





















Protecting the integrity of survey research

Kathleen Hall Jamieson ^{a,*}, Arthur Lupia ^b, Ashley Amaya ^c, Henry E. Brady ^d, René Bautista ^e, Joshua D. Clinton ^f, Jill A. Dever ^g, David Dutwin ^h, Daniel L. Goroff ⁱ, D. Sunshine Hillygus ^j, Courtney Kennedy ^k, Gary Langer ^l, John S. Lapinski ¹, Michael Link ^m, Tasha Philpot ⁿ, Ken Prewitt ^o, Doug Rivers ^p, Lynn Vavreck ^q, David C. Wilson ^r and Marcia K. McNutt ^s

^aAnnenberg Public Policy Center, University of Pennsylvania, Philadelphia, PA 19104, USA

^bDepartment of Political Science, University of Michigan, 505 South State Street, Ann Arbor, MI 48106, USA

^cPew Research Center, 1615 L Street NW #800, Washington, DC 20036, USA

^dGoldman School of Public Policy, University of California, 2607 Hearst Ave, Berkeley, CA 94720, USA

^eMethodology and Quantitative Social Science Department, NORC at the University of Chicago, Chicago, IL 60637, USA

^fDepartment of Political Science, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203, USA

^gDivision for Statistical and Data Sciences, RTI International, Washington, DC 20005-3967, USA

^hCenter for Panel Survey Sciences, NORC at the University of Chicago, 55 East Monroe Street, Chicago, IL 60618, USA

ⁱAlfred P. Sloan Foundation, 630 Fifth Avenue, New York, NY 10111, USA

^jDepartment of Political Science, Duke University, 140 Science Dr, Durham, NC 27707, USA

^kLanger Research Associates, Carmel, NY 10512, USA

^lDepartment of Political Science, University of Pennsylvania, 133 South 36th Street, Philadelphia, PA 19104, USA

^mIpsos, Canton, GA 30115, USA

ⁿDepartment of Government, University of Texas at Austin, 158 W 21st Street, Austin, TX 78712, USA

^oDepartment of Political Science, Columbia University, New York City, NY 10012, USA

^pDepartment of Political Science, Stanford University, Stanford, CA 94305, USA

^qDepartment of Political Science, University of California Los Angeles, Los Angeles, CA 90095, USA

^rGoldman School of Public Policy, University of California, Berkeley, Berkeley, CA 94720, USA

^sNational Academy of Sciences, 500 5th Street NW, Washington, DC 20001, USA

*To whom correspondence should be addressed: Email: kathleen.jamieson@asc.upenn.edu

Edited By: Michele Gelfand

Abstract

Although polling is not irredeemably broken, changes in technology and society create challenges that, if not addressed well, can threaten the quality of election polls and other important surveys on topics such as the economy. This essay describes some of these challenges and recommends remediations to protect the integrity of all kinds of survey research, including election polls. These 12 recommendations specify ways that survey researchers, and those who use polls and other public-oriented surveys, can increase the accuracy and trustworthiness of their data and analyses. Many of these recommendations align practice with the scientific norms of transparency, clarity, and self-correction. The transparency recommendations focus on improving disclosure of factors that affect the nature and quality of survey data. The clarity recommendations call for more precise use of terms such as “representative sample” and clear description of survey attributes that can affect accuracy. The recommendation about correcting the record urges the creation of a publicly available, professionally curated archive of identified technical problems and their remedies. The paper also calls for development of better benchmarks and for additional research on the effects of panel conditioning. Finally, the authors suggest ways to help people who want to use or learn from survey research understand the strengths and limitations of surveys and distinguish legitimate and problematic uses of these methods.

Keywords: scientific integrity, survey research, weighting of data

Introduction

The type of field study known as a survey “involves the collection of data from a sample of elements drawn from a well-defined population through the use of a questionnaire” (1). Because they provide critical data about people, communities, and nations,

surveys are used to inform policy, clarify stakeholder needs, and improve accountability and customer service in the private and public sectors. The scientific activity termed “survey research” provides the methodological and organizational foundations of this work and is a source of its credibility.

OXFORD
UNIVERSITY PRESS

Competing interest: In addition to his Stanford University affiliation, Doug Rivers is Chief Scientist at YouGov. David Dutwin is senior vice president for NORC at the University of Chicago, a nonpartisan survey and social research organization, and was president of AAPOR in 2018–19. Gary Langer is founder and president of a for-profit company that provides survey research design, management and analysis services to nonprofits, foundations, businesses, and government agencies, as well as a sister company that provides knowledge management software to survey practitioners. Langer also is vice chair of the Roper Center for Public Opinion Research and lead author of its transparency policy, implemented in 2018. Marcia K. McNutt is the President of the National Academy of Sciences. The other authors report no competing interests.

Received: July 26, 2022. **Revised:** January 26, 2023. **Accepted:** February 3, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Scholars at universities and professionals at a wide range of public opinion and survey research organizations share findings and methodological advances in journals such as *Public Opinion Quarterly*, *The Journal of Survey Statistics and Methodology*, and *Survey Methodology*. Professional organizations, such as the American Association for Public Opinion Research (AAPOR), the World Association for Public Opinion (WAPOR), the European Society for Opinion and Marketing Research (ESOMAR), the Insights Association, and the American Statistical Association promulgate best practices designed to improve data collection and data quality, and promote professional standards and ethics.

In recent years, questions have been raised about whether this way of knowing is as reliable as it once was. Some who question its reliability point to trends in refusal to participate in surveys, a phenomenon that increases the difficulty and cost of securing samples that can produce reliable and precise inferences about a population of interest. The advent of alternative, less costly “non-probability” or “opt-in” samples and a range of methodological challenges associated with changes in society and technology (2, 3) raise related concerns. Questions about the reliability of survey research also appear in political contexts. In addition to instances in which some survey researchers have inaccurately forecasted high-profile election outcomes, skepticism in some parts of the population has been fueled by political polarization (4), partisan attacks on ideologically inconvenient survey findings, declining trust in governments and media institutions that fund major surveys (5), and attacks on expertise and experts, including those in academe (6).

