 14.1 per cent) and an almost as large a decline in the per cent of Australians who thought that governments themselves should use data 'within government to evaluate the effectiveness of government programs' from (41.6 per cent to 25.5 per cent). The data collection methodology for these two surveys is almost exactly the same, the sample is similar, and these declines are far in excess of what is expected in terms of random variation in data. Australians are less comfortable in their data being shared and used than they were four short years ago.

One of the reasons for this reluctance to use data is that a low percentage of Australians and fewer Australians than in 2018 agreed that governments 'could respond quickly and effectively to a data breach' – down from 34.0 per cent in October 2019 to 30.3 per cent in August 2022, with a halving of those who strongly agree – from 10.0 to 5.5 per cent.

Part of the response to these trends is to make sure that when data is used, it is done so in a way that maximises benefits to society. Collectively, the Australian research and policy

community also needs to better understand who is reluctant for their data to be used, why they are reluctant, and what the possible responses and safeguards might be to make better use of such resources whilst still maintaining a social licence.

References

Crato, N. and P. Paruolo (2019). *Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design*. *Springer Open*.

Jain, P., M. Gyanchandani and N. Khare (2016). 'Big data privacy: a technological perspective and review.' *Journal of Big Data*, 3(1): 1-25.

Newman, N. (2014). 'Search, antitrust, and the economics of the control of user data.' *Yale Journal on Regulation*, 31(3):: 401-454.

Reid, G., F. Zabala. and A. Holmberg (2017). 'Extending TSE to Administrative Data: A Quality Framework and Case Studies from Stats NZ.' *Journal of Official Statistics (JOS)*, 33(2): 477-511.

Robert, S. *(2020).* Doorstop interview regarding changes to the Income Compliance Program, 29 May 2020 [Transcript]. https://minister.servicesaustralia.gov.au/transcripts/2020-05-29-doorstop-interview-regarding-changes-income-compliance-program

Sakshaug, J.W. and M. Antoni (2017). 'Errors in linking survey and administrative data.' In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, N. Tucker, B. West (eds) *Total Survey Error in Practice*, pp.557-573.

Thodey, D. I., (Chair). (2019). Department of Prime Minister and Cabinet, Canberra.

Zaveri, A., A. Rula, A. Maurino, R. Pietrobon, J. Lehmann. and S. Auer (2016). 'Quality assessment for linked data: A survey.' *Semantic Web*, 7(1): 63-93.

Endnotes

1 https://www.credit-suisse.com/about-us-news/en/articles/news-and-expertise/mark-cuban-data-is-the-new-gold-201706.html

2 https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

3 https://www.theguardian.com/technology/2018/jan/31/data-laws-corporate-america-capitalism

4 https://www.datacommissioner.gov.au/law/dat-act

5 https://www.pc.gov.au/inquiries/current/productivity/terms-of-reference

6 https://www.pc.gov.au/inquiries/current/productivity/interim2-data-digital

7 https://theconversation.com/optus-says-it-needed-to-keep-identity-data-for-six-years-but-did-it-really-191498

8 https://csrm.cass.anu.edu.au/research/publications/covid-19

9 The unit record survey data is available for download through the Australian Data Archive.

10 The contact methodology adopted for the online Life in Australia™ members is an initial survey invitation via email and SMS (where available), followed by multiple email reminders and a reminder SMS. Telephone follow up of panel members who have not yet completed the survey commenced in the second week of fieldwork and consisted of reminder calls encouraging completion of the online survey. The contact methodology for offline Life in Australia™ members was an initial SMS (where available), followed by an extended call-cycle over a two-week period. A reminder SMS was also sent in the second week of fieldwork.

11 The cumulative response rate (CUMRR2) takes account of non-response at each point. It is the product of the recruitment rate (RECR), the profile rate (PROR), the retention rate (RETR) and the completion rate: CUMRR2 = RECR × PROR × RETR × COMR. The recruitment rate is the rate at which eligible individuals agree to join the panel. The profile rate is the rate at which initially consenting individuals complete the panel profile, thus joining the panel. The retention rate is the proportion of active panellists at the time of this survey out of all those who joined the panel. The cumulative response rate for this survey is around 6.8 per cent (0.124 * 0.917 * 0.726 * 0.817 * 100).

12 The Eigenvalue for the first component 4.35 (explaining 54.4 per cent of the variation) and the Eigenvalue or the second component is 1.25. All eight of the variables correlated at a similar level with the first component (minimum Eigenvalue of 0.27 and a maximum of 0.40).

13 These differences are statistically significant when tested using the linked longitudinal sample

14 https://www.abc.net.au/news/2022-10-01/optus-data-hack-australians-waiting/101486874