Trappmann, Mark; Schwarz, Stefan; Haas, Georg-Christoph; Malich [...]

Augmenting survey data with digital trace data : Is there a threat to panel retention?

Date of secondary publication: 29.06.2023

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-599005

Primary publication

Trappmann, Mark; Schwarz, Stefan; Haas, Georg-Christoph; Malich [...]: Augmenting survey data with digital trace data : Is there a threat to panel retention?. In: Journal of survey statistics and methodology : JSSAM. 11 (2022), 3, pp. 541–552. DOI: 10.1093/jssam/smac023

Journal of Survey Statistics and Methodology (2023) ***11***, 541–552

AUGMENTING SURVEY DATA WITH DIGITAL TRACE DATA: IS THERE A THREAT TO PANEL RETENTION?

MARK TRAPPMANN*
GEORG-CHRISTOPH HAAS
SONJA MALICH
FLORIAN KEUSCH
SEBASTIAN BÄHR
FRAUKE KREUTER
STEFAN SCHWARZ

Linking digital trace data to existing panel survey data may increase the overall analysis potential of the data. However, producing linked products often requires additional engagement from survey participants through consent or participation in additional tasks. Panel operators may worry that such additional requests may backfire and lead to lower panel retention, reducing the analysis potential of the data. To examine these concerns, we conducted an experiment in the German PASS panel survey after wave 11. Three quarters of panelists ($n = 4{,}293$) were invited to install a research app and to provide sensor data over a period of 6

MARK TRAPPMANN is a Head of Research Unit with the "Panel Study Labour Market and Social Security," Institute for Employment Research, Nuremberg, Germany and Professor of Sociology, Especially Survey Methodology, University of Bamberg, Bamberg, Germany. GEORG-CHRISTOPH HAAS is a Senior Researcher, and SONJA MALICH is a Doctoral Student with the Institute for Employment Research, Nuremberg, Germany. FLORIAN KEUSCH is Professor of Social Data Science and Methodology with the University of Mannheim, Mannheim, Germany. SEBASTIAN BÄHR is a Senior Researcher with the Institute for Employment Research, Nuremberg, Germany. FRAUKE KREUTER is Professor at the Joint Program in Survey Methodology and Co-Director of the Social Data Science Center with the University of Maryland, College Park, MD, USA, and Professor of Statistics and Data Science for the Social Sciences and Humanities at the Ludwig-Maximilians-University of Munich, Munich, Germany. STEFAN SCHWARZ is a Doctoral Student with the Institute for Employment Research, Nuremberg, Germany.The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Georg-Christoph Haas' time was supported by the German Research Foundation (DFG) through the Collaborative Research Center SFB 884 "Political Economy of Reforms" (Project A9) [139943784 to Markus Frölich, Florian Keusch, and Frauke Kreuter].

*Address correspondence to Mark Trappmann, "Panel Study Labour Market and Social Security," Institute for Employment Research, Germany; E-mail: mark.trappmann@iab.de.

https://doi.org/10.1093/jssam/smac023 Advance access publication 28 June 2022

months, while one quarter ($n = 1{,}428$) did not receive an invitation. We find that the request to install a smartphone app and share data significantly decreases panel retention in the wave immediately following the invitation by 3.3 percentage points. However, this effect wears off and is no longer significant in the second and third waves after the invitation. We conclude that researchers who run panel surveys have to take moderate negative effects on retention into account but that the potential gain likely outweighs these moderate losses.

KEYWORDS: Attrition; Digital trace data; Panel survey; Privacy; Respondent burden.

Statement of Significance

Our research provides evidence of the effects on panel retention when existing panels are used as sampling frame and recruitment pool for app-based research studies. Fears of decreased panel retention, held by many panel providers, are not unjustified, however the effects we observed leveled off over time.

1. INTRODUCTION

One of the most promising new developments in social science is the combination of survey data and digital trace data, for example, smartphone usage data (Japec, Kreuter, Berg, Biemer, Decker, et al. 2015; Lazer and Radford 2017; Callegaro and Yan 2018).

From the perspective of survey research, digital trace data can be used to enrich existing survey data with nonreactive behavioral data of high velocity and frequency. Digital trace data are valued for nonintrusive data collection, high precision, and granularity (Stier, Breuer, Siegers, and Thorson 2020). This approach allows for the collection of data in realms that have traditionally been difficult to collect in surveys due to the burden placed on respondents and the difficulty of correct recall. Typical examples are data on locations and activities, communication, or app use.

