

spaced nitrogen atoms at the edges seems to be crucial for this phenomenon to occur.

Simulations that Blackwell *et al.* performed to complement their experiments suggest that the nitrogen atoms at the edges of the nanoribbons form strong bonds with the substrate. The most stable configuration of the ribbons seems to be one with substantial corrugation, with the nitrogen atoms sitting at the lowest positions. As a result, the edge states, which tend to localize in the carbon segments, are decoupled from the surface. These edge states can thus be clearly visualized, even when the nanoribbons are adsorbed on gold.

Although the edges of graphene nanoribbons have previously been altered to decouple them from the substrate<sup>9</sup>, the technique of substituting the carbon atoms with nitrogen ones has the advantage of leaving the electronic and magnetic properties of the nanoribbon almost unmodified. And it has another benefit: nitrogen atoms can be used to probe the magnetism of the edge states. The spins of the electrons are all aligned along the edges of the nanoribbon, including those at the nitrogen sites. This affects other electron states, particularly those of the outer orbitals of the nitrogen atoms (Fig. 1), which do not participate in binding to the carbon structure. The spin directions of the two electrons in these outer orbitals have different energies – one is higher than the other. This difference can be detected with the scanning tunnelling microscope and represents direct evidence of the magnetism of the long zigzag edges in these graphene nanoribbons.

The work by Blackwell and colleagues offers exciting opportunities to study the magnetic properties of graphene nanostructures grown on metal surfaces. Future research will determine whether the observed surface-decoupling mechanism also applies to other structures and substrates. Similarly, the use of nitrogen as a local probe will probably inspire a search for other substitutions that could offer spectroscopic signatures of magnetism, while preserving the electronic structure of the zigzag edge.

**Aran Garcia-Lekue** and **Daniel Sánchez-Portal** are at the Donostia International Physics Center, 20018 San Sebastian, Spain. **A.G.-L.** is also at Ikerbasque, Basque Foundation for Science, Bilbao, Spain. **D.S.-P.** is also at the Materials Physics Centre (CSIC-UPV/EHU), San Sebastian. e-mails: wmbgalea@ehu.eus; daniel.sanchez@ehu.eus

1. Wang, H. *et al.* *Nature Rev. Phys.* **3**, 791–802 (2021).
2. Blackwell, R. E. *et al.* *Nature* **600**, 647–652 (2021).
3. Son, Y.-W., Cohen, M. L. & Louie, S. G. *Nature* **444**, 347–349 (2006).
4. Mishra, S. *et al.* *Nature Nanotechnol.* **15**, 22–28 (2020).
5. Li, J. *et al.* *Nature Commun.* **10**, 200 (2019).
6. Lawrence, J. *et al.* *ACS Nano* **14**, 4499–4508 (2020).
7. Mishra, S. *et al.* *Nature* **598**, 287–292 (2021).
8. Cai, J. *et al.* *Nature* **466**, 470–473 (2010).
9. Ruffieux, P. *et al.* *Nature* **531**, 489–492 (2016).

The authors declare no competing interests.

Sociology

# What surveys really say

Frauke Kreuter

Increasing the sample size of a survey is often thought to increase the accuracy of the results. However, an analysis of big surveys on the uptake of COVID-19 vaccines shows that larger sample sizes do not protect against bias. **See p.695**

The accuracy of survey results is often thought to increase with sample size. However, Bradley *et al.*<sup>1</sup> show on page 695 that this is not always the case. Although ‘big’ surveys can, under certain conditions, be useful for tracking changes in a population measure over time and across space, their estimates of population variables can be considerably biased.

Early in the COVID-19 pandemic, many nations lacked essential epidemiological data – even those with well-developed public-health monitoring infrastructures. There was a scarcity of timely information on regional increases in SARS-CoV-2 infections, on adherence to physical-distancing measures and on the social and economic effects of the pandemic. The state-sponsored data collections that existed at the time were often too slow to meet the demands generated by the pandemic.

As a result, some private companies jumped in to offer data; for example, Google, in Mountain View, California, provided anonymized,

**“These findings are a reminder to researchers that statistical precision does not equate to unbiased population estimates.”**

aggregated data on people’s mobility (go.nature.com/3htjccv), and Facebook in Menlo Park, California, presented anonymized and aggregated data about the development of connections between different geographical regions (go.nature.com/3lwknox). The London-based lifestyle company ZOE built the ZOE COVID Study app in collaboration with academic partners (go.nature.com/3i7ypxj). The app surveyed participants who downloaded it, to identify infection hotspots and track the effect of mitigation measures. And when vaccination programmes were rolled out, it was used to record COVID-19 vaccine side effects. In addition, various private-sector surveys – many of which were archived by the US-based Societal Experts

Action Network (go.nature.com/3rcmkwh) – produced data on changes in the public’s response to the pandemic.

