

Ethical Considerations for Augmenting Surveys with Auxiliary Data Sources

Bella Struminskaya^{1,*} , Joseph W. Sakshaug²

¹Associate Professor, Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands

²Distinguished Researcher, Department of Statistical Methods, Institute for Employment Research, Nuremberg, Germany; and Professor, Department of Statistics, Ludwig-Maximilian University of Munich, Munich, Germany

Abstract Survey researchers frequently use supplementary data sources, such as paradata, administrative data, and contextual data to augment surveys and enhance substantive and methodological research capabilities. While these data sources can be beneficial, integrating them with surveys can give rise to ethical and data privacy issues that have not been completely resolved. In this research synthesis, we review ethical considerations and empirical evidence on how privacy concerns impact participation in studies that collect these novel data sources to supplement surveys. We further discuss potential approaches for safeguarding participants' data privacy during data collection and dissemination that may assuage their concerns. Finally, we conclude with open questions and suggested avenues for future research.

In 2011, Robert Groves argued that we had recently transitioned into the third era of survey research, an era characterized by designed data supplemented with organic data from the internet and other digital systems (Groves 2011). The abundance of continuously produced process data from digital systems as well as data from smart devices is viewed as a promise to mitigate the downward trend of response rates and increasing risk of nonresponse bias observed across countries and over time (Luiten et al. 2020). Researchers are drawn to the rich and fine-grained auxiliary data sources available, such as survey paradata, administrative data, and contextual data derived from digital traces, apps, sensors, wearables, and geodata. These data are seen as particularly attractive when augmented with self-reported data collected through surveys or in-the-moment questionnaires, such as

*Corresponding author: Bella Struminskaya, Department of Methodology and Statistics, Utrecht University, Padualaan 14, Utrecht 3584CH, The Netherlands; email: b.struminskaya@uu.nl.

Advance Access publication July 12, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of American Association for Public Opinion Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com <https://doi.org/10.1093/poq/nfad030>

Ecological Momentary Assessments (EMA). Augmenting these data with surveys has the potential to improve data quality and overcome recall and social desirability errors of survey self-reports and potentially reduce respondent burden, as some tasks that require commitment and diligence from respondents can be substituted by passive measurement (e.g., tracking travel patterns via smartphone apps instead of relying on self-reported travel diaries on paper or the web).

Recent years have witnessed a surge in “third-era” social and behavioral research that addresses substantive research questions by combining auxiliary data with surveys. The applications are wide ranging: from studying job search behavior using native apps and sensors available on smartphones (Sugie 2018), linking administrative data to investigate social disadvantage (Pattaro, Bailey, and Dibben 2020), exploring the effects of air quality on health through the linkage of survey data and environmental sensor data (English et al. 2022), examining happiness by utilizing contextual data of participants’ geographic locations and EMA data collected through smartphones (MacKerron and Mourato 2013), and many more (see, for example, an overview by Keusch and Conrad 2021). Administrative data is frequently used alongside survey data by National Statistical Offices, such as the US Census Bureau and Statistics Netherlands, and research institutes, such as the German Institute for Employment Research (IAB). Additionally, several infrastructures that link administrative data and survey data have been established, such as Administrative Data Research (ADR) UK in the United Kingdom, the Secure Data Access Center (CASD) in France, and the Open Data Infrastructure for Social Science and Economic Innovations (ODISSEI) in the Netherlands.

The role and demand for auxiliary data in social and behavioral research will continue to grow, as the breadth of substantive research questions that can be addressed through the use of these data sources is plentiful. However, the greater accessibility and usage of these data sources raises potential ethical and data privacy issues, including whether research participants have a clear understanding of what they are consenting to, the purposes for which their data will be used, and how their data privacy will be protected. Such issues can impact the willingness of individuals to participate in survey research studies that use auxiliary data sources. This synthesis complements an earlier one in these pages that focused more on how response to ethical and privacy concerns impacted data quality (Plutzer 2019; see also Fobia et al. 2019; Keusch et al. 2019; Sakshaug et al. 2019a). Here, we highlight recent scholarship in order to provide an overview of the ethical considerations raised by utilizing these data sources in survey research and various approaches that scholars have employed to both understand and address respondents’ data privacy concerns. We conclude by suggesting directions for future development.

Ethics and the Use of Auxiliary Data

Social and behavioral research rests on the premise that research participants should be able to provide accurate and relevant information without fear of negative consequences (Joe et al. 2016). Ethical principles for protecting the rights of research participants are rooted in several documents such as the Declaration of Helsinki (World Medical Association General Assembly 1964), the Belmont Report (1979), and the Menlo Report on ethical principles guiding information and communication technology research issued by the US Department of Homeland Security (Bailey et al. 2012). Some of these reports were created in response to violations of research subjects' rights. In general, the three basic ethical principles for conducting research are: (1) beneficence—minimizing harm and maximizing benefit for individual subjects; (2) justice—ensuring research burdens are shared equitably among groups of subjects; and (3) autonomy—requiring informed consent from research subjects for their participation (Joe et al. 2016, p. 79). The purpose of the autonomy principle is to give research participants agency over the sharing of their collected information.