Flipping the perspective, a majority of studies using only digital trace data focus on special populations such as Twitter users (Vaccari, Valeriani, Barberá, Bonneau, Jost, et al. 2015), students (Wang, Chen, Chen, Li, Harari, et al. 2014; Sapiezynski, Stopczynski, Lassen, and Lehmann 2019), or patients with a specific illness (Wang, Wang, Aung, Ben-Zeev, Brian, et al. 2017; Goodspeed, Yan, Hardy, Vydiswaran, Berrocal, et al. 2018), limiting the external validity of their results to the general population. Other studies rely on convenience samples of volunteers who self-select into the study (MacKerron and Mourato 2013; Lathia, Sandstrom, Mascolo, and Rentfrow 2017), making it impossible to assess coverage and participation biases. One obvious advantage of combining digital trace data studies with survey data collection using probability samples of the general population is that it increases the external validity and facilitates statistical inference of the digital trace data (Lazer and Radford 2017). Furthermore, design can be introduced to otherwise organic data, such as the ability to assess errors of representation and separate them into coverage and nonparticipation errors (Keusch, Bähr, Haas, Kreuter, and Trappmann 2020).

However, collecting and linking digital trace data to survey data, if performed on an individual level, requires informed consent. For panel surveys with repeated measurements, requests to collect and link digital trace data may lead to a decrease in retention rates. Some panel members may feel that a line has been crossed when there are requests for access to potentially highly sensitive data or they may find an additional request too burdensome. Thus, digital trace data collection may undermine the statistical power of longitudinal analyses and introduce attrition bias within a panel survey. Our study evaluates the existence and magnitude of the potential consequences of such requests.

We conducted an experiment with panelists from the German "Panel Study Labour Market and Social Security" (PASS, Trappmann, Bähr, Beste, Eberl, Frodermann et al. 2019) to investigate whether panel retention is affected by the request to participate in a smartphone app study. While 1,428 eligible panelists did not receive an invitation, we invited 4,293 panelists to install the app from the Institute for Employment Research (IAB-SMART app) that requested access to sensor and log file data over 6 months (Kreuter, Haas, Keusch, Bähr, and Trappmann 2020).

In the following sections, we review the few studies that evaluate the effects of similar requests on panel retention and derive expectations on how our request affects retention. We briefly present the PASS data and the design of the IAB-SMART study, describe the analysis methods, and report results. Finally, we discuss our findings concerning limitations and implications for future studies combining survey data with digital trace data.

2. EFFECTS OF ADDITIONAL BURDEN AND CONSENT REQUESTS ON PANEL RETENTION

Past research provides several suggestions on why inviting panelists to a smartphone data collection and requesting multiple data collection consents can affect retention. Several researchers hypothesized that an additional or large burden can decrease respondents' motivation, resulting in a decreasing likelihood of response and panel retention (e.g., Hartmann and Krug 2009; Marmot and Steptoe 2008; Pashazadeh, Cernat, and Sakshaug 2021; Silberstein and Jacobs 1989; Weir 2008). However, with respect to nonresponse in subsequent waves of panel surveys, studies do not find that interview length has an effect on panel retention (e.g. Hill and Willis 2001; Hart, Rennison, and Gibson 2005; Lynn 2014).

Conceptualizing burden as a function of multiple factors, such as requested time, effort, and potential interest (Kantorowitz 1998; Deeg, Van Tilburg, Smit, and De Leeuw 2002) is likely better suited to foreshadow the effects on retention with additional participation requests than mere survey length. Burden due to supplemental surveys requests did not predict panel retention overall (e.g., Kantorowitz 1998; Deeg et al. 2002; Kleinert, Christoph, and Ruland 2015), though respondents who declined to participate in a supplemental study or an additional task were found to be less likely to be retained (Deeg et al. 2002; Sastry, Fomby, and McGonagle 2016; Kleinert et al. 2015). An underlying mechanism may be the panelists' self-selection into a group of committed and therefore "faithful" panelists and a group who attrites more easily (Hoogendoorn and Sikkel 1998; Sastry et al. 2016).