In this paper, we offer recommendations for protecting the integrity of survey research in light of these and other challenges. While many surveys are designed to answer questions about corporate reputations and marketing options, we focus on protecting the integrity of studies intended to advance a public interest. A quick scan of the national survey landscape reveals some of the ways in which survey research is used to improve quality of life in the United States. These include the US Census Bureau’s annual American Community Survey (ACS) and Annual Social and Economic Supplement to the Community Population Survey (7), the Bureau of Labor Statistics’ Current Employment Statistics survey (8), the National Science Foundation’s three large “infrastructure surveys” that track Americans’ attitudes about society (the General Social Survey), the economy (the Panel Study of Income Dynamics), and elections (the American National Election Studies), the University of Michigan’s Surveys of Consumers (9), the Center for Disease Control’s Behavioral Risk Factor Surveillance System and National Intimate Partner Violence Survey, the SEAN COVID-19 Survey Archive (10), and work by the Pew Research Center, among others. Collectively, these and other high-quality surveys and data providers inform leaders and the populace on the state of the nation, the substance and meaning of public attitudes and experiences, and public opinion about critical issues facing our society and the globe. So, too, does the survey-based research published in scholarly journals by researchers in academic fields, such as political science, communication, sociology, and public health.

Factors raising questions about the trustworthiness of survey research

Questions about the trustworthiness of survey research can arise when surveys produce contradictory results or ones belied by other data, such as the certified vote count in an election. While possibly a consequence of error, these types of outcomes instead can reflect dissimilar operationalizations or methodological

approaches. When researchers choose noncomparable ways to measure attitudes and behaviors, or employ distinctive sampling frames, modes, field periods, or question wordings, their survey data can produce different conclusions.

Survey researchers also can arrive at dissimilar conclusions because some have produced inaccurate estimates. Paths to inaccuracy include a misunderstanding of who is, and is not, participating in a survey. Decades ago, adults in the US could be reached at a particular landline phone number, and most received very few requests to participate in surveys of any kind. When they were invited to take a survey, many people agreed to do so. These assumptions no longer hold true. Cell phones have replaced landlines, and many people avoid calls from unfamiliar numbers. At the same time, even when reached by phone, some parts of the population are now much less likely to accept invitations to be interviewed than they once were. Where a survey researcher was, in earlier times, expected to elicit answers from more than 60% of the target sample, response rates now rarely reach 10%. Consequently, it is now more expensive, time-consuming, and difficult for survey researchers to secure the types of representative samples for which the most reputable surveys have long been known. Declining response rates (11) and changing patterns of “nonresponse” (12, 13) are among the factors that affect the quality of population estimates. It has been reported, for example, that Democrats are more likely than Republicans and Independents to agree to be interviewed (14).

Because elections produce certified outcomes, the accuracy of pre-election polls is particularly susceptible to year-to-year comparisons and report card-like assessments of performance. So, for example, an AAPOR 2020 election post-mortem observed that “national presidential polls had their worst performance in 40 years and state-level presidential polls had their worst overall performance in 20 years” (15). And a post-2022 midterm election piece in *The Conversation* observed that “As compiled by the widely followed RealClearPolitics site, polls collectively missed the margins of victory by more than 4 percentage points in key 2022 Senate races in Arizona, Colorado, Florida, New Hampshire, Pennsylvania and Washington.... In gubernatorial races, deviations from polling averages of 4 percentage points or more figured in the outcomes in Arizona, Colorado, Florida, Michigan, Pennsylvania and Wisconsin” (16).

Forecasting election outcomes is a particularly complicated endeavor. Survey researchers are often uncertain about who will turn out to vote from election to election. Because many factors can affect individual decisions on whether to participate in a particular election, survey producers use models to estimate which survey respondents are more and which are less likely to vote in the upcoming contest. The models mix current information and historical trends to generate predictions about who will vote and who will stay home. When researchers use different models, or when turnout varies in unexpected ways, pre-election polls may provide inaccurate estimates about an election’s outcome.

Having noted some of the challenges facing survey researchers, our question becomes:

How can the survey research and associated communities better safeguard integrity and increase the utility of surveys on which scholars, leaders, and the public rely to understand the attitudes and behaviors of important populations?

The protecting the integrity of survey research retreat

To address this question, on November 18 and 19, 2021, Marcia McNutt, president of the National Academy of Sciences, convened

a virtual retreat to explore ways to protect the integrity of survey research, increase understanding of the limitations and strengths of individual surveys, incentivize disclosure of information needed to evaluate findings from surveys, and help the public recognize distinguishing features of credible surveys. The retreat was cohosted by the Annenberg Foundation Trust at Sunnylands and the Annenberg Public Policy Center (APPC) of the University of Pennsylvania. The proceedings were coordinated by Arthur Lupia of the University of Michigan and Kathleen Hall Jamieson, APPC director and Sunnylands Trust program director. Included among the 20 participants were the crafters of this document, a list that includes current and past editors of major academic journals, past presidents of the American Political Science Association and AAPOR, and a past director of the US Census Bureau, as well as scholars who have led some of the nation's largest university-based election surveys and individuals responsible for the creation and maintenance of large governmental and 501(c)(3) survey datasets.

The convening's main outcome is an understanding that safeguarding the integrity of survey research, including political polls, is possible. A path to that end includes renewed commitments to scientific norms of transparency, precise specification of key methodological decisions, dedication to disclosure and self-correction when errors are identified, and improved reporting practices by researchers and the media. In service of these goals, we offer 12 actions that key stakeholders can take now, and in the near future, to improve the integrity, utility, and public understanding of surveys.

