Editorial: From web surveys to mobile web to apps, sensors, and digital traces

Bella Struminskaya, Utrecht University, The Netherlands
Florian Keusch, University of Mannheim, Germany

10.12.2020

How to cite this article: Struminskaya, B. & Keusch, F. (2020). From web surveys to mobile web to apps, sensors, and digital traces in *Survey Methods: Insights from the Field, Special issue: 'Advancements in Online and Mobile Survey Methods'.* Retrieved from https://surveyinsights.org/?p=15106.
DOI:10.13094/SMIF-2020-00015

Over the past three decades, web surveys have come a long way from being based on self-selected samples of mainly Internet-savvy volunteers to an established mainstream tool that is used in all spheres of the social and behavioral sciences. The discussion around online surveys as a research tool is no longer focused on whether it is a legitimate data collection mode or whether online surveys will make traditional modes obsolete as has been noted at the inception of web data collection methods (e.g., Couper, 2000). In the late 1990s and early 2000s, online surveys were mostly based on non-probability samples, with the use of probability-based samples restricted to populations with high coverage such as employees or students; probability-based panels of Internet users or the full population were an exception. The primary way of completing online questionnaires was on a PC.

The fragmentation of the online survey industry, first described by Couper (2000) and characterized by a wide array of approaches applied by online sample vendors and varying levels of quality at varying costs, has continued since. A decade after Couper's article was published, the proliferation of mobile communication devices, including smartphones, has given rise to the investigation of device effects on survey data quality (e.g., Peytchev & Hill, 2010).

Online panels, both the ones using probability-based methods for recruitment and those using non-probability approaches, have contributed to the rapid increase of web surveys (Callegaro et al., 2014). Both from an operations and research perspective, online panels have focused on reducing measurement error such as identifying the effects of professional respondents on data quality (e.g., Hillygus, Jackson & Joung, 2014), speeding (e.g., Greszki, Meyer & Schoen, 2014) and other forms of satisficing (e.g., Zhang & Conrad, 2016), as well as other attempts to eliminate fraudulent respondents (Baker et al., 2014). Questions about biases in non-probability online panels (Yeager et al., 2011) and techniques to correct for them by using statistical adjustment procedures, including weighting and imputation, and sample matching (Bethlehem, 2016), received increased attention.

Probability-based panels of the general population such as the Dutch LISS Panel, the German GESIS Panel and the GIP, and the French ELIPSS panel have sprung in multiple countries, providing non-Internet respondents with alternative modes or Internet access (Blom et al., 2016). Mixing modes to achieve target population coverage has been adapted in academia and has found its use in industry for some consumer panels. If the population estimates are of interest to the researchers, probability-based surveys are still the gold standard to achieve accurate estimates: Non-probability online panels show greater biases compared with probability-based panels when compared with the population benchmark data, and weighting the data from non-probability panels is shown to not be effective in the majority of cases when the goal is to achieve unbiased population statistics (Cornesse et al., 2020).

Recent research has focused on whether and how probability-based and non-probability surveys can be combined to achieve accurate population estimates (Sakshaug, Wiśniowski, Perez Ruiz & Blom, 2019), how the online mode can be integrated with other data collection modes (e.g., De Leeuw, 2005), and how online surveys can be administered on different devices with minimal measurement error (Couper, Antoun & Mavletova, 2017).

In 2020, the online mode is increasingly used in government surveys, large-scale general population panels, and establishment surveys. For some countries, such as the Netherlands, several household survey programs have been re-designed to include the online mode as part of the mode mix (Van der Laan & Van Nunspeet, 2009). The European Social Survey has launched the cross-national online CRONOS panel in three countries, and is expanding the number of participating countries presently. Most of the EU countries have the online mode as part of the mode mix for data collection used by national statistical institutes (Murgia, Lo Conte & Gravem, 2018). Furthermore, online survey data are increasingly linked to auxiliary data such as administrative records (Sakshaug & Antoni, 2017), social media data (Stier, Breuer, Siegers & Thorson, 2020), and (smartphone) sensor data (Struminskaya, Lugtig, Keusch & Höhne, forthcoming).