Digital trace data from mobile phones, such as geolocation and app usage data, can be considered particularly sensitive, and their collection potentially threatens respondents' privacy (Keusch, Struminskaya, Kreuter, and Weichbold 2021). Keusch, Struminskaya, Antoun, Couper, and Kreuter (2019) find that asking respondents a hypothetical question about participating in a six-month passive mobile data collection (compared to one-month) duration leads to reduced willingness to participate. Experimentally including information about whether the research app additionally asks survey questions in the participation request had no effect on respondents' willingness. The authors interpret this as an indication that the participation decision is driven by privacy and security concerns about the data collection rather than by the additional effort required. Other studies also report privacy and security concerns as most often mentioned reasons for not being willing to participate (Jäckle, Burton, Couper, and Lessof 2019; Revilla, Couper, and Ochoa 2019; Struminskaya, Lugtig, Toepoel, Schouten, Giesen, et al. 2021).

Most of the existing literature does not consider the long-term effects of actual requests on subsequent panel attrition, as we aim to do, and is therefore only partially comparable to our research. One exception is Eisnecker and Kroh (2017) who conducted the only experimental study we are aware of in the first and second waves of data collection in a migrant refreshment sample of the German Socio-Economic Panel (SOEP). They found that the consent request to link administrative employment and benefits records increased attrition between waves one and two by approximately 1.5 percentage points compared to a group that was not asked to consent to data linkage.

Despite differences with respect to the exact type of data, the target population and the maturity of the panel—our panel being more than a decade old, our research question is similar to Eisnecker and Kroh (2017) in that both studies investigate the effect of asking for permission to link potentially highly sensitive data on panel retention.

In summary, we expect that requesting panelists to install a smartphone app that collects a broad range of sensitive data for a period of six months decreases panel retention in subsequent waves.

3. METHODS AND DATA

To answer our research question whether the invitation to an app data collection has a negative effect on panel retention, we combine data from the PASS (Trappmann et al. 2019) and the IAB-SMART study (Kreuter et al. 2020).

3.1 PASS

The PASS is a mixed-mode panel survey of the residential population in Germany with annual waves since 2007 that uses computer-assisted telephone (CATI) and personal interviews (CAPI). PASS is a major data source for research into unemployment, poverty, and the evaluation of the welfare benefits system. For all cases, the mode of the previous interview becomes the starting mode in the following wave. From wave 4 on, CAPI became the starting mode for all new samples (Trappmann, Müller, and Bethmann 2013). PASS is operated as a household survey; that is, within each household, all eligible respondents aged 15 and older are targeted for an interview.

PASS uses a dual-frame sampling design to oversample households receiving welfare benefits. A register sample of benefit recipients is refreshed yearly by a sample of new entries to benefit receipt (Trappmann et al. 2019). The questionnaire addresses labor market participation, job search, income, deprivation, and social inclusion. After two consecutive waves of nonresponse, a household is no longer issued for fieldwork (for further details, see https://fdz.iab.de/en/FDZ_Individual_Data/PASS/PASS-SUF0619v2.aspx).

3.2 IAB-SMART

By using an Android app, the IAB-SMART study passively collected a broad range of sensor data from the user's smartphone and enabled regular short surveys (some of them targeted by location, e.g., Haas, Trappmann, Keusch, Bähr, and Kreuter 2020). Passive data collection was grouped into five permissions: (1) mobile phone network quality and location information, (2) interaction history, (3) characteristics of the social network, (4) activity data, and (5) smartphone usage. Before any data could be collected, participants had to actively consent to each permission (see Kreuter et al. 2020 for more detail).

Invitations to the IAB-SMART study were sent out to PASS wave 11 respondents who had indicated that they owned an Android smartphone. The sample was limited to adults who had not reached the legal age for retirement in Germany, that is, aged between 18 and 64, and who were interviewed in the German language.

Among those eligible for the IAB-SMART study ($n = 5{,}721$), a randomly selected 75 percent sample ($n = 4{,}293$) was invited to participate in the IAB-SMART study (treatment group). The other 25 percent ($n = 1{,}428$) were not invited (control group). In total, 687 of 4,293 invited PASS participants installed the app (15.9 percent).

We used relatively generous incentives to motivate PASS participants to install the app and share their data. The exact incentive amount was assigned experimentally (compare Haas, Kreuter, Keusch, Trappmann, and Bähr 2021 for design and results). Participants could earn between 40 and 80 Euros by installing the app and giving all permissions for passive data collection, while the regular PASS incentive was 10 Euros for a 45 minute interview.