Changes in technology and society create challenges that, if not addressed well, can threaten the quality of important surveys on topics such as public health, the economy, and elections. In what follows, we describe some of them and recommend remedies designed to protect the integrity of survey research. These recommendations are the product of presentations and conversations at the retreat and email discussions in the months that followed. They reflect points of agreement among a diverse group of stakeholders. Collectively, the recommendations are designed to increase the research community's ability to independently assess survey-based research claims and to share those assessments with a broader audience. If followed, these recommendations will add to the existing menu of good practices and strengthen the ability of researchers to draw properly qualified, reliable inferences from survey research.

Twelve recommendations

Since survey research is a form of scientific inquiry, many of our recommendations focus on ways to better align current practice

with scientific norms. Workshop organizers organized their discussions and grounded their recommendations in three of those interrelated norms: *transparency*, *clarity*, and *correcting the record*.

- **Transparency** about methods and practices helps generate constructive critiques and critical insights and fuels science's norm of self-correction. Transparency makes it possible for other researchers to reproduce past work and determine whether it is replicable (17). Transparency also makes it possible to compare methods across surveys that have produced dissimilar results.
- **Clarity** ensures that scholarly methods and objects of inquiry are expressed in apt, carefully defined terms. When this norm is honored, assumptions about data transformations are presented intelligibly, and findings from these inquiries and analyses are expressed in ways that align with the underlying data.
- **Correcting the record** is a multi-stage process that involves flagging problems, assessing the viability of proposed solutions, and determining how well the ones that are implemented are working. Science's culture of transparency, clarity, and critique increases the likelihood that this multi-stage process will occur and succeed.

With a goal of translating these norms into tangible outcomes, we offer 12 recommendations to an audience that includes public opinion scholars and practitioners, survey vendors, leaders of AAPOR and related professional associations, journal editors, reporters and publishers, and others who use or report on survey findings. Each recommendation includes a course of action and people or organizations we consider well-positioned to demonstrate its feasibility and value.

Tables 1–5 summarize our recommendations.

The recommendations vary in their resource requirements. Some are relatively easy to implement, while others entail costs. In each case, we have concluded that the benefit of implementing the proposed action is worth the cost.

Recommendations and rationales

1. **TRANSPARENCY: A norm of transparency requires access to datasets and disclosure of how respondents were recruited,**

Table 1. Transparency recommendations.

Transparency recommendations	Implementer(s)	Location
1. Improve disclosure of sampling design, modeling, and weighting assumptions for all surveys – probability and non-probability alike.	Vendors, authors, and researchers.	Disclosed in all documents associated with data distribution and in all publications that use the data.
2. Disclose question wording and order.		
3. Improve disclosure of respondent recruitment and question-related panel conditioning factors.		
4. Disclose known or expected consequences of attrition on panel surveys.		
5. Create client expectation checklists to include newly recommended forms of disclosure.	Professional associations that include large numbers of survey vendors, researchers, or clients.	A publicly available template for clients to use when contracting with an organization to conduct a survey.
6. Engage entities such as AAPOR to incorporate new disclosure recommendations into its Code of Ethics	AAPOR & other public-facing survey research organizations.	Reported in AAPOR publications and survey-focused venues.

Table 2. Clarity recommendations.

Clarity recommendations	Implementer(s)	Location
7. When survey data are weighted, the phrase “representative sample” should not be used without explicit acknowledgment of the underlying assumptions, including weighting and modeling assumptions. Survey vendors should not release data without this information, and publishers of content that use survey data should publicly commit to cite or use data only from sources that provide this information.	Vendors, authors, editors, and publishers.	Report in all documents associated with data distribution and in all publications that use the data.
8. Incentivize clarity by asking authors to populate, and dissemination outlets to publish, a template that, like the Roper Center’s Transparency and Acquisition Policy, clearly describes survey attributes (e.g. information about the sample, wording, and coding decisions) that can influence accuracy.		

Table 3. Correcting the record recommendation.

Correcting the record recommendation	Implementer(s)	Location
9. Create an online resource center to archive and make accessible information about technical problems, sources of data corruption, and solutions that survey researchers have uncovered when trying to conduct surveys rigorously and responsibly. The National Academy of Engineering’s Online Ethics Center provides a template.	AAPOR, the Roper Center for Public Opinion Research, ICPSR, or a similar organization--ideally with funding support from the National Science Foundation or equivalent funder.	Publicly available, professionally curated archive.

Table 4. Increasing recognition of the value and limitations of survey research recommendations.

Increasing recognition of the value and limitations of survey research recommendations	Implementer(s)	Location
10. Professional organizations and universities should develop and disseminate a guide to survey research that can be used in high school courses.	AAPOR, other survey-focused organizations, and universities.	Various forms of content that expand public understanding of survey data Curricular offerings.
11. Draw greater attention to organizations that join AAPOR’s Transparency Initiative.	Professional organizations, universities, and media outlets.	Reported in AAPOR publications and survey-focused venues.

Table 5. Improving the quality of survey data recommendations.

Improving the quality of survey data recommendations	Implementer(s)	Location
12. A. Facilitate efforts to create better benchmarks and other data that can improve survey quality. B. Federal funders of survey research, private philanthropists, and companies that recognize the public importance of maintaining the integrity of survey research should prioritize support of widely usable research that identifies and shows how to mitigate negative consequences of panel conditioning.	Government science agencies, science philanthropy, and private sector.	12A. Publicly available on professionally curated websites. 12B. Competitively selected scholars produce widely shared research addressing the question.

sources of possible respondent conditioning, the existence, effects, and researcher responses to attrition bias, and weighting and modeling assumptions. This information should be available at every stage of survey research and publication processes.