Paradata, administrative records, smartphone sensor data, app data, and digital trace data are characterized by large-volume and highly granular information on individuals. These features make these kinds of auxiliary data particularly susceptible to confidentiality breaches. Furthermore, it is challenging to anticipate all the purposes for which the information collected about individuals may be used. Due to their relative newness, researchers may be uncertain about the appropriate conduct in digital social research. Salganik (2018) provides three examples in which people's privacy was violated by enrolling them in unethical experiments due to such uncertainty. One controversial study he cites investigated emotional contagion, in which about 700,000 Facebook users were unwittingly subjected to an experiment that may have altered their emotions. Negative words (or positive ones in another experimental condition) were blocked from participants' Facebook feeds without specific consent or third-party ethical review. Since then, the Cambridge Analytica scandal in 2016, in which a British consultancy firm built an app allowing them to construct psychological profiles of Facebook users without their consent and use those profiles for political advertising, resulted in Facebook revising its privacy settings (BBC 2018). Moreover, the EU General Data Protection Regulation or GDPR (EU Data Protection Regulation 2016) changed the way companies can collect and store data about their users, giving individuals knowledge and control over their data, such as the right to be forgotten.

While the ethical principles discussed above are not exclusive to the big-data era, the nature of the data and their unique characteristics (longitudinal, detailed, rich), coupled with their potential unforeseen uses, can pose distinct

risks to research participants both at the stage of collecting auxiliary data and combining them with surveys and at the stage of data dissemination. In the following discussion, we focus on ethical and privacy-related issues for both of these stages.

Informed Consent, Willingness, and Privacy Concerns

As individuals increasingly use smartphones, apps, and online platforms for their daily activities, their knowledge and attitudes toward data collection and uses of such data can influence their willingness to share such data for research purposes. In a recent study, [Turow et al. \(2023\)](#) show that Americans do not understand what it means to consent to companies' use of their data, feel that they have no control over it, and believe they are being harmed by the companies using the data. The study, which used data from a mixed-mode panel of internet users in the United States, found that 80 percent of respondents feel they have little control over how marketers learn about them online, and believe that information that can be learned about them from their online behavior can cause harm. The authors refer to this phenomenon as "digital resignation," wherein individuals desire control over the data companies have on them but feel powerless to do so ([Turow et al. 2015](#)). These beliefs may explain why individuals can be reluctant to provide auxiliary data for research purposes.

[Singer and Couper \(2011\)](#) conducted a vignette experiment to study the effects of informing participants about paradata collection on their willingness to participate in survey data collection. The study used several conditions, ranging from no mention of paradata collection, a simple description of what data was being collected (keystrokes, timestamps, browser information), a simple description with information about the purposes of the data collection, and a simple description with a hyperlink to additional information. The results indicated that none of the conditions fully informed respondents about what paradata is, and requiring informed consent reduced willingness to participate in a survey. Similarly, a recent study by [Henninger et al. \(2023\)](#) in this special issue found that participants were less willing to provide mouse-tracking paradata when the survey participation and paradata collection were requested sequentially as opposed to when the requests were bundled together. The study also found that mentioning specific purposes for the paradata collection did not increase willingness to participate and, in some cases, even decreased it.

Privacy and Confidentiality Concerns and Linkage to Administrative Data

In the context of linking surveys with administrative data, [Sala, Knies, and Burton \(2014\)](#) found that respondents who declined to consent to such

linkages did so primarily due to privacy and data confidentiality concerns. The level of risk perceptions to data linkage appears to be driven by respondents' own privacy beliefs and attitudes toward the data collector and relevant stakeholders. For instance, [Fobia et al. \(2019\)](#) demonstrated that trust in federal statistics, the belief that the federal statistical system keeps information about them confidential and respects people's privacy, and the belief that policymakers need federal statistics to make informed decisions were associated with greater favorability toward data linkage. Additionally, studies have explored how the framing of the data linkage request may impact respondents' attitudes toward linkage and their willingness to link their own data. Various frames have been investigated, including those emphasizing the benefits of data linkage, such as cost savings, improved data accuracy, and reduced respondent burden, among others ([Singer, Bates, and Van Hoewyk 2011](#); [Bates, Wroblewski, and Pascale 2012](#); [Sakshaug, Tutz, and Kreuter 2013](#); [Sakshaug and Kreuter 2014](#)), and those emphasizing the negative aspects of non-linkage, such as reducing the value of the survey data collection ([Kreuter, Sakshaug, and Tourangeau 2016](#); [Sakshaug et al. 2019a](#)). However, the results of framing studies are mixed, not all frames have the same effect on respondents, and some frames have more leverage than others for specific subgroups ([Sakshaug et al. 2019b](#); [Fobia et al. 2020](#)).