Online surveys are once again at the verge of a new era of survey research, that is, the integration of 'big' data into data collection. Data collection methods adapt to the communication patterns of participants and technological changes (e.g., messenger-like interviewing, see Toepoel, Lugtig, Struminskaya, Elevelt & Haan, 2020). We can distinguish several trends. First, survey questions in online surveys are increasingly being replaced by sensor measurement using smartphones and other mobile devices. Sensor measurement using smartphones is integrated into data collection by official statistics: For example, the IAB SMART study in which app-based data informs labour market research in Germany (Kreuter, Haas, Keusch, Bähr & Trappmann, 2020), the travel TABI app in the Netherlands (McCool, Schouten & Lugtig, forthcoming), and other plans to incorporate app and sensor measurement (Salemnik, Dufour & Van der Steen, 2020). Second, in-situ measurement using Ecological Momentary Assessment (EMA) or Experience Sampling Method (ESM) (Höhne, 2020) questions requesting participants to provide self-reports multiple times a day on participants' devices or loaner phones. Such measurement is close to the behavior of the participant and allows us to collect context-rich information. Third, donation of data collected on social media or from user's devices, such as smartwatches, fitness trackers, and smart household appliances. With the implementation of the European General Data Protection Regulation (GDPR) in 2018 that legally obliged companies to provide the data to their users upon request, researchers can use the data download packages (DDPs) obtained from users. For example, the Instagram DDPs provide insights into user's behaviors on Instagram and allow researchers to study social media behavior (Boeschoten, Oberski, Van Driel & Pouwels, 2020). The DDPs donated by research participants do not require programming of special apps or in-browser measurements as would smartphone sensor-based research.

As these methods move forward and smartphone use increases, concerns about the selectivity in each step of obtaining consent from the study participants, their privacy concerns, and willingness to share smartphone-sensor data gain attention in the literature (Keusch, Struminskaya, Antoun, Couper & Kreuter, 2019; Struminskaya et al. forthcoming; Wenz, Jäckle & Couper, 2019).

Biemer and Lyberg (2003) have proposed thinking about surveys as data collection systems, that is, entire data collection processes designed around a specific mode (see also Struminskaya, Kaczmirek & De Leeuw, 2015). Today, such data collection systems that include online survey components become more complex and multi-source: mixed-mode multi-device surveys with high-frequency EMS measurement and potentially always-on sensor components, with an option to link to auxiliary data. Such augmentation of surveys and replacement of questions results in complex designs with multi-stage recruitment and consent procedures poses challenges of potential selectivity, complex missing data patterns, and measurement validity.

The data quality of the components of such data collection systems can be assessed by applying the Total Survey Error framework (Groves, 1989; Biemer 2010) that has been adapted to incorporate digital traces (e.g., Sen, Floeck, Weller, Weiss & Wagner, 2019) and big data (e.g., Amaya, Biemer & Kinyon, 2020). The TSE framework and its modifications allow us to decompose the data collection processes and focus on individual error sources related to measurement and representation. The papers of this special issue provide a sneak peek into the future of online data collection. While we cannot predict future (technological) developments, we can focus on the quality of individual parts of the data collection systems as well as "stitching" these parts together to achieve high-quality data that is fit for the intended (research) purpose.

This special issue is sponsored by the German Society for Online Research (DGOF), an organization that continuously facilitates research on online surveys in academia, industry, and market research since the 1990s. The goal of this special issue is to contribute to knowledge about the current state of online and mobile survey methods, focusing on the components of the Total Survey Error framework for cross-sectional and panel surveys as well as augmenting survey data with other data types. The issue includes eight papers that cover different aspects of recent technological advancements that drive innovation and enable new measurement capabilities that potentially allow new and deeper insights into human interactions, attitudes, and behaviors.

Kaplan and Edgar (2020) deal with issues of recruitment through crowdsourcing for pretesting of web surveys. The authors use a mix of pretesting methods to gauge the impact of confidentiality pledge wordings that can be used for governmental surveys. Traditional interviewer-administered cognitive interviews are followed by online pretests in which an online crowdsourcing platform (in this case, Amazon Mechanical Turk) is used. Complementary use of interviewer-administered and online pretesting methods allowed for the initial assessment of respondents' reactions, comprehension, recollection, and potential impact of the confidentiality pledge language through open-ended probes. The small-scale study showed that respondents had few concerns with the confidentiality language. Online pretesting allowed for larger group sizes and recruitment of respondents who were unfamiliar with the study sponsor. While the findings on confidentiality claims should be interpreted with caution since they are based on the non-probability sample of MTurk respondents, this article demonstrates how to effectively combine multiple-mode questionnaire pretesting methods, design a series of pretesting studies that build upon each other, and leverage the benefits of the crowdsourced platforms for survey research.

