10th GESIS Summer School in Survey Methodology

[2nd Virtual GESIS Summer School]

28 July – 20 August 2021

Syllabus for Course 7: (Non-)Probability Samples in the Social Sciences

Lecturer: Dr. Carina Cornesse
E-mail: carina.cornesse@uni-mannheim.de
Homepage: https://www.sowi.uni-mannheim.de/en/blom/team/carina-cornesse/

Date: 09-13 August 2021
Time: 09:00-12:00 + 13:30-16:30
Time zone: CEST/CEDT, course starts Monday at 09:30 am
Venue: Online via Zoom

About the Instructor:

Carina Cornesse is a post-doctoral researcher in survey methodology at the University of Mannheim and co-director of the project "Recruiting Probability-Based Online Panels: Cost-Efficiency and Data Quality" at the Mannheim Centre for European Social Research (MZES). She also serves as an associate editor of the Journal of Survey Statistics and Methodology (JSSAM). Her research focuses on social science research design topics, in particular the collection, evaluation, and enhancement of innovative high-quality survey data. She is particularly interested in probability and nonprobability samples, online panel recruitment and maintenance, and augmenting survey data using other data sources.

Selected Publications:

- Cornesse, Carina, Annelies G. Blom, David Dutwin, Jon A. Krosnick, Edith D. de Leeuw, Stéphane Legleye, Josh Pasek, Darren Pennay, Benjamin Philips, Joseph W. Sakshaug, Bella Struminskaya, Alexander Wenz (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.
- Cornesse, Carina, Annelies Blom (2020). Response Quality in Nonprobability and Probability-based Online Panels. Sociological Methods & Research. doi:10.1177/0049124120914940.
- Cornesse, Carina, Ines Schaurer (2021). The Long-Term Impact of Different Offline Population Inclusion Strategies in Probability-Based Online Panels: Evidence from the German Internet Panel and the GESIS Panel. Social Science Computer Review. doi:10.1177/0894439320984131.

Short Course Description:

The main objective of the course is to provide students with a full overview of the history, theoretical foundations, critical arguments, and accumulated empirical evidence surrounding the debate about probability and nonprobability sample surveys. A focus will be on real-world examples of why and how the choice of sample type matters, including topics such as gender data gap, election polling debacles, and the role that surveys can have in supporting versus debunking fake news. In addition to discussing these topics, the course will provide students with an in-depth understanding of the conditions under which probability and nonprobability samples can provide useful data to answer social scientific research questions (e.g., so-called "fit-for-purpose" designs), including hands-on recommendations and exercises on how to design your own (hypothetical) research study.

Keywords:

Research design, sample selection, fit-for-purpose designs, probability samples, nonprobability samples

Course Prerequisites:

- Basic knowledge of introductory statistics (e.g., descriptive statistics, basic regression analysis)
- Basic conceptual understanding of survey data collection (e.g., survey lifecycle, Total Survey Error framework)
- Basic understanding of sampling theory and/or survey weighting procedures is desirable, but not strictly necessary.

Target Group:

Participants will find the course useful if:

- they would like to get a full picture of the debate about probability and nonprobability sample surveys.
- they plan to design their own research study and need to choose a sample type.

Course and Learning Objectives:

By the end of the course participants will:

- Have a full overview of the history, theoretical foundations, critical arguments, and accumulated empirical evidence surrounding the debate about probability and nonprobability sample surveys.
- Possess the necessary skills to evaluate whether any given sample is fit for the purpose of answering a particular research question.
- Be able to choose an appropriate sample type when designing their own social scientific research studies.

Organizational Structure of the Course:

This is a five-day course with a total amount of 30 hours of virtual class time. Participants can expect a mix of interactive teaching, group exercises, and opportunity for individual consultation. Exercises
are designed to deepen the understanding of the course material and to apply it to hypothetical and real-world research settings. This will not require using any statistical software (although participants will have the opportunity to work with data voluntarily using their preferred software). To get the most out of the course, students are strongly encouraged to participate pro-actively in small-group projects, present their thoughts and work output in class, and critically discuss the course material with the lecturer and classmates.

Software and Hardware Requirements:

None.