Privacy Concerns in Data Collection via Smartphone Apps and Wearables

In studies that collect smartphone sensor data, app data, and wearables data, and link to social media data, research has demonstrated that privacy concerns are a significant barrier to participation. Privacy concerns were the main reason against participation when survey respondents were asked about stated willingness to perform additional tasks, such as provide data from smartphones and accelerometers about their physical activity, provide access to their social media profiles, track geolocation, use measurement devices provided for data collection, etc. ([Revilla et al. 2019](#)), download a research app that would track geolocation, app use on smartphone and send additional questionnaires ([Keusch et al. 2019](#)), share their geolocation, photos, or videos, and download an app for the collection of household expenditure data ([Jäckle et al. 2019](#)). Studies have found that higher privacy concerns are associated with lower willingness to participate in app-based and smartphone sensor data collection ([Keusch et al. 2019](#); [Revilla et al. 2019](#); [Wenz et al. 2019](#); [Struminskaya et al. 2020, 2021](#); [Roberts et al. 2022](#); [Silber et al. 2022](#)), as well as linking survey data to social media data ([Mneimneh et al. 2021](#)). Despite efforts to emphasize privacy and confidentiality protection, such as assuring respondents that their data will be anonymized, will not be accessed by third parties, and will only be used for the purposes of the

research, there was no increase in stated willingness or actual sharing of data collected by smartphone sensors (Struminskaya et al. 2020, 2021). Participants can have multiple concerns regarding data sharing, such as data streams being intercepted by unauthorized parties, multiple streams of data being connected to reidentify anonymous users, and information being used to impact credit, employment, or insurability. Keusch et al. (2019) demonstrated that the more situations respondents perceived as a violation of privacy (e.g., by banks, government, social media), the less willing they were to share smartphone sensor data.

Do Privacy Concerns Vary by the Type of Task?

For tasks that involve passive measurement, such as tracking geolocation and app usage, privacy concerns are higher than for tasks that require active participation, such as taking pictures with a smartphone camera or filling out surveys using a smartphone app (Keusch et al. 2020). This is likely due to the respondents' perceived control over the data collection process. Indeed, research has shown that when participants have more control (perceived or actual) over the data collection process, such as the ability to stop the data collection, view the data before it is shared, and revoke consent if desired, this increases stated willingness to share sensor data (Keusch et al. 2019; Struminskaya et al. 2020) and actually do so (Struminskaya et al. 2021). However, Kreuter et al. (2020) found that only 30 percent of participants who were willing to install a research app that collected information about participants' geolocation, call and messaging logs, physical activity data, social network information, and information on smartphone usage reviewed the descriptions of the individual functions before consenting to data collection. Similar to studies from the data linkage consent literature, studies examining the influence of framing requests on willingness to share sensor data have yielded mixed results. While some studies found no effect of framing the request in terms of time-saving benefits (Struminskaya et al. 2021; Beuthner et al. 2023), others have shown a small negative effect of emphasizing time-saving benefits for sharing of sensor data (Struminskaya et al. 2020).

How Well Do Respondents Understand Consent Requests?

Given that many surveys employ procedures to obtain informed consent from respondents to link their survey data to auxiliary data sources, particularly administrative records, a key question is to what extent respondents understand the consent request and the specific aspects of what they are consenting to. While detailed information may be presented regarding data processing, storage, and security, it's unclear how much of this information is comprehended by respondents and whether their consent is truly informed,

aligning with legal and ethical standards. Qualitative studies indicate that respondent understanding of linkage requests is often low, with participants frequently misunderstanding the purpose of the linkage and the accessibility of their survey responses and linked data to third parties and data custodians (Bates and Pascale 2005; Jäckle et al. 2018; Thornby et al. 2018). Thornby et al. (2018) noted that some participants only skim-read the information about the linkage process and did not always understand the terminology used. However, their study also indicated that respondents who improved their knowledge about data linkage over the course of the interview became more receptive to the idea of consenting to link their own data. Jäckle et al. (2021) found that using easy-to-understand language in the consent question increases understanding of the linkage request. Burton et al. (2021) found that respondents use different decision-making processes when considering linkage consent, and that “reflective” processes are associated with a higher likelihood of consent, better comprehension, and greater confidence in their decision compared to “instinctive” (or “gut feeling”) processes.