A commitment to transparency means that those who conduct, analyze, and report on surveys should disclose key properties of the data on which they report as well as limitations in the data and should indicate the ways in which the researcher addressed those limitations. In keeping with this norm, the retreat’s first set of recommendations builds upon the AAPOR Code of Professional Ethics and Practices’ principle that “good

professional practice imposes the obligation upon all public opinion and survey researchers to disclose sufficient information about how the research was conducted to allow for independent review and verification of research claims, regardless of the methodology used in the research.” Among the practices in service of that goal are those listed in the current version of AAPOR’s Code (18). A key component of the Code is “access to data”, which the Code operationalizes by saying that:

Reflecting the fundamental goals of transparency and replicability, AAPOR members share the expectation that access to datasets and related documentation will be provided to allow for independent review and verification of research claims upon request.

This principle should be honored to the extent possible, with an understanding that risks to privacy and other ethical concerns will limit the types of data that can be shared.

Beyond data access, the norms of transparency and reproducibility require disclosure. Section III of the Code, on which our next set of recommendations builds, calls for any report or article that uses survey research to immediately disclose:

1. The data collection strategy.
2. Who sponsored the research and who conducted it.
3. Tools and instruments that can influence responses.
4. Which population the survey is designed to study.
5. Which methods were used to generate and recruit the sample.
6. Method(s) and mode(s) of data collection.
7. The dates on which data were collected.
8. Sample sizes and expected precision of results.
9. How the data were weighted and, otherwise, reasons for unweighted estimates.
10. Steps taken to assess and assure data quality.
11. A general statement acknowledging limitations of the data.¹

The AAPOR Code also specifies additional items for disclosure after results are reported. To its “Procedures for managing participation in surveys whose participants are interviewed multiple times or at different times”, we would add: **disclosure of modeling and weighting assumptions at all stages of the data generation, analysis, and dissemination process; disclosure of question wording and order; improved disclosure of sources of respondent conditioning; disclosure of attrition; enhanced client expectation checklists to include newly recommended forms of disclosure.**

Improve disclosure of modeling and weighting assumptions

Since there is not a one-to-one correspondence between the salient characteristics of the sample and the population from which it is drawn, modeling and weighting are regular features of survey research. Among other uses, they are designed to compensate for different rates of survey nonresponse and nonparticipation.

Recommendation 1A: *To facilitate evaluations of assumptions, such as these, and to facilitate more accurate interpretations of the resulting data and analyses, survey vendors should disclose their modeling and weighting assumptions to all users of survey data in ways that are consistent with the FAIR (findable, accessible, interoperable, and reusable) principles for open data.*

Recommendation 1B: *All publications that include survey research findings should require that modeling and weighting assumptions be disclosed as part of an article’s methods section or in supplementary material to which the publication offers direct links. When no weights have been applied, that fact should be disclosed as well.*

Today, the extent of disclosure by survey organizations varies widely. We call on survey vendors who are not fully disclosing their modeling and weighting assumptions to do so. This action will empower researchers to analyze, and reporters to more accurately interpret, the corresponding data. Such transparency will make it possible for reporters and researchers to explore how and why different data or analyses produce different results.

¹ For more details, read the full version of the AAPOR Code (18).

Margins of error and credibility intervals reported in survey-based articles tend to assume that the survey estimates are unbiased and only subject to errors due to incomplete sampling. In other words, this way of reporting survey results depends on the often-unrealistic assumption that survey samples contain no asymmetries. But asymmetries often exist in which parts of a population are covered in a sample or in factors that can cause differential rates of participation or otherwise influence a sample’s ability to represent a larger population. When the assumption that survey estimates are unbiased does not hold, particularly in the case of surveys that depend on participating “opt-in” rather than random selection, reported margins of error understate the likely magnitude of errors in survey estimates.

Recommendation 1C: *Since some practitioners (particularly those outside of academia) may not know how to assess or disclose all these phenomena, professional associations and groups that oversee the integrity of the federal statistical agencies that use survey research ought to provide templates and education that can improve the extent and utility of disclosure of modeling and weighting assumptions.*

These templates could take the form of a one-page document that lists questions to ask or a brief report that describes common types of decisions made when developing models or weighting schemes.

Improve disclosure of question wording and order

Among the ways in which the answers of survey respondents can be biased is sensitization or conditioning because of exposure to earlier questions in a survey. Since these sorts of exposures can affect subsequent responses, they should be disclosed by vendors and reported in all forms in which the results are disseminated.

The AAPOR Code’s description of tools and instruments that should be immediately disclosed requires survey vendors to share “questionnaires with survey questions and response options, show cards, vignettes, or scripts used to guide discussions or interviews. The exact wording and presentation of any measurement tool from which results are reported as well as any preceding contextual information that might reasonably be expected to influence responses to the reported results and instructions to respondents or interviewers should be included.” However, not all vendors comply with this expectation, particularly when it comes to disclosing question order. Because journals that publish scholarly work based on surveys and public accounts of survey findings could incentivize this process by requiring disclosure of this information as a condition of publication and by linking it to publications, Recommendation 2 urges them to do so.

Recommendation 2: *All publications that include survey research findings should require question wording and order disclosures as part of an article’s methods section or in supplementary material to which the publication offers direct, permanent links.*

Improve disclosure of respondent or panel conditioning factors

Panel studies are surveys in which the same respondents are interviewed multiple times and that can be conducted on samples of individuals who are randomly selected to participate or samples of those who opt-in. Respondent or panel conditioning occurs when respondents’ responses to prior questions or experiences in an earlier survey affect their later responses (see, e.g. 19). Panel conditioning does not always occur, but greater disclosure about the types of questions that a survey previously asked a panel’s

respondents can help researchers, reporters, and the public better understand whether this type of effect could be influencing responses.

To alert researchers and interested members of the public to such conditioning effects, survey vendors should be expected to disclose information about any past participation by a respondent in their surveys that might affect that individual's responses to the survey being analyzed. To understand the full context of potential conditioning effects within a survey or across surveys, this information should include both how the survey vendor recruited panelists and relevant data about factors that distinguish those who decided to participate from those who opted not to do so. In sum, vendors should disclose the nature and extent of respondents' exposure to previous material that may have prejudiced the survey data, and scholarly journals should require that those publishing survey data include such disclosures in the methods sections of their articles.