The US Census Bureau, in partnership with various federal agencies, and the Delphi group at Carnegie Mellon University, based in Pittsburgh, Pennsylvania, in partnership with Facebook, designed and performed massive surveys to forecast the spread of COVID-19 and measure its effects; questions about vaccination were added in early 2021. With more than 3 million and 25 million responses collected, respectively (as of November 2021; see go.nature.com/3dg0qvy and go.nature.com/3y2r1bk), these are now probably the largest US surveys relating to the pandemic. However, using a subset of responses, Bradley and colleagues demonstrate that the US Census Bureau–federal agencies survey (dubbed the Census Household Pulse survey) and the Delphi–Facebook survey overestimated the vaccination uptake compared with the benchmark data from the US Centers for Disease Control and Prevention (CDC) (Fig. 1).

The authors conclude that having more data does not necessarily lead to better estimates. They discuss how design choices in survey-data collection can lead to error – in this case, the overestimation of vaccination uptake. Their findings are a reminder to researchers that statistical precision does not equate to unbiased population estimates.

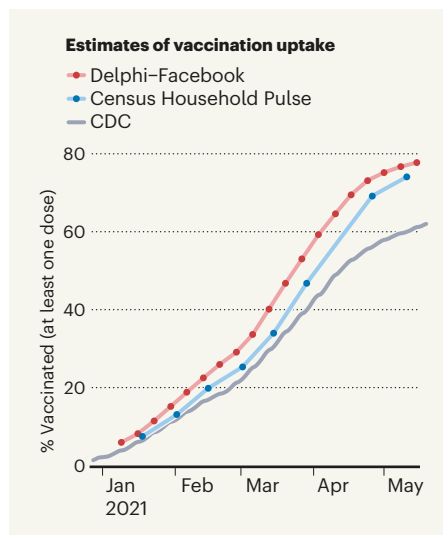
Bradley and co-workers focus on three elements that can contribute to the size of the error – that is, the difference between estimates from big surveys and actual population values. These elements are data quantity (the fraction of a population that is captured in the sample), problem difficulty (how much variation in the outcome of interest there is in the population) and data quality. The quality is very difficult to assess, because there is usually no independently verified ‘ground truth’ or ‘gold standard’ with which to compare survey data. In this case, the CDC’s reports of the numbers of vaccines administered provide benchmark data with which the estimates reported in the surveys could be compared. Under

the strong assumption that these reports are indeed the gold standard and reflect the correct vaccination rates, the survey estimates can be compared with these official numbers (which the CDC frequently updates; state-level estimates updated more recently than those used by Bradley *et al.* can be found at [go.nature.com/3dtrdit](https://go.nature.com/3dtrdit)). Using this approach, Bradley *et al.* evaluated estimates from several surveys and found that they did not match the CDC's reported rates of vaccination uptake.

However, what the metric used by Bradley and colleagues does not enable us to answer – at least, not quantitatively – is the cause of the differences in data quality. To address this issue, the authors used a conceptual framework from survey methodology<sup>2</sup> called the total survey error (TSE) framework<sup>3</sup>, which can help to optimize survey-data quality in three key ways.

First, the TSE framework seeks to ensure that the population of interest and the members included in the 'frame' from which the sample is drawn are aligned. Facebook's active user base is an example of a population that is not aligned with the entire population of the United States. Therefore, if Facebook users have different vaccination habits from those who do not use Facebook, estimates from a survey of Facebook users will be biased. Second, the framework aims to minimize the extent to which those who are sampled and respond differ from the sample members who do not respond. For example, some people who don't trust the government might be less likely to respond to a government survey. Third, the accordance between the survey measure and the construct of interest should be maximized, and the respondents need to answer in the way intended. For example, questions about vaccination are at risk of being answered positively if respondents feel that they need to present themselves in a favourable light.