3.3 Analysis Strategy

Our focus lies on identifying causal effects of the invitation to the IAB-SMART study on panel retention. We measure panel retention as the proportion of wave 11 participants that participated in waves 12, 13, and 14, respectively. By including all wave 11 respondents in the denominator and only complete interviews in the numerator, we apply a strict definition of panel retention in the spirit of the minimum response rate RR1 according to AAPOR standards (American Association for Public Opinion Research 2016). As we employed an experimental design, the difference in outcomes between the treatment and control groups directly informs us about the treatment effect. To test whether inviting panelists to the IAB-SMART study has a substantial effect on panel retention, we compared retention rates in waves 12–14 between invited and not invited wave 11 participants. We compute retention rates as the proportion of wave 11 participants who participated in the consequent wave. This enables us to assess the immediate effect in the wave directly after the invitation (wave 12) as well as cumulative effects over all three waves.

As response in each wave is highly correlated within households, we control for clustering of respondents within households by using regressions of response on a treatment indicator with cluster-robust standard errors using the sandwich-estimator (Huber 1967; White 1980, 1982) as implemented in Stata Version 15.1 (option vce(cluster) of the regress command) and assess statistical significance based on two-sided *t*-tests. We use a linear probability model (and not logistic regression), as its effect estimate is equivalent to the percentage point difference between the probability of a positive outcome in the treatment and the control group (see Angrist and Pischke 2008; Gomila 2021).

4. RESULTS

Figure 1 shows the retention rates for wave 11 PASS participants in waves 12–14 by treatment and control group. We found that compared to the control group, retention rates in all three waves were lower for those who were invited to IAB-SMART. While the difference of −3.3 percentage points (p.p.) for

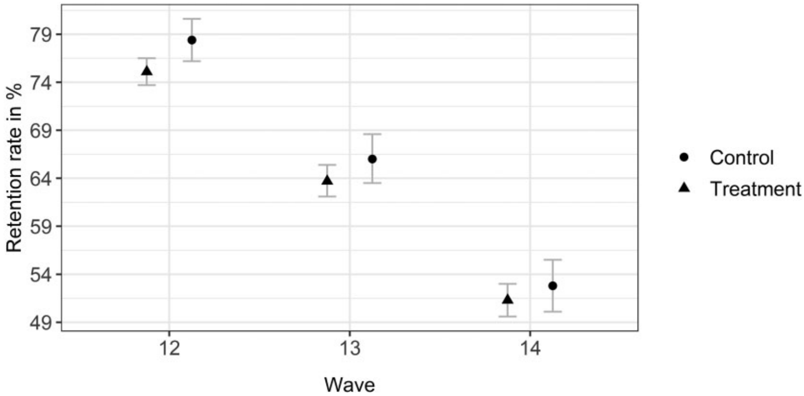

Figure 1. Retention Rates for All Participants in Waves 12, 13, and 14 by Treatment.

wave 12 is statistically significant ($p < .01$), the differences of -2.3 p.p. for wave 13 ($p = .114$) and -1.5 p.p. for wave 14 ($p = .331$) are not.

The pattern supports our hypothesis that the invitation to participate in a smartphone study collecting sensitive data harmed participation in wave 12. The effect, however, diminishes with waves 13 and 14, indicating that no long-term effect exists.

A deeper look into the data (table 1) suggests that the negative short-term effect of the invitations does not seem to be driven by a negative experience of those who participate in the IAB-SMART study but rather by those who were invited and decided not to participate. In all three subsequent waves, IAB-SMART participants were the group that was most likely to participate in PASS, followed by those who were not invited. Those who were invited and did not participate showed the lowest panel retention rate. While all these differences are highly significant ($p < .01$), they may not be interpreted causally but are most likely the result of the increased motivation of those who participate in supplementary studies, as has been observed in earlier studies (Deeg et al. 2002; Sastry et al. 2016; Kleinert et al. 2015).