Recommendation 3: *Survey vendors should disclose both panel recruitment and retention practices and questions on a survey that have the potential to influence responses given to subsequent questions in that survey. They should also disclose that a particular respondent has been asked questions in a previous survey that can have the same effect. To the extent possible, they should explain the types of bias that exposure to previous questions may have introduced. When information about prior exposure does not exist, that fact should be explicitly disclosed as well.*

Because we recognize that there is much that scholars do not know about conditioning effects, Recommendation 12B, below, calls for funders to prioritize additional research on those effects within surveys. Randomized experiments could go a long way toward clarifying the effects of panel and other kinds of conditioning in different survey contexts.

Disclose attrition bias

Panel surveys and longitudinal ones (with subjects interviewed over a defined period of time) rarely retain all of their original respondents in subsequent waves. When the effects of attrition bias are not disclosed in such studies, the resulting data can be misinterpreted.

Recommendation 4: *For panel surveys, vendors and researchers should, in an Annual Non-Response Analysis and Attrition Report, disclose attrition rates, report any estimated biases that result from the change in a panel's composition, and explain what they did to take those changes into account.*

Beyond augmenting AAPOR's Code of Ethics in the ways described above, our next recommendations describe actions that key stakeholders can take to incentivize adherence to best practices.

Client expectation checklists should include new recommended forms of disclosure

Recommendation 5: *Professional associations and others that have a stake in the integrity of survey research should draw greater attention to the list of organizations that have committed to the disclosures and to each of the items on existing disclosure checklists. Survey vendors should act on commitments to follow the disclosure recommendations in these checklists as well as any new disclosures that come to be required, such as those that we have recommended.*

By developing and distributing a checklist indicating the information clients should require of vendors, AAPOR, through its Code of Ethics, and the Roper Center for Public Opinion Research, through its acquisitions policy (20), have sought to

improve public, media, and client understanding of what makes a survey rigorous and reliable. In addition, AAPOR's Transparency Initiative (TI) gives organizations the opportunity to publicly commit to disclosing "its basic research methods and make them available for public inspection" (21). The list of organizations that have agreed to follow these practices can be found on the AAPOR TI web page.²

Incorporate these new disclosure recommendations in the AAPOR Code

Recommendation 6: *Engage AAPOR to consider ways to augment its Code of Professional Ethics and Practices to include disclosure of the sources from which samples are drawn, attrition rates in panels and longitudinal surveys, and the extent to which respondents have been exposed to related surveys or survey questions in the recent past, if known.*

Some vendors already provide information of these types. The purpose of Recommendation 6 is to aid those who use survey data or published results by raising the visibility as well as the frequency of these disclosures. We single out AAPOR for its pioneering work and focal position within the survey research field in the hope that organizations with a comparable mission will take equivalent actions.

2. **CLARITY: A norm of clarity dictates that clear, accurate language be used to characterize the nature of the survey process and its outcomes. A commitment to clarity involves being forthright about the types of precision that surveys can and cannot produce.**

To advance the norm of clarity, our next recommendations expand upon best practices in the AAPOR Code of Professional Ethics and Practices. The relevant part of the code states:

1. We will not knowingly make interpretations of research results that are inconsistent with the data available, nor will we tacitly permit such interpretations. We will ensure that any findings we report, either privately or for public release, are a balanced and accurate portrayal of research results.
2. We will not knowingly imply that interpretations are accorded greater confidence than the data warrant. When we generalize from samples to make statements about populations, we will only make claims of precision and applicability to broader populations that are warranted by the sampling frames and other methods employed (18).

The US Census Bureau's decision not to release the 2020 estimates from one of the nation's premier governmental surveys demonstrates a commitment to ensuring sample quality consistent with such best practices. After surveying a sample of 290,000 people monthly, the Bureau's ACS then "combines the monthly responses into a set of 1-year estimates for the nation, states and communities with populations of 65,000 or more" (22). The ACS is widely used by researchers, governments, and various private sector organizations. However, by disrupting the lives of various subgroups of the US population in different ways in 2020, the COVID-19 pandemic created new "nonresponse bias" challenges that made it more difficult to produce representative survey samples. Although the Bureau sought many ways to adapt to unprecedented circumstances, its researchers concluded that the 2020 ACS data failed to meet the Statistical

² Read the full list of organizations here: (21).

Data Quality Standards established “to ensure the utility, objectivity and integrity of the statistical information” (22). Rather than publish data in a compromised and potentially misleading form, the Bureau announced that it would not release its 1-year estimates from the 2020 survey.

At the same time, the use of clear language and definitions increases the likelihood that researchers speak a shared language when addressing each other and the public.

To increase the clarity with which survey data are reported, we offer the following recommendations:

Recommendation 7: *When survey data are weighted, the phrase “representative sample” should not be used without explicit acknowledgment of the underlying assumptions, including weighting and modeling assumptions. Survey vendors should not release data without including this information, and publishers of content who use survey data should publicly commit to cite or use data only from sources that provide such information.*

There remains an active debate in the survey community about the threshold test a sample must satisfy to be called “representative.” Our hope is that Recommendation 7 will focus that debate on the articulation of explicit standards that can help analysts and the public draw more accurate inferences about the population a survey sample is more, and less, likely to represent.

Of particular importance in these discussions is clarifying the most effective use of probability and non-probability samples. These diverse forms of data collection have distinct advantages and disadvantages. Probability samples minimize the risk of systematic bias. Non-probability samples are easier and less costly to generate. On some matters, there is a consensus on when one method of gathering a sample is more effective at achieving these types of goals. In other cases, there is less consensus on what these different types of surveys can and cannot do. Helping a broader set of stakeholders understand trade-offs associated with these methods could produce significant public benefits.