Studies using quantitative data have also established a link between understanding the linkage request and the likelihood of giving consent. In one such study, Das and Couper (2014) experimented with providing members of a probability-based online panel in the Netherlands with varying amounts of information regarding the proposed data linkage in an advance letter or email. The information provided included: (1) a standard description of the purpose of the linkage and data protection assurances, or (2) an extended description that contained more details and examples of the linkage process, data access and storage, and planned uses of the data. The respondents were then asked several knowledge questions about the proposed linkage. The study found that respondents who opted out of the linkage and those who received only the standard linkage description tended to give fewer correct answers than those who did not opt out and those who received the extended description. Understanding of the consent request was particularly low with respect to the purpose of the linkage and storage of the linked data and personal identifiers.

More recent studies that have used knowledge questions to assess respondents' comprehension of the consent request have yielded similar findings. Sakshaug et al. (2021) found that the percentage of correct answers was significantly higher among those who consented to linkage compared to those who did not in a German survey. Notably, only about one-third of non-consenters answered correctly that their survey responses would not be merged with administrative data, suggesting that the majority of non-consenters believed their data would be linked anyway, irrespective of their wishes. Jäckle et al. (2021) also found that respondents who had a better understanding of the linkage request were more likely to consent, and that web

respondents had lower levels of understanding compared to face-to-face respondents.

According to early studies on the donation of digital trace data, the level of understanding of consent requests among participants is not very high. For example, [Struminskaya and Boeschoten \(2022\)](#) asked participants of a Dutch online panel to download their Google Semantic Location History and share data on their distance traveled and time spent on different modes of transportation or on foot. About 30 percent stated their willingness to donate their data, and only about 14 percent actually did so. The study also included seven knowledge statements that checked participants' understanding of the request, with an average of 3.23 statements answered correctly. Only 5.5 percent of participants answered all seven statements correctly. Those who answered more correct statements were more likely to agree to donate their data.

Data Privacy-Utility Trade-Off

The literature cited above indicates that privacy concerns are a significant issue for research participants when it comes to sharing their auxiliary data. As the demand and accessibility of these data continue to grow, the magnitude of such concerns may also increase. [Bender et al. \(2017\)](#) argue that individual privacy and confidentiality should be balanced against the social benefits of research access and use. The level of data access must be carefully considered, as the greater the access, the greater the risk to individuals. Additionally, the usefulness of the data must be measured against disclosure risks ([Duncan et al. 2011](#)). This can be thought of in terms of a privacy-utility trade-off, which refers to the inverse relationship between data utility and privacy protection of the study participants. At the two extremes, all data could be accessible and available for analysis, resulting in high utility but low privacy protection, or no one could have access to the data, resulting in low utility but high privacy protection. Researchers must therefore strike a balance between these two extremes: increasing access to the data, and thus increasing the utility of these data, and protecting participants' data privacy.

Privacy risks and data utility are closely connected to the aforementioned ethical principles of beneficence, justice, and autonomy. [Bowen \(2022\)](#) argues that personal privacy loss is not evenly distributed in society, with underrepresented individuals, minorities, and socially and economically disadvantaged individuals experiencing higher levels of privacy insecurity and being at greater risk of privacy breach and reidentification. However, removing their data from publicly released datasets would also mean they miss out on the benefits and impacts of the research performed on the data.

Promising Approaches for Protecting Data Privacy

To protect participants' privacy in individual research projects, researchers can consider privacy-preserving methods during data collection by aggregating or minimizing the amount of data that's collected. For example, if researchers are interested in collecting information about conversation frequency and duration using a smartphone microphone, they might choose to implement privacy-preserving aggregation on the device by using privacy-sensitive classifiers (Rabbi et al. 2011). This allows researchers to obtain only binary information, such as whether a conversation is happening or not, without recording the content of the conversation (see, for example, Wang et al. 2014). In the case of digital trace data, researchers are establishing privacy-preserving infrastructures that locally extract information from data download packages (DDPs) provided by digital platforms, such as Google, Facebook (Meta), Netflix, and others, and then aggregate the data into a relevant form (e.g., distance traveled instead of GPS coordinates, or the number of liked social media posts without revealing the specific content being liked) before transmitting them to researchers (see Boeschoten et al. 2022). Moreover, this infrastructure allows participants to view the aggregated digital trace data before sharing them with researchers. This approach is similar to the data reduction principle in the context of linking social media data to survey data (Sloan et al. 2020), which states that only necessary data should be linked, with "high-risk" variables that may increase the risk of re-identification excluded or aggregated from the raw data.