5. DISCUSSION

Asking panelists to provide additional data increases the analysis potential of the main study. However, each additional data request may be perceived by respondents as an additional burden or as a threat to their privacy. As a result, panelists may decide to attrite, leading to lower case numbers and a decrease in analytical power of the main study. We implemented a request to install a research app that collects sensitive information such as location, interaction history, and app usage shortly after the 11th wave of the German PASS panel. By

Table 1. Retention Rates of PASS Panelists in PASS Wave 12, 13, and 14, Broken Out by Those Panelists Participating in IAB-SMART, Those Invited but Declining to Participate, and Those Never Invited

	Retention rate of PASS panelists participating in IAB-SMART (P)	Retention rate of PASS panelists invited to IAB-SMART, but declined participation ($\bar{P}$)	Retention rate of those PASS panelists not invited ($\bar{I}$) to participate in IAB-SMART	Difference in retention between participating PASS panelists and those not invited to IAB-SMART (P) − ($\bar{I}$) (*p* value from two-sided *t*-test in parentheses)	Difference in retention between non-participating PASS panelists and those not invited to IAB-SMART ($\bar{P}$) − ($\bar{I}$) (*p* value from two-sided *t*-test in parentheses)
Wave 12	87.6	73.0	78.4	9.2 (0.000)	−5.4 (0.000)
Wave 13	79.5	61.1	66.0	13.4 (0.000)	−5.1 (0.001)
Wave 14	68.5	48.4	52.8	15.7 (0.000)	−4.4 (0.005)

applying an experimental design, we were able to evaluate the effect of the invitation on panel retention.

In summary, our results indicate that inviting panelists to the IAB-SMART study has a short-term effect on retention rates in the subsequent wave. The good news for survey practitioners, however, is that the initial negative effect vanishes in later waves. One possible explanation for this is that those who attrite were less committed to the panel even prior to the request and would have attrited sooner or later anyway. Therefore, long-run consequences are rather small and the potential gain from additional data likely outweighs the small loss of panelists.

The experimental design does not allow us to answer the question of whether the participation experience in the additional data collection efforts itself has a positive or negative effect on attrition. We focus on the invitation because participation is systematically confounded by motivation. IAB-SMART participants were much more likely to participate in the subsequent panel wave than nonparticipants. However, this is most likely the case because they are more motivated and devoted study participants and is not due to the positive experience of participating in the extra study.

Unfortunately, our case numbers are not sufficient to differentiate our results by subgroups. Therefore, we cannot address otherwise interesting research questions such as whether the effect of our invitation on panel retention differs between subgroups defined by target variables of the survey and thus decreases or increases bias. With four equal-sized subgroups, an effect of approximately 6 percentage points would be required to achieve statistical significance. Our invitation to an additional study does not create effects of this magnitude.

A related question that we are not able to answer is whether the invitation effect differs between groups with different past panel experiences (e.g., duration, response propensity, survey mode). It is nevertheless interesting that despite the overall maturity of the panel, PASS being in its eleventh wave at the time of the request, effects similar to that of Eisnecker and Kroh (2017) were observed who conducted their experiment after the first recruitment wave.

Another question we have to leave to future research is whether the design of invitations to such additional studies can be optimized with respect to panel retention. The timing and the amount of incentives offered and whether the incentive is conditional or unconditional might play a crucial role. Larger samples and experiments designed to detect the differences between these conditions are needed to guide future studies in optimal design decisions.

DATA AVAILABILITY

Data used in this article are available at the Institute for Employment Research (IAB). Up-to-date access information can be found at https://www.iab.de/en/

daten.aspx. For all analyses, author-originated code under Stata Version 15.1 was used. The code is available and archived at IAB. Our study and data were not preregistered before analysis.

REFERENCES

American Association for Public Opinion Research. (2016), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, 9th edition, AAPOR.

Angrist, J. D., and J. S. Pischke (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ: Princeton University Press.

Callegaro, M., and Y. Yang (2018), "The Role of Surveys in the Era of "Big Data"," in *The Palgrave Handbook of Survey Research*, ed. D. Vannette and J. Krosnick, pp. 175–192, Cham: Palgrave Macmillan.

Deeg, D. J., T. Van Tilburg, J. H. Smit, and E. D. De Leeuw (2002), "Attrition in the Longitudinal Aging Study Amsterdam: The Effect of Differential Inclusion in Side Studies," *Journal of Clinical Epidemiology*, 55, 319–328.

Eisnecker, P. S., and M. Kroh (2017), "The Informed Consent to Record Linkage in Panel Studies: Optimal Starting Wave, Consent Refusals, and Subsequent Panel Attrition," *Public Opinion Quarterly*, 81, 131–143.