Kühne and Zindel (2020) demonstrate how hard-to-reach and hard-to-enumerate populations can be recruited through social media. While probability-based surveys are gold-standard for population estimates, for other purposes or when the population is unknown or unreachable, non-probability methods prove to be valuable. The paper provides guidance on how to prepare and launch Facebook and Instagram ad campaigns to recruit study participants. The authors describe a web survey for which they recruited lesbian, gay, bisexual, transgender, and queer (LGBTQ) participants in Germany. They focus on preparation of the study, campaign creation, and monitoring and evaluation of the recruitment campaign, including its costs. Furthermore, they compare the sample composition of the social media-recruited sample to the sample of LGB households that was drawn from the probability-based German Socio-Economic Panel (SOEP) based on a telephone screening followed by face-to-face interviews. While the non-probabilistic social-media-recruited sample was heavily biased in age and gender, the researchers could reach a greater number of individuals from this rare target population. This paper illustrates the promise of combining data from probability and non-probability data sources, and it helps researchers to set up social media-based participant recruitment.

Dealing with issues of weighting and adjustment of web surveys, Irimata et al. (2020) study the properties of web survey estimates from a probability sample for health outcomes. They calibrate the web survey estimates based on propensity score weighting techniques using an existing national survey, and they test the influence of size and collection timeline of the reference data set on the outcomes. The adjusted health estimates vary little when using quarterly or yearly data, suggesting that there is flexibility in selecting the reference dataset. The study has a number of practical implications for constructing reference data for web surveys, including the reduced cost and burden of a smaller sample size and a more flexible timeline.

Moving to papers that deal with mobile devices in web surveys, Andreadis (2020) presents a web survey in Greece that used exclusively text messaging (SMS) for invitations and reminders. The paper examines the impact of various design study features such as pre-notifications, time and day of SMS delivery, lag between invitations, and reminders on survey response. He finds that an SMS pre-notification significantly improves the response rates to the web survey, even more so than a reminder SMS. The timing of the SMS message had no influence on response behavior. Not surprisingly, the majority of respondents used their smartphone to complete the web survey, and only few switched to a PC when receiving the invitation SMS. This finding stresses the importance of a mobile-friendly questionnaire design when inviting respondents via SMS.

Clement et al. (2020) use the Danish cross-sections of the International Social Survey Programme (ISSP) 2018 and 2019 to study device effects in web surveys optimized for mobile devices. They find no difference in self-reported survey engagement by device for respondents who self-selected the device to complete the survey (i.e., smartphone, tablet, or PC). In addition, no evidence that responding on smartphones or tablets causes lower data quality (acquiescence, nonsubstantive answers, midpoint responding, primacy effects, straightlining) than responding on PCs was found. The completion time also did not differ significantly across devices. This study demonstrates the importance of studying device effects using surveys that have different length and content to be able to validate findings from one cross-section with the data from another cross-section.

Baier and Fuchs (2020) study the prevalence of page-switching during the completion of a (mobile) web survey and the effects of page-switching on data quality. They find that the prevalence of page-switching is low; if it occurs, it is short and less likely to occur for respondents who use smartphones for survey completion compared to respondents who use PCs or tablets. There is no evidence that page-switching leads to lower quality data, judging by the absence of significant differences in item missings, non-differentiation, and the number of characters in open-ended questions. These findings should be interpreted with caution since they are based on an online access panel and two convenience samples of university applicants in Germany. However, given that respondents from non-probability online panels usually cause concern about data quality, these results are important for researchers considering data collection among non-probability samples. Moreover, since page-switching is an indicator of multitasking, researchers need to worry less about respondents of mobile web surveys being distracted during survey completion.

Two papers go beyond using smartphones for self-reporting in mobile web surveys, and they employ the sensors built into smartphones to enrich data collection. Eckman and colleagues (2020) study the feasibility of always-on geolocation data collection from smartphones. In a pilot study with 24 iPhone users who shared their geolocation data over two weeks, the authors try to determine specific locations that participants visited by comparing the smartphone geolocation coordinates with points of interest from three publicly available databases (Google Places, Yelp, Foursquare). They find both too few and too many matches between the smartphone geolocations and the databases. In addition, the agreement between the identified places and survey data vary by type of location. One location that could be identified particularly well using the always-on geolocation data was the workplace, but there was also relatively high alignment between survey data and geolocation data for daycare centers. The finding raises legal and ethical questions about privacy and (re-)identification when collecting these types of data.