For the dissemination of privacy-sensitive data to the research community, one solution is to add uncertainty or noise to the data. One such approach is data synthesis, which involves generating synthetic datasets that replicate the structure and statistical properties of the original confidential dataset using statistical modeling and imputation methods. Well-known applications of synthetic data include a synthetic version of the Longitudinal Business Database at the US Census Bureau (Kinney, Reiter, and Miranda 2014) and a synthetic version of the IAB Establishment Panel at the German Institute for Employment Research (Drechsler et al. 2008). Another privacy-preserving approach is differential privacy (see, e.g., Oberski and Kreuter 2020; Bowen 2022), which is a framework that involves adding noise to the data in a way that ensures the outcome of any statistical analysis "is nearly equally likely" regardless of whether any single unit is included or excluded from the dataset (US Census Bureau 2021). Since 2020, the US Census Bureau has adopted formal differential privacy protections in the Decennial Census (see, e.g., the special issue by Gong, Groshen, Vadhan 2022). However, for some research questions, researchers may need access to the original data. One option is to provide them with restricted access. For this option, secure data access platforms are available, such as the Coleridge

Initiative (<https://coleridgeinitiative.org/>) that facilitates data sharing across agencies and states in the United States or the Secure Analysis Environment (SANE) in the Netherlands, a virtual data vault in which researchers can analyze sensitive data while data owners retain complete control (<https://www.surf.nl/en/news/sane-secure-data-environment-for-social-sciences-and-humanities>).

Conclusions

When supplementing survey data with paradata, administrative data, and contextual data from auxiliary sources, apps, sensors, and digital platforms, it is the responsibility of researchers to be aware of the ethical implications of collecting and using these data sources. It is essential to adhere to basic ethical principles, such as collecting and using the data responsibly and ensuring that participants endure no harm during any phase of the research cycle, including data collection, storage, analysis, and dissemination. It is also important to ensure that certain groups are not disadvantaged, and that the burdens of the research are spread evenly. As the sources and uses of these auxiliary data continue to expand, it is crucial for researchers to consider potential ethical implications of the research early in the study design stage. In this research synthesis, we have outlined ethical considerations pertinent to supplementing these data with surveys, including individuals' willingness to participate in such research studies and their comprehension of consent requests, as well as approaches to addressing potential data privacy risks, such as aggregating data during data collection, carefully introducing uncertainty/noise into the data prior to public release, and the use of secure data access platforms. We anticipate that further development of such platforms and new data access modalities will facilitate the transition from highly restricted access options to more open science principles.

In addition to developing new access models and privacy-preserving data collection software and methods, it is essential to gain a deeper understanding of privacy and confidentiality concerns to effectively address them. Helen Nissenbaum's theoretical framework provides a useful perspective, positing that information is not inherently private or public, but is governed by context-specific norms that determine to whom it can be appropriately transmitted and for what purpose (Nissenbaum 2010). Further research is needed to explore the norms, preferences, and interests of all actors involved, including data providers and recipients. A good example is the PERVADE (Pervasive Data Ethics) Project at the University of Maryland, which examines questions related to quantifying risks to data subjects, public attitudes about data reuse, and adapting ethical codes for computational research (<https://pervade.umd.edu/>).

Gaining a better understanding of privacy concerns should go hand-in-hand with communicating to research participants how their privacy is being protected and the safeguards put in place to prevent data disclosure. However, questions remain regarding whether respondents understand the sophisticated measures and approaches taken to protect their privacy, such as differential privacy, and whether providing this level of information will effectively alleviate their concerns. We encourage further research in this domain. We know from previous research that respondent understanding of data requests is associated with willingness to participate and that privacy concerns are the main cause of nonparticipation; thus, if researchers can effectively communicate the approaches that protect participants' privacy, this may increase understanding, transparency, trust, and possibly willingness to participate.

References

- Bailey, M., A. Burstein, K. C. Claffy, S. Clayman, D. Dittrich, J. Heidemann, E. Kenneally, D. Maughan, J. McNeill, P. Neumann, C. Scheper, L. Tien, C. Papadopoulos, W. Visscher, and J. Westby. 2012. *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*. U.S. Department of Homeland Security. https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf.
- Bates, N., and J. Pascale. 2005. "Development and Testing of Informed Consent Questions to Link Survey Data with Administrative Records." In *Proceedings of the Survey Research Methods Section*, 3786–93. Alexandria, VA: American Statistical Association.
- Bates, N., M. J. Wroblewski, and J. Pascale. 2012. "Public Attitudes Toward the Use of Administrative Records in the US Census: Does Question Frame Matter?" *Survey Methodology Series No. 2012–04*. Washington, DC: U.S. Census Bureau.
- BBC. 2018. "Facebook Privacy Settings Revamped After Scandal." <https://www.bbc.com/news/technology-43557803>. Accessed March 28, 2018.
- Belmont Report. 1979. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. hhs.gov/ohrp/humansubjects/guidance/belmont.html. Accessed December 13, 2011.
- Bender, S., R. Jarmin, F. Kreuter, and J. Lane. 2017. "Privacy and Confidentiality." In *Big Data and Social Science*, edited by I. Foster, R. Ghani, R. S. Jarmin, F. Kreuter, and J. Lane, 299–311. Boca Raton, FL: CRC Press.
- Beuthner, C., B. Weiß, H. Silber, F. Kusch, and J. Schröder. 2023. "Consent to Data Linkage for Different Data Domains—The Role of Question Order, Question Wording, and Incentives." *International Journal of Social Research Methodology* 1. <https://doi.org/10.1080/13645579.2023.2173847>.
- Boeschoten, Laura, Jef Ausloos, Judith E. Möller, Theo Araujo, and Daniel L. Oberski. 2022. "A Framework for Privacy Preserving Digital Trace Data Collection Through Data Donation." *Computational Communication Research* 4:388–423.
- Bowen, C. M. 2022. *Protecting your Privacy in a Data-Driven World*. Boca Raton, FL: CRC Press.
- Burton, J., M. P. Couper, T. F. Crossley, A. Jäckle, and S. Walzenbach. 2021. "How Do Survey Respondents Decide Whether to Consent to Data Linkage?" *Understanding Society Working Paper 2021-05*. Colchester: University of Essex.