For certain inferential tasks, surveys with deficiencies can be useful<sup>4</sup>. The usefulness of a data set can be evaluated only in the context of a specific research question. For example, data from samples that are known to be biased have provided useful information for monitoring inflation rates, as exemplified by the Billion Prices Project ([go.nature.com/3i6qock](https://go.nature.com/3i6qock))<sup>5</sup> – which, for years, used prices of online goods and services to estimate alternative indices of inflation. The project was able to do this because, even though not all goods and services were online, online and offline price changes tracked each other. Similarly, the data produced by the US Census Bureau and its partner agencies, and by the Delphi–Facebook partnership, can help to create early-warning systems when administrative data are lacking, as well as help to track cases<sup>6</sup> and evaluate the effectiveness of measures designed to mitigate the spread of SARS-CoV-2 infections, if



**Figure 1 | Big surveys can give biased estimates of population variables.** Bradley *et al.*<sup>1</sup> compared estimates of the uptake of SARS-CoV-2 vaccines among US adults, as reported by large surveys, with numbers of administered vaccine doses reported by the US Centers for Disease Control and Prevention (CDC) on 26 May 2021. Results from a survey carried out by the US Census Bureau in partnership with various federal agencies (Census Household Pulse), and another survey by the Delphi group at Carnegie Mellon University in Pittsburgh, Pennsylvania, in partnership with Facebook (Delphi–Facebook), overestimated vaccine uptake, but were useful in tracking the increase in vaccination over time in the first half of 2021. Bradley and colleagues explain how design choices in these surveys could account for the bias in the surveys' absolute estimates of vaccine uptake.

the errors of these surveys stay constant over time.

Large sample sizes can also reveal relationships between variables – such as reasons for vaccine hesitancy in subgroups of the population, and changes in these reasons over time – unless these relationships for survey respondents differ from those for people who do not respond. Samples collected at a high frequency over time and across relatively small geographical areas, such as some of the samples discussed by Bradley and colleagues, can also be used to evaluate the need for and effectiveness of policy interventions, such as mask-wearing mandates, lockdowns and school-based measures to limit COVID-19 spread<sup>7–9</sup>.

The world is moving towards making decisions on the basis of data – as reflected, for example, in the US Foundations for Evidence-Based Policymaking Act of 2018 and the European Data Strategy ([go.nature.com/3cp1f7o](https://go.nature.com/3cp1f7o)). In response to these changes, we will probably see more data from all kinds of sources, not just surveys. Strong hopes rest on having more available administrative data, such as those from the CDC, that can in some instances replace survey data<sup>10</sup> and, in

others, improve survey estimates<sup>11</sup>.

However, as with survey data, we will need robust frameworks and metrics to assess the quality of the data provided by governments, academic institutions and the private sector, and to guide us in using such data. The work by Bradley and colleagues reminds us that, alongside the studies themselves, research is needed on how best to use data – and on their quality and relevance to the question being asked.

**Frauke Kreuter** is in the Department of Statistics, Ludwig Maximilian University, Munich 80539, Germany, and in the Joint Program in Survey Methodology, the University of Maryland, College Park, USA. e-mail: [fkreuter@umd.edu](mailto:fkreuter@umd.edu)

- Bradley, V. C. *et al.* *Nature* **600**, 695–700 (2021).
- Groves, R. M. *et al.* *Survey Methodology* Vol. 56, 1 Ch. 2 (Wiley, 2011).
- Biemer, P. P. *et al.* (eds) *Total Survey Error in Practice* (Wiley, 2017).
- Kohler, U. *Surv. Methods Insights Field* <https://doi.org/10.13094/SMIF-2019-00014> (2019).
- Cavallo, A. & Rigobon, R. *J. Econ. Perspect.* **30**, 151–178 (2016).
- Reinhart, A. *et al.* Preprint at medRxiv <https://doi.org/10.1101/2021.07.12.21259660> (2021).
- Lessler, J. *et al.* *Science* **372**, 1092–1097 (2021).
- Bilinski, A., Emanuel, E., Salomon, J. A. & Venkataramani, A. *J. Gen. Intern. Med.* **36**, 1825–1828 (2021).
- Adjodah, D. *et al.* *PLoS ONE* **16**, e0252315 (2021).
- Lane, J. *Democratizing our Data* (MIT Press, 2020).
- National Academies of Sciences, Engineering, and Medicine. *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps* (National Academies, 2017).

The author declares competing interests. See [go.nature.com/3d7ewhx](https://go.nature.com/3d7ewhx) for details.

This article was published online on 8 December 2021.