Gomila, R. (2021), "Logistic or Linear? Estimating Causal Effects of Experimental Treatments on Binary Outcomes Using Regression Analysis," *Journal of Experimental Psychology: General*, 150, 700–709.

Goodspeed, R., X. Yan, J. Hardy, V. V. Vydiswaran, V. J. Berrocal, P. Clarke, D. M. Romero, I. N. Gomez-Lopez, and T. Veinot (2018), "Comparing the Data Quality of Global Positioning System Devices and Mobile Phones for Assessing Relationships between Place, Mobility, and Health: Field Study," *JMIR mHealth and uHealth*, 6, e168.

Haas, G. C., F. Kreuter, F. Keusch, M. Trappmann, and S. Bähr (2021), "Effects of Incentives in Smartphone Data Collection," in *Big Data Meets Survey Science: A Collection of Innovative Methods*, ed. C. A. Hill, P. P. Biemer, T. D. Buskirk, L. Japec, A. Kirchner, S. Kolenikov and L. E. Lyberg, pp. 387–414, Hoboken, NJ: John Wiley & Sons.

Haas, G. C., M. Trappmann, F. Keusch, S. Bähr, and F. Kreuter (2020), "Using Geofences to Collect Survey Data: Lessons Learned from the IAB-SMART Study," *Survey Methods: Insights from the Field*, 2020, 1–12.

Hart, T. C., C. M. Rennison, and C. Gibson (2005), "Revisiting Respondent 'Fatigue Bias' in the National Crime Victimization Survey," *Journal of Quantitative Criminology*, 21, 345–363.

Hartmann, J., and G. Krug (2009), "Verknüpfung von Personenbezogenen Prozess-und Befragungsdaten–Selektivität Durch Fehlende Zustimmung der Befragten?," *Zeitschrift für ArbeitsmarktForschung*, 42, 121–139.

Hill, D. H., and R. J. Willis (2001), "Reducing Panel Attrition: A Search for Effective Policy Instruments," *Journal of Human Resources*, 36, 416–438.

Hoogendoorn, A. W., and D. Sikkel (1998), "Response Burden and Panel Attrition," *Journal of Official Statistics*, 14, 189.

Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221–233.

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.

Japec, L., F. Kreuter, M. Berg, P. Biemer, P. Decker, C. Lampe, J. Lane, C. O'Neil, and A. Usher (2015), "Big Data in Survey Research: AAPOR Task Force Report," *Public Opinion Quarterly*, 79, 839–880.

Kantorowitz, M. (1998), "Is It True That Nonresponse Rates in a Panel Survey Increase When Supplement Surveys Are Annexed?," *International Workshop on Household Survey Nonresponse*, 4, 121–137.

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, 210–235.

Keusch, F., B. Struminskaya, F. Kreuter, and M. Weichbold (2021), "Combining Active and Passive Mobile Data Collection: A Survey of Concerns," in *Big Data Meets Survey Science: A Collection of Innovative Methods*, ed. C. A. Hill, P. P. Biemer, T. D. Buskirk, L. Japec, A. Kirchner, S. Kolenikov, and L. E. Lyberg, pp. 657–682, Hoboken, NJ: Wiley.

Keusch, F., S. Bähr, G. C. Haas, F. Kreuter, and M. Trappmann (2020), "Coverage Error in Data Collection Combining Mobile Surveys with Passive Measurement Using Apps: Data from a German National Survey," *Sociological Methods and Research*, 1–38, DOI: 10.1177/0049124120914924.

Kleinert, C., B. Christoph, and M. Ruland (2015), "Auswirkungen der Administration von Kompetenztests im Rahmen einer Panelerhebung für Erwachsene. Ergebnisse eines Experiments in der Startkohorte 6 des Nationalen Bildungspanels (NEPS)," in *Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen*, ed. J. Schupp and C. Wolf, pp. 359–381, Wiesbaden: Springer VS.

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, 533–549.

Lathia, N., G. M. Sandstrom, C. Mascolo, and P. J. Rentfrow (2017), "Happier People Live More Active Lives: Using Smartphones to Link Happiness and Physical Activity," *PLoS One*, 12, e0160589.

Lazer, D., and J. Radford (2017), "Data Ex Machina: Introduction to Big Data," *Annual Review of Sociology*, 43, 19–39.