- Das, M., and M. P. Couper. 2014. "Optimizing Opt-Out Consent for Record Linkage." *Journal of Official Statistics* 30:479–97.
- Drechsler, J., A. Dundler, S. Bender, S. Rässler, and T. Zwick. 2008. "A New Approach for Disclosure Control in the IAB Establishment Panel—Multiple Imputation for a Better Data Access." *AStA Advances in Statistical Analysis* 92:439–58.
- Duncan, G., M. Elliot, and J. J. Salazar-Gonzalez. 2011. *Statistical Confidentiality: Principles and Practice*. New York, NY: Springer.
- English, N., C. Zhao, K. L. Brown, C. Catlett, and K. Cagney. 2022. "Making Sense of Sensor Data: How Local Environmental Conditions Add Value to Social Science Research." *Social Science Computer Review* 40:179–94. <https://doi.org/10.1177/0894439320920601>.
- EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>.
- Fobia, A. C., J. J. Childs, and C. Eggleston. 2020. "Attitudes Toward Data Linkage." In *Big Data Meets Survey Science*, edited by C. A. Hill, P. P. Biemer, T. D. Buskirk, L. Japac, A. Kirchner, S. Kolenikov, and L. E. Lyberg, 683–712. Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9781118976357.ch23>.
- Fobia, A. C., J. Holzberg, C. Eggleston, J. J. Childs, J. Marlar, and G. Morales. 2019. "Attitudes Towards Data Linkage for Evidence-Based Policymaking." *Public Opinion Quarterly* 83: 264–79. <https://doi.org/10.1093/poq/nfz002>.
- Gong R., E. L. Groshen, and S. Vadhan. 2022. "Harnessing the Known Unknowns: Differential Privacy and the 2020 Census." *Harvard Data Science Review* 1–14. <https://doi.org/10.1162/99608f92.cb06b469>.
- Groves, R. M. 2011. "Three Eras of Survey Research." *Public Opinion Quarterly* 75:861–71. <https://doi.org/10.1093/poq/nfr057>.
- Henninger, F., P. Kieslich, A. Fernandez-Fontelo, S. Greven, and F. Kreuter. 2023. "Privacy Attitudes Toward Mouse-Tracking Paradata Collection." *Public Opinion Quarterly*.
- Jäckle, A., K. Beninger, J. Burton, and M. P. Couper. 2018. "Understanding Data Linkage Consent in Longitudinal Surveys." Understanding Society Working Paper Series 2018-07, Understanding Society at the Institute for Social and Economic Research.
- Jäckle, A., J. Burton, M. P. Couper, T. F. Crossley, and S. Walzenbach. 2021. "Understanding and Improving Data Linkage Consent in Surveys." <https://www.understandingsociety.ac.uk/sites/default/files/downloads/library/working-papers/2021-07/USWp-2021-07.pdf>.
- Jäckle, A., J. Burton, M. P. Couper, and C. Lessof. 2019. "Participation in a Mobile App Survey to Collect Expenditure Data as Part of a Large-Scale Probability Household Panel: Coverage and Participation Rates and Biases." *Survey Research Methods* 13:23–44. <http://doi.org/10.18148/srm/2019.v1i1.7297>.
- Joe, Kathy, Finn Raben, and Adam Phillips. 2016. "The Ethical Issues of Survey and Market Research." In *The SAGE Handbook of Survey Methodology*, edited by C. Wolf, D. Joye, T. W. Smith, and Y. Fu, 77–86. London: SAGE Publications Ltd. <https://doi.org/10.4135/9781473957893>.
- Keusch, F., and F. G. Conrad. 2021. "Using Smartphones to Capture and Combine Self-Reports and Passively Measured Behavior in Social Research." *Journal of Survey Statistics and Methodology* 10:863–85. <https://doi.org/10.1093/jssam/smab035>.
- Keusch, F., B. Struminskaya, C. Antoun, M. P. Couper, and F. Kreuter. 2019. "Willingness to Participate in Passive Mobile Data Collection." *Public Opinion Quarterly* 83:210–35. <https://doi.org/10.1093/poq/nfz007>.
- Keusch, F., B. Struminskaya, F. Kreuter, and M. Weichbold. 2020. "Combining Active and Passive Mobile Data Collection: A Survey of Concerns." In *Big Data Meets Survey Science*,