For example, non-probability surveys are quite efficient for tracking changes in public sentiment (e.g. presidential approval) over time, provided that the estimates need not be extremely precise. Non-probability surveys have also proven useful for estimating treatment effects across randomly assigned groups (e.g. 23, 24). For other research purposes, however, non-probability surveys may not be fit for use. Studies have shown that the positivity bias associated with bogus respondents can lead non-probability surveys to systematically overestimate rare outcomes, such as ingesting bleach to protect against COVID-19 (25), belief in conspiracies like PizzaGate (26), support of political violence (27), or favorable views of Vladimir Putin (28). In terms of scale, Geraci (29) estimates that researchers should anticipate removing 35–50% of non-probability completes due to poor data quality. More broadly, non-probability surveys are not fit for use in federal surveys that are expected to yield highly precise estimates for the country as a whole, but also for harder to reach subgroups.

Our final recommendation in this section pertains to reporting standards. Consistent with our earlier discussions about weighting and modeling, we recommend that publishers require that researchers who use survey data report their decisions about how observations are weighted relative to one another, how this weighting affects margin of error estimates, and which variables are used in attempts to determine whether a population of respondents is sufficiently aligned with a population of interest.

Recommendation 8: *Publishers and editors of scholarly journals should incentivize clarity by adopting reporting standards that better reflect the realities of modern survey research. In particular, they should ask authors*

to populate, and make available for readers to view, a template that, like the Roper Center’s Transparency and Acquisition Policy, clearly describes survey attributes that can influence accuracy.

Completing a well-designed template, particularly a standardized one in hand when work on a survey begins, does not have to be burdensome and can provide important information in an effective way. This simple instrument, in turn, may increase researchers’ and other stakeholders’ ability to accurately interpret survey data and published results.

3. CORRECTING THE RECORD. A norm of self-correction requires that when errors are identified, they be disclosed in forms accessible to other researchers and that protections be put in place to minimize their recurrence in other surveys or subsequent analyses.

Self-correction is a key scientific norm. When scientists are uncertain about the correspondence between an observation and a research claim, the expectation is that they will report that uncertainty. When a scientist discovers an error, a parallel expectation arises.

Because surveys are complex, involve coding human response to language, and often are administered in challenging environments, unanticipated forms of error occur. When an error is identified in published work, existing practice involves reporting it to the journal in which the work appeared and offering a correction. So, for example, Kathleen Hall Jamieson, Marcia McNutt, Veronique Kiermer, and Richard Sever appended a correction to “Signaling the trustworthiness of science” (30), which read: “A coding error was uncovered in the survey vendor’s computer-assisted telephone interviewing (CATI) programming of the 2019 Annenberg Science Knowledge (ASK) survey used in our study. To minimize response order bias, the scale items ranging from 1 to 5 were programmed to reverse from 5 to 1 for a random half of the sample. Instead of recoding the responses to 1, 2, 3, 4, and 5, the programmer recoded only 1 and 5. As a result, the data reported in the article underrepresented the percentage saying that the reported statements mattered.” Although those who search for the article will find a prominent link to the correction, there is currently no ready way for authors to alert others to the need to check to make sure that a comparable problem has not occurred in their vendor’s programming.

Recommendation 9: *We recommend that an online resource center, modeled on the National Science Foundation-supported Online Ethics Center for Engineering and Science (established by the National Academy of Engineering and now run by the University of Virginia (31)), be established to archive and make accessible information about technical problems, sources of data corruption, and solutions that survey researchers have uncovered when trying to conduct surveys rigorously and responsibly.*

Increasing recognition of the distinguishing characteristics of this way of knowing

Increase public understanding of the nature, utility, strength, and limitations of survey research

In decades past, the survey professionals and associated communities from whom we invite specific forms of action have implemented important steps to protect public understanding of survey research. By standardizing the expectation that survey researchers should report their surveys’ margins of error, sample sizes, dates of fielding, and question wording, for example,

AAPOR and other groups have enhanced public access to useful information about how to interpret surveys. The recommendations that we offer next build upon such efforts.

Many survey-based reports give an estimate of the margin of error but fail to mention other potential sources of bias and error, such as those we noted earlier. Here learning materials for high school students and a relatively brief “users guide” for use by laypeople/journalists without statistical or survey research training would be useful. Several excellent textbooks (see, for example, 32), provide such information for those in college classrooms. AAPOR offers a similar resource entitled “5 Tips for Writing About Polls” (<https://medium.com/pew-research-center-decoded/5-tips-for-writing-about-polls-9cb0596ff28>). Such documents could serve as the foundation for materials for use in media literacy courses and in civics education both in and outside the classroom.

Recommendation 10: Professional organizations and universities should develop and disseminate a guide to survey research that can be used in high school courses.

Increase the visibility of organizations that join AAPOR’s Transparency Initiative

The AAPOR’s TI web page identifies the organizations that have committed to publicly disclosing their basic research methods. The names of the subscribers to the TI are posted on the AAPOR website.³ Were academic journals and media outlets that report survey results to note whether the vendor in question has agreed to honor the transparency standards and, were reviewers for scholarly journals to take that evidence into account in evaluating manuscripts for publication, this form of publisher badging could incentivize other firms to adopt them as well. Drawing attention to publishers and media organizations that value the kinds of disclosures specified in the TI would give them a “badge of honor” to use to gain competitive advantages while also distinguishing the features of high-integrity survey work which they disseminate or on which they report.

Recommendation 11: Journals and media outlets that use or report on surveys not only should note whether the data on which they are relying comes from vendors who have joined the AAPOR Transparency Initiative but also should include links to the data and relevant modeling, weighting, attrition, and related information. AAPOR and other organizations should publicly recognize publishers and media outlets that agree to do so.