Lynn, P. (2014), "Longer Interviews May Not Affect Subsequent Survey Participation Propensity," *Public Opinion Quarterly*, 78, 500–509.

MacKerron, G., and S. Mourato (2013), "Happiness Is Greater in Natural Environments," *Global Environmental Change*, 23, 992–1000.

Marmot, M., and A. Steptoe (2008), "Whitehall II and ELSA: Integrating Epidemiological and Psychobiological Approaches to the Assessment of Biological Indicators," in *Biosocial Surveys*, ed. M. Weinstein, J. W. Vaupel, and K. W. Wachter, pp. 42–95, Washington, DC: National Academies Press.

Pashazadeh, F., A. Cernat, and J. Sakshaug (2021), "The Effects of Biological Data Collection in Longitudinal Surveys on Subsequent Wave Cooperation," in *Advances in Longitudinal Survey Methodology*, ed. P. Lynn, pp. 100–121, Hoboken, NJ: John Wiley & Sons.

Revilla, M., M. P. Couper, and C. Ochoa (2019), "Willingness of Online Panelists to Perform Additional Tasks," *Methods, Data, Analyses*, 13, 223–252.

Sapiezynski, P., A. Stopczynski, D. D. Lassen, and S. Lehmann (2019), "Interaction Data from the Copenhagen Networks Study," *Scientific Data*, 6, 1–10.

Sastry, N., P. Fomby, and K. McGonagle (2016), "Effects on Core Panel Attrition of Selection for a Supplemental Study: Evidence from the Panel Study of Income Dynamics," Available at: https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.535777.de/psmw2016_sastry_fomby_mcgonagle_effects%20on%20core%20panel%20attrition%20of%20selection%20for%20a%20supplemental%20study.pdf. Accessed April 28, 2022.

Silberstein, A. R., and C. A. Jacobs (1989), "Symptoms of Repeated Interview Effects in the Consumer Expenditure Interview Survey," in *Panel Surveys*, ed. D. Kasprzyk, G., Duncan, G. Kalton, and M. P. Singh, pp. 289–303, New York: Wiley.

Stier, S., J. Breuer, P. Siegers, and K. Thorson (2020), "Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field," *Social Science Computer Review*, 38, 503–516.

Struminskaya, B., P. Lugtig, V. Toepoel, B. Schouten, D. Giesen, and R. Dolmans (2021), "Sharing Data Collected with Smartphone Sensors: Willingness, Participation, and Nonparticipation Bias," *Public Opinion Quarterly*, 85, 423–462.

Trappmann, M., G. Müller, and A. Bethmann (2013), "Design of the Study," in *User Guide "Panel Study Labour Market and Social Security" (PASS): Wave 6*, ed. A. Bethmann, B. Fuchs and A. Wurdack, pp. 13–22, Nuremberg: Institute for Employment Research.

Trappmann, M., S. Bähr, J. Beste, A. Eberl, C. Frodermann, S. Gundert, S. Schwarz, N. Teichler, S. Unger, and C. Wenzig (2019), "Data Resource Profile: Panel Study Labour Market and Social Security (PASS)," *International Journal of Epidemiology*, 48, 1411–1411g.

Vaccari, C., A. Valeriani, P. Barberá, R. Bonneau, J. T. Jost, J. Nagler, and J. A. Tucker (2015), "Political Expression and Action on Social Media: Exploring the Relationship between Lower- and Higher-Threshold Political Activities among Twitter Users in Italy," *Journal of Computer-Mediated Communication*, 20, 221–239.

Wang, R., F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell (2014), "StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones," *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 3–14.

Wang, R., W. Wang, M. S. Aung, D. Ben-Zeev, R. Brian, A. T. Campbell, T. Choudhury, M. Hauser, J. Kane, E. A. Scherer, and M. Walsh (2017), "Predicting Symptom Trajectories of Schizophrenia Using Mobile Sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1, 1–24.

Weir, D. (2008), "Elastic Powers: The Integration of Biomarkers into the Health and Retirement Study," in *Biosocial Surveys*, ed. M. Weinstein, J. Vaupel, and K. Wachter, pp. 78–95, Washington, DC: National Academies Press.

White, H. L. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica: Journal of the Econometric Society*, 48, 817–838.

———. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica: Journal of the Econometric Society*, 50, 1–25.