Haas and colleagues (2020) describe a feasibility study that uses geofencing technology. Geofence is a geographical area that, when entered, exited, or dwelled in, triggers a survey invitation on a participant's smartphone. The authors geofenced over 400 German job centers, so that study participants who stayed over 25 minutes within the geofence received surveys about the purpose of the visit to the job center and their experiences. The study participants were Android phone owners recruited from a probability-based general population panel that collects information on labor market behavior. The paper describes study design choices, including the incentive regime, demonstrating how high response to geofenced surveys could be achieved and, more importantly, focus on design challenges that can cause measurement errors. For example, the consequences of not considering opening hours of job centers are discussed and how many surveys could be falsely triggered if geolocation is not validated.

References

1. Amaya, A., Biemer, P. P., & Kinyon, D. (2020). Total error in a big data world: Adapting the TSE framework to big data. *Journal of Survey Statistics and Methodology*, *8*(1), 89-119. https://doi.org/10.1093/jssam/smz056
2. Baker, R., Miller, C., Kachhi, D., Lange, K., Wilding-Brown, L., & Tucker, J. (2014). Validating respondents' identity in online samples. In P. J. Lavrakas, J. A. Krosnick, A. S. Göritz, R. P. Baker, J. Bethlehem, & M. Callegaro (Eds.), *Online Panel Research: A Data Quality Perspective* (1st ed., pp. 441–456). Wiley.
3. Bethlehem, J. (2016). Solving the nonresponse problem with sample matching? *Social Science Computer Review*, *34*(1), 59–77. https://doi.org/10.1177/0894439315573926
4. Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, *74*(5), 817–848. https://doi.org/10.1093/poq/nfq058
5. Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Wiley.
6. Blom, A. G., Bosnjak, M., Cornilleau, A., Cousteaux, A. S., Das, M., Douhou, S., & Krieger, U. (2016). A comparison of four probability-based online and mixed-mode panels in Europe. *Social Science Computer Review*, *34*(1), 8-25. https://doi.org/10.1177/0894439315574825
7. Boeschoten, L., Oberski, D. L., Van Driel, I. I., Pouwels, L. (2020). Instagram Use and the Well-Being of Adolescents: Using Deep Learning to Link Social Scientific Self-reports with Instagram Data Download Packages. Paper presented at ICMI '20 Companion, October 25–29, 2020, Virtual event, The Netherlands. https://doi.org/10.1145/3395035.3425185
8. Callegaro, M., Baker, R., Bethlehem, J., Göritz, A. S., Krosnick, J. A., and Lavrakas, P. J. (2014). Online panel research: History, concepts, applications, and a look at the future. In P. J. Lavrakas, J. A. Krosnick, A. S. Göritz, R. P. Baker, J. Bethlehem, & M. Callegaro (Eds.), *Online Panel Research: A Data Quality Perspective* (1st ed., pp. 1–22). Wiley.
9. Cornesse, C., Blom, A. G., Dutwin, D., Krosnick, J. A., De Leeuw, E. D., Legleye, S., Pasek,