Improving the quality of benchmarks and related resources

Using weighting to correct imbalances between samples and the underlying population requires reliable benchmarks that document the characteristics of the population. The ACS conducted by the US Census Bureau has traditionally served as such a reliable national population benchmark for core demographics, such as sex, age, race and ethnicity, educational attainment, and geographic region. Although these demographic variables are correlated with many attitudes and behaviors of interest to survey researchers, they are inadequate for some important other ones. As a recent National Science Foundation-funded Duke University conference put it, “A critical resource is a large-scale national sample survey that obtains benchmark estimates of non-demographic characteristics on key dimensions, such as religiosity and social trust, making it possible to assess—and potentially

adjust—for the representativeness of other, less high-end surveys” (33).

Recommendation 12A: Federal funders of survey research, philanthropies, and companies that recognize the importance of safeguarding the integrity of survey research should prioritize support for new benchmarking resources that improve the quality of the surveys that collect data on matters of significance to the nation.

Building a benchmark that provides needed information while protecting privacy requires skill, an understanding of diverse research needs, and significant planning. Indeed, 12A is likely our highest cost recommendation. At the same time, the potential benefits are substantial. A benchmark of this kind would allow survey vendors and analysts to conduct a much wider and deeper array of analyses on critical questions about the economy, elections, and society. A single new benchmark could improve the quality of thousands of important data collections. Although costly, creating such a benchmark is both technologically feasible and of great public benefit.

Our final recommendation focuses on panel conditioning. Earlier, we noted the challenges associated with it and recommended greater disclosure of it. Here, to improve the integrity of future surveys, we recommend additional research on this topic.

Recommendation 12B: Federal funders of survey research, private philanthropists, and companies that recognize the public importance of maintaining the integrity of survey research should prioritize support of widely usable research that identifies, and shows how to mitigate, negative consequences of panel conditioning.

In this case, we call for action by federal funders and private philanthropists because we recognize that research on this topic is what economists call a “public good”. Since everyone can benefit without paying the cost of such public goods, governments and philanthropies are often the social actors called upon to fund needed infrastructure such as streetlights. For the many social entities that benefit from survey research that tracks the attitudes and reported behaviors of people over time, a better understanding of panel conditioning is both critical infrastructure and a public good.

Overcoming barriers to adoption

One might argue that by increasing vendor costs, industry adoption of our disclosure recommendations about weighting assumptions and panel sensitization would drive vendors whose work cannot sustain the resulting scrutiny out of business. If these forms of disclosure help protect the integrity of the research process, as we believe they do, that outcome is a benefit not a downside of adopting them. We believe that these disclosures are likely to reveal, and hopefully reduce, methods of panel assembly that are difficult to defend and, at the same time, will help researchers better interpret panel data.

However, since the increased costs will be passed on by surviving vendors, it is fair to say that some studies may prove cost-prohibitive and not be undertaken. Others will be based on less data than otherwise would have been collected. We believe that improvements in the quality of published research and in the reliability of inferences grounded in the survey data are worth the trade-off and costs. But the market will ultimately determine whether the increased quality of the data, analysis, and reliability of inferences is worth the cost.

Because costs associated with recommendations reduce the likelihood of adoption, many of our recommendations incentivize it by making adoption a signal of greater trustworthiness. The large and

³ See the AAPOR Transparency Initiative web page here: (21).

growing number of journals whose editors or publishers have asked to be listed as subscribers to the International Committee of Medical Journal Editors' (ICMJE's) Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals shows this process at work (<https://www.icmje.org/journals-following-the-icmje-recommendations/>). Among other topics, the ICMJE recommendations focus on Defining the Role of Authors and Contributors, Disclosure of Financial and Non-Financial Relationships and Activities, and Conflicts of Interest, and Responsibilities in Submission and Peer-Review. If the publishers of high-impact journals presuppose disclosures as a condition of publication, and media outlets that report on survey research do the same, researchers will require them from vendors. Because journals are judged in part by their reputation, when those known as high quality adopt a practice, others follow suit. The same logic applies to vendors. If signing on to the AAPOR Transparency Initiative is a signal of commitment to protect the data gathering and reporting process, then vendors who do so have a competitive advantage.

Conclusion

Surveys offer a unique and powerful form of evidence. Safeguarding this important way of collecting data and ensuring that it adheres to scientific norms of transparency, clarity, and correcting the record should be a priority of the scholarly and professional communities and audiences that rely on survey findings. Factors that increase the likelihood that the aspirations embodied in the recommendations will become accepted practice include the extent to which they complement best practices and materials already championed by respected entities in the survey research community, have already been implemented by some gold standard vendors and are leveraged by incentives that the scientific community has successfully employed in the past. Broadly, our set of 12 recommendations calls for a culture change in the research community in which fuller and more open disclosure of survey practices and limitations becomes the norm.

Funding

No Funders. Expenses for the retreat were underwritten by the Annenberg Foundation Trust at Sunnylands and the Annenberg Public Policy Center of the University of Pennsylvania from endowment funds provided to each by the Annenberg Foundation.

Data availability

There are no data underlying this work.