Long Course Description:

In the social sciences, probability sample surveys have been the gold-standard approach to studying the general population for many decades. Since the rise of the internet as a mass medium in the late 20th century, however, more and more researchers use nonprobability internet samples for such research purposes. This has, once again, sparked a decades-old debate about whether and when it is appropriate to use nonprobability samples for social science research and when it may be necessary to rely on probability samples instead. Fuelled by the news media, this debate tends to get particular heated in the weeks and months following an important political election, such as the Brexit referendum in 2016 and, recently, the US presidential election in 2020. However, the debate about probability and nonprobability samples is multi-faceted and extends far beyond the area of election polling. In our digital age, probability and nonprobability samples are everywhere and concern many areas of our lives, including all social and epidemiological aspects surrounding the COVID-19 pandemic, via our use of voice-recognition technology in "smart" devices, to our likelihood of surviving a heart attack. When the basic assumptions underlying the use of probability and nonprobability samples for general population research are violated, this can have harmful consequences. These may include the implementation of the wrong policy measures, diminishing trust in our democracy, our exclusion from promising new technologies, and even avoidable deaths.

In this course, we shed a light on the past, present, and future of probability and nonprobability samples with a particular focus on practical applications in the social sciences. We critically discuss the theoretical foundations, empirical evidence, and practical relevance of probability and nonprobability samples. By the end of the course, participants will feel confident to participate in the debate about probability and nonprobability samples, will be empowered to design their own studies following best-practice recommendations, assess offers from survey companies, and evaluate new research studies.

The course will include a number of individual and small group excercises to help participants to get into the habit of thinking about how to design studies that fit different research goals, assess studies that have a particular research design, evaluate survey products offered by data providers on the survey market, look for the relevant study design information in research publications, and make sure to include all relevant information in their own publications. The output from these excercises will be presented and discussed by the course participants in class to accelerate, expand, and deepen the learning progress.

Day-to-day Schedule and Literature:

Day	Topic(s)
1	**Introduction: (Non-)Probability Samples in the Social Sciences** ▪ The Debate ▪ The (Sampling) Theory ▪ The (Accumulated) Evidence
	<u>Compulsory reading (have to be read before the session):</u> ▪ Cornesse, C., Blom, A. G., Dutwin, D., Krosnick, J. A., De Leeuw, E. D., Legleye, S., Pasek, J.,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. ▪ Kohler, U. (2019). Possible Uses of Nonprobability Sampling for the Social Sciences. Survey Methods: Insights from the Field. https://surveyinsights.org/?p=10981. <u>Suggested reading (suggested, yet do not have to be read before the session):</u> ▪ Dutwin, D., Buskirk, T.D. (2017). Apples to Oranges or Gala Versus Golden Delicious? Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples. Public Opinion Quarterly 81(S1), pp.213-239. ▪ MacInnis, B., Krosnick, J. A., Ho, A. S., Cho, M. J. (2018). The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension. Public Opinion Quarterly, 82(4), 707-744. ▪ Mercer, A. W., Kreuter, F., Keeter, S., & Stuart, E. A. (2017). Theory and Practice in Nonprobability Surveys: Parallels Between Causal Inference and Survey Inference. Public Opinion Quarterly, 81(S1), 250-271. ▪ 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.
2	**Election (Mis-)Predictions and (Non-)Probability Samples.** ▪ The Problem ▪ The Role of the Media ▪ The (Suggested) Solutions ▪ The Evidence ▪ Survey Weighting Approaches
	<u>Compulsory reading:</u> ▪ Sturgis, P., Kuha, J., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., Lauderdale, B.E. & Smith, P. (2018). An Assessment of the Causes of the Errors in the 2015 UK General Election Opinion Polls. Journal of the Royal Statistical Society: Series A (Statistics in Society), 181(3), 757-781. ▪ Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting Elections with Non-Representative Polls. International Journal of Forecasting, 31(3), 980-991. <u>Suggested reading:</u> ▪ Blumenthal, Mark, S. Clement, J. D. Clinton, C. Durand, C. Franklin, L. Miringoff, K. Olson, D. Rivers, Y. L. Saad, and G. E. Witt. (2017). An Evaluation of 2016 Election Polls in the US. Available from https://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx#POLLING%20AND%20PROBABILISTIC%20FORECASTING.

	▪ Enns, P. K., Rothschild J. (2021). Revisiting the 'Gold Standard' of Polling: New Methods Outperformed Traditional Ones