- edited by C. Hill, P. Biemer, T. Buskirk, L. Japac, A. Kirchner, S. Kolenikov, and L. Lyberg, 657–82. Hoboken, NJ: John Wiley & Sons.
- Kinney, S. K., J. P. Reiter, and J. Miranda. 2014. “Synlbd 2.0: Improving the Synthetic Longitudinal Business Database.” *Statistical Journal of the IAOS* 30:129–35.
- Kreuter, F., G.-C. Haas, F. Keusch, S. Bähr, and M. Trappmann. 2020. “Collecting Survey and Smartphone Sensor Data with an App: Opportunities and Challenges Around Privacy and Informed Consent.” *Social Science Computer Review* 38:533–49. <https://doi.org/10.1177/0894439318816389>.
- Kreuter, F., J. W. Sakshaug, and R. Tourangeau. 2016. “The Framing of the Record Linkage Consent Question.” *International Journal of Public Opinion Research* 28:142–52. <https://doi.org/10.1093/ijpor/edv045>.
- Luiten, A., Joop Hox, and Edith de Leeuw. 2020. “Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys.” *Journal of Official Statistics* 36:469–87. <https://doi.org/10.2478/jos-2020-0025>.
- MacKerron, G., and S. Mourato. 2013. “Happiness Is Greater in Natural Environments.” *Global Environmental Change* 23:992–1000.
- Mneimneh, Zeina N., Colleen McClain, Ronny Bruffaerts, and Yasmin A. Altwaijri. 2021. “Evaluating Survey Consent to Social Media Linkage in Three International Health Surveys.” *Research in Social & Administrative Pharmacy: RSAP* 17:1091–100.
- Nissenbaum, H. 2010. *Privacy in Context*. Stanford, CA: Stanford University Press.
- Oberski, D. L., and F. Kreuter. 2020. “Differential Privacy and Social Science: An Urgent Puzzle.” *Harvard Data Science Review* 2. <https://doi.org/10.1162/99608f92.63a22079>.
- Pattaro, S., N. Bailey, and C. Dibben. 2020. “Using Linked Longitudinal Administrative Data to Identify Social Disadvantage.” *Social Indicators Research* 147:865–95. <https://doi.org/10.1007/s11205-019-02173-1>.
- Plutzer, E. 2019. “Privacy, Sensitive Questions, and Informed Consent: Their Impacts on Total Survey Error, and the Future of Survey Research.” *Public Opinion Quarterly* 83:169–84.
- Rabbi, M., S. Ali, T. Choudhury, and E. Berke. 2011. “Passive and In-Situ Assessment of Mental and Physical Well-Being using Mobile Sensors.” *Proceedings of the ACM International Conference on Ubiquitous Computing, Ubicomp (Conference)* 2011:385–94. <https://doi.org/10.1145/2030112.2030164>.
- Revilla, M., M. P. Couper, and C. Ochoa. 2019. “Willingness of Online Panelists to Perform Additional Tasks.” *Methods, Data, Analyses* 13:223–52. <http://doi.org/10.12758/mda.2018.01>.
- Roberts, Caroline, Jessica M. E. Herzing, Jimena Sobrino Piazza, Philip Abbet, and Daniel Gatica-Perez. 2022. “Data Privacy Concerns as a Source of Resistance to Complete Mobile Data Collection Tasks Via a Smartphone App.” *Journal of Survey Statistics and Methodology* 10:518–48.
- Sakshaug, J. W., and F. Kreuter. 2014. “The Effect of Benefit Wording on Consent to Link Survey and Administrative Records in a Web Survey.” *Public Opinion Quarterly* 78: 166–76.
- Sakshaug, J. W., A. Schmucker, F. Kreuter, M. P. Couper, and E. Singer. 2019a. “The Effect of Framing and Placement on Linkage Consent.” *Public Opinion Quarterly* 83:289–308.
- Sakshaug, J. W., J. Stegmaier, M. Trappmann, and F. Kreuter. 2019b. “Does Benefit Framing Improve Record Linkage Consent Rates? A Survey Experiment.” *Survey Research Methods* 13:289.
- Sakshaug, J. W., A. Schmucker, F. Kreuter, M. P. Couper, and L. Holtmann. 2021. “Respondent Understanding of Data Linkage Consent, Survey Methods: Insights from the Field.” <https://surveyinsights.org/?p=15407>.
- Sakshaug, J., V. Tutz, and F. Kreuter. 2013. “Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data.” *Survey Research Methods* 7: 133–44.