References

- 1 Visser PS, Krosnick JA, Lavrakas PJ. 2000. Survey research. In: Reis HT, Judd CM, editors. *Handbook of research methods in social and personality psychology*. New York (NY): Cambridge University Press. p. 223–252.
- 2 Santos R. 2014. Presidential address: borne of a renaissance—a metamorphosis for our future. *Public Opin Quart*. 78(3):769–777.
- 3 Link M. 2015. Presidential address: AAPOR2025 and the opportunities in the decade before US. *Public Opin Quart*. 79(3): 828–836.
- 4 Pew Research Center. 2014. *Political polarization in the American public*. Pew Research Center. <https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/>.
- 5 Rainie L, Perrin A. 2019, July 22. Key findings about Americans' declining trust in government and each other. Pew Research Center. <https://www.pewresearch.org/fact-tank/2019/07/22/key-findings-about-americans-declining-trust-in-government-and-each-other/>.
- 6 Nichols T. 2017. *The death of expertise: the campaign against established knowledge and why it matters*. New York (NY): Oxford University Press.
- 7 US Census Bureau. 2022. About the American community survey. Census.Gov. <https://www.census.gov/programs-surveys/acs/about.html>.
- 8 US Bureau of Labor Statistics. 2022, February 4. Labor force statistics from the Current Population Survey. BLS.Gov. https://www.bls.gov/web/empsit/ces_cps_trends.htm#intro.
- 9 University of Michigan. 2022. Surveys of consumers - University of Michigan home page. Ann Arbor (MI). <https://data.sca.isr.umich.edu/>.
- 10 Societal Experts Action Network. 2022. Welcome to the Societal Experts Action Network (SEAN) COVID-19 Survey Archive. SEAN COVID-19 Survey Archive. <https://covid-19.parc.us.com/client/index.html#/>.
- 11 Czajka JL, Beyler A. 2016. Declining response rates in federal surveys: trends and implications (Background Paper Volume 1). Mathematica Policy Research. <https://aspe.hhs.gov/sites/default/files/private/pdf/255531/Decliningresponserates.pdf>.
- 12 Bernhardt R, Munro D, Wolcott E. 2021. How does the dramatic rise of CPS non-response impact labor market indicators? (Working Paper No. 781). GLO Discussion Paper. <https://www.econstor.eu/handle/10419/229653>.
- 13 Williams D, Brick JM. 2018. Trends in US face-to-face household survey nonresponse and level of effort. *J Survey Stat Methodol*. 6(2): 186–211.
- 14 Clinton J, Lapinski JS, Trussler MJ. 2022. Reluctant Republicans, eager Democrats? Partisan nonresponse and the accuracy of 2020 presidential pre-election telephone polls. *Public Opin Quart*. 86(2):247–269.
- 15 Clinton J, et al. 2021. AAPOR Task Force on 2020 Pre-Election Polling Report FNL. AAPOR. https://www.researchgate.net/publication/353343195_AAPOR_Task_Force_on_2020_Pre-Election_Polling_Report_FNL.
- 16 Campbell WJ. 2022, November 17. Some midterm polls were on-target—but finding which pollsters and poll aggregators to believe can be challenging. *The Conversation*. <https://theconversation.com/amp/some-midterm-polls-were-on-target-but-finding-which-pollsters-and-poll-aggregators-to-believe-can-be-challenging-194700>.
- 17 National Academies of Science, Engineering, and Medicine. 2019, April 7. New report examines reproducibility and replicability in science, recommends ways to improve transparency and rigor in research. Washington (DC). <https://www.nationalacademies.org/news/2019/05/new-report-examines-reproducibility-and-replicability-in-science-recommends-ways-to-improve-transparency-and-rigor-in-research>.
- 18 American Association for Public Opinion Research. 2021, April. AAPOR Code of Professional Ethics and Practices. <https://www.aapor.org/Standards-Ethics/AAPOR-Code-of-Ethics.aspx>.
- 19 Halpern-Manners A, Warren JR. 2012. Panel conditioning in longitudinal studies: evidence from labor force items in the current population survey. *Demography*. 49(4):1499–1519.
- 20 Roper Center for Public Opinion Research. 2018, June 22. Roper Center transparency and acquisition policy. Ithaca (NY). <https://ropercenter.cornell.edu/roper-center-transparency-and-acquisitions-policy>.

- 21 American Association for Public Opinion Research. 2015, March 20. What is the TI? <https://aapor.org/standards-and-ethics/transparency-initiative/>.
- 22 Census.Gov. 2021, July 29. Census Bureau Announces Changes for 2020 American Community Survey 1-Year Estimates. <https://www.census.gov/newsroom/press-releases/2021/changes-2020-acs-1-year.html>.
- 23 Coppock A, Leeper TJ, Mullinix KJ. 2018. Generalizability of heterogeneous treatment effects estimates across samples. *Proc Natl Acad Sci U S A*. 115(49):12441–12446.
- 24 Mullinix KJ, Leeper TJ, Druckman JN, Freese J. 2015. The generalizability of survey experiments. *J Exp Political Sci*. 2(2): 109–138.
- 25 Litman L, et al. 2021. Did people really drink bleach to prevent COVID-19? A tale of problematic respondents and a guide for measuring rare events in survey data. MedRxiv. 2020-12. <https://doi.org/10.1101/2020.12.11.20246694>, preprint: not peer reviewed.
- 26 Lopez J, Hillygus DS. 2018. Why so serious?: survey trolls and misinformation (SSRN Scholarly Paper No. 3131087). Rochester (NY). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3131087
- 27 Westwood SJ, Grimmer J, Tyler M, Nall C. 2022. Current research overstates American support for political violence. *Proc Natl Acad Sci U S A*. 119(12):e2116870119.
- 28 Kennedy C, et al. 2021. Strategies for detecting insincere respondents in online polling. *Public Opin Quart*. 85(4):1050–1075. Oxford (UK).
- 29 Geraci J. 2022. *POLL-ARIZED: why Americans don't trust the polls and how to fix them before it's too late*. Houndstooth Press.
- 30 Jamieson KH, McNutt M, Kiermer V, Sever R. 2019. Signaling the trustworthiness of science. *Proc Natl Acad Sci U S A*. 116(39): 19231–19236.
- 31 Online Ethics Center. 2022. History and funding. <https://onlineethics.org/history-and-funding>.
- 32 Traugott MW, Lavrakas PJ. 2016. *The voter's guide to election polls*. Lanham (MD): Rowman & Littlefield Publishers.
- 33 Madson G, Cooper A. 2021. 2021 future of survey research conference. Durham (NC): Duke University. <https://sites.duke.edu/surveyresearch/report/>.