J., Pennay, D., Phillips, B., Sakshaug, J. W., Struminskaya, B., & Wenz, A. (2020). A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research. *Journal of Survey Statistics and Methodology*, *8*(1), 4–36. https://doi.org/10.1093/jssam/smz041
10. Couper, M. P. (2000). Web surveys: A review of issues and approaches. *The Public Opinion Quarterly*, *64*(4), 464-494. https://doi.org/10.1086/318641
11. Couper, M. P., Antoun, C., & Mavletova, A. (2017). Mobile web surveys: A total survey errorperspective. In Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., Tucker, N. C., West, B. T., de Leeuw, E. D., Leeuw, E. D., & Biemer, P. P. (Eds), *Total Survey Error in Practice: Improving Quality in the Era of Bid Data.* (1st ed, pp. 133–154). Wiley.
12. De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of official statistics*, *21*(5), 233-255.
13. Greszki, R., Meyer, M., & Schoen, H. (2014). The impact of speeding on data quality in nonprobability and freshly recruited probability-based online panels. In P. J. Lavrakas, J. A. Krosnick, A. S. Göritz, R. P. Baker, J. Bethlehem, & M. Callegaro (Eds.), *Online Panel Research: A Data Quality Perspective* (1st ed., pp. 238–262). Wiley.
14. Groves, R. M. (1989). *Survey costs and survey errors.* Wiley.
15. Hillygus, D. S., Jackson, N., & Young, M. (2014). Professional respondents in non-probability online panels. In P. J. Lavrakas, J. A. Krosnick, A. S. Göritz, R. P. Baker, J. Bethlehem, & M. Callegaro (Eds.), *Online Panel Research: A Data Quality Perspective* (1st ed., pp. 219-237). Wiley.
16. Höhne, J. (2020). Experience sampling. In P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, & R.A. Williams (Eds.), *SAGE Research Methods Foundations.* doi: 10.4135/9781526421036910198
17. Keusch, F., Struminskaya, B., Antoun, C., Couper, M., Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly* 83(S1), 210-235. https://doi.org/10.1093/poq/nfz007
18. Kreuter, F., Haas, G. C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*, *38*(5), 533-549. https://doi.org/10.1177/0894439318816389
19. McCool, D., Schouten, J.G. & Lugtig, P. (forthcoming). An app-assisted travel survey in official statistics. Possibilities and challenges. *Journal of Official Statistics*
20. Murgia, M., Lo Conte, M., Gravem, D. F. (2018). Report on MIMOD survey on the state of the art of mixed mode for EU social surveys. Work Package 1 “Investigation of mode organization” Deliverable 1. Cooperation on Multi-Mode Data Collection (MMDC) Mixed Mode Designs for Social Surveys – MIMOD. Istat. https://www.istat.it/it/files//2011/07/WP1-deliverables.zip
21. Peytchev, A., & Hill, C. A. (2010). Experiments in mobile web survey design: Similarities to other modes and unique considerations. *Social Science Computer Review, 28(3),* 319-335. https://doi.org/10.1177/0894439309353037
22. Sakshaug, J. W., & Antoni, M. (2017). Errors in linking survey and administrative data. In Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., Tucker, N. C., West, B. T., de Leeuw, E. D., Leeuw, E. D., & Biemer, P. P. (Eds), *Total Survey Error in Practice: Improving Quality in the Era of Bid Data.* (1st ed, pp. 133–154). Wiley.
23. Sakshaug, J. W., Wiśniowski, A., Ruiz, D. A. P., & Blom, A. G. (2019). Supplementing small probability samples with nonprobability samples: A Bayesian approach. *Journal of Official Statistics*, *35*(3), 653-681. https://doi.org/10.2478/jos-2019-0027
24. Salemink, I., Dufour, S., & Van der Steen, M. (2020). A vision on future advanced data collection. *Statistical Journal of the IAOS*, 3, 685-699. https://ddoi.org/10.3233/SJI-200658
25. Sen, I., Floeck, F., Weller, K., Weiss, B., & Wagner, C. (2019). A total error framework for digital traces of humans. *arXiv preprint arXiv:1907.08228.*
26. Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: key issues in developing an emerging field. *Social Science Computer Review, 38*(5), 503-516. https://doi.org/10.1177/0894439319843669
27. Struminskaya, B., de Leeuw, E. D., & Kaczmirek, L. (2015). Mode system effects in an online panel study: Comparing a probability-based online panel with two face-to-face reference surveys. *Methods, Data, Analyses*, *9*(1), 3-56. https://doi.org/10.12758/mda.2015.001
28. Struminskaya, B., Lugtig, P., Keusch, F., & Höhne, J. K. (forthcoming). Augmenting surveys with data from sensors and apps: Opportunities and challenges. Editorial for the Special Issue of Social *Science Computer Review* “Using Mobile Apps and Sensors in Surveys”

29. Struminskaya, B., Toepoel, V., Lugtig, P., Haan, M., Luiten, A., & Schouten, B. (forthcoming). Understanding willingness to share smartphone-sensor data. *Public Opinion Quarterly.*
30. Toepoel, V., Lugtig, P., Struminskaya, B., Elevelt, A., & Haan, M. (2020). Adapting Surveys to the Modern World: Comparing a Research Messenger Design to a Regular Responsive Design for Online Surveys. *Survey Practice* 13 (1). https://doi.org/10.29115/SP-2020-0010
31. Van der Laan, P., & Van Nunspeet, W. (2009). Modernising household surveys in the Netherlands: Design, Efficiency Gains and Perspectives. *Statistics Netherlands*.
32. Wenz, A., Jäckle, A., & Couper, M. P. (2019). Willingness to use mobile technologies for data collection in a probability household panel. *Survey Research Methods*, *13*(1), 1-22. https://doi.org/10.18148/srm/2019.v1i1.7298
33. Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public opinion quarterly*, *75*(4), 709-747. https://doi.org/10.1093/poq/nfr020
34. Zhang, C., & Conrad, F. G. (2018). Intervening to reduce satisficing behaviors in web surveys: Evidence from two experiments on how it works. *Social Science Computer Review*, *36*(1), 57-81. https://doi.org/10.1177/0894439316683923