- Sala, E., G. Knies, and J. Burton. 2014. "Propensity to Consent to Data Linkage: Experimental Evidence on the Role of Three Survey Design Features in a UK Longitudinal Panel." *International Journal of Social Research Methodology* 17:455–73. <https://doi.org/10.1080/13645579.2013.876809>.
- Salganik, M. 2018. *Bit by Bit: Social Research in the Digital Age*. Princeton: Princeton University Press.
- Silber, H., F. Gerdon, R. Bach, C. Kern, F. Keusch, and F. Kreuter. 2022. "A Preregistered Vignette Experiment on Determinants of Health Data Sharing Behavior: Willingness to Donate Sensor Data, Medical Records, and Biomarkers." *Politics and the Life Sciences: The Journal of the Association for Politics and the Life Sciences* 41:161–81. <https://doi.org/10.1017/pls.2022.15>.
- Singer, E., N. Bates, and J. Van Hoewyk. 2011. "Concerns about Privacy, Trust in Government, and Willingness to Use Administrative Records to Improve the Decennial Census." In Annual Meeting of the American Association for Public Opinion Research, Phoenix.
- Singer, E., and M. P. Couper. 2011. Ethical Considerations in Internet Surveys. In *Social and Behavioral Research and the Internet*, edited by M. Das, P. Ester, and L. Kaczmirek, 133–62. London: Routledge Taylor & Francis Group.
- Sloan, Luke, Curtis Jessop, Tarek Al Baghal, and Matthew Williams. 2020. "Linking Survey and Twitter Data: Informed Consent, Disclosure, Security, and Archiving." *Journal of Empirical Research on Human Research Ethics* 15:63–76.
- Struminskaya, B., and L. Boeschoten. 2022. "Willingness and Nonparticipation Biases in Data Donation." *Paper presented at the ODISSEI Conference*, Utrecht, Netherlands, November 3.
- Struminskaya, B., P. Lugtig, V. Toepoel, B. Schouten, D. Giesen, and R. Dolmans. 2021. "Sharing Data Collected with Smartphone Sensors: Willingness, Participation, and Non-Participation Bias." *Public Opinion Quarterly* 85:423–62. <https://doi.org/10.1093/poq/nfab025>.
- Struminskaya, B., V. Toepoel, P. Lugtig, M. Haan, A. Luiten, and B. Schouten. 2020. "Understanding Willingness to Share Smartphone-Sensor Data." *Public Opinion Quarterly* 84: 725–59. <https://doi.org/10.1093/poq/nfaa044>.
- Sugie, Naomie. 2018. "Work as Foraging: A Smartphone Study of Job Search and Employment after Prison." *American Journal of Sociology* 123:1453–91.
- Thornby, M., L. Calderwood, M. Kotecha, K. Beninger, and A. Gaia. 2018. "Collecting Multiple Data Linkage Consents in a Mixed-Mode Survey: Evidence from a Large-Scale Longitudinal Study in the UK." *Survey Methods: Insights from the Field* 1–14. <https://doi.org/10.13094/SMIF-2018-00006>.
- Turow, J., M. L. Hennessy, and N. Draper. 2015. "The Tradeoff Fallacy: How Marketers Are Misrepresenting American Consumers and Opening Them Up to Exploitation." June 26. <https://ssrn.com/abstract=2820060> or <http://dx.doi.org/10.2139/ssrn.2820060>.
- Turow, J., Y. Lelkes, N. A. Draper, and A. E. Waldman. 2023. "Americans Can't Consent to Companies' Use of Their Data: They Admit They Don't Understand It, Say They're Helpless to Control It, and Believe They're Harmed When Firms Use Their Data—Making What Companies Do Illegitimate." Report, Annenberg School for Communication, University of Pennsylvania. February 15. <https://ssrn.com/abstract=4391134> or <http://dx.doi.org/10.2139/ssrn.4391134>.
- U.S. Census Bureau. 2021. *Disclosure Avoidance for the 2020 Census: An Introduction*. U.S. Government Publishing Office, Washington, DC.
- Wang, R., F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and T. A. Campbell. 2014. "Student Life: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones." In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3–14.

Wenz, A., A. Jäckle, and M. P. Couper. 2019. "Willingness to Use Mobile Technologies for Data Collection in a Probability Household Panel." *Survey Research Methods* 13:1–22. <https://doi.org/10.18148/srm/2019.v1i1.7298>.

World Medical Association General Assembly. 1964. "Declaration of Helsinki: Recommendations Guiding Doctors in Clinical Research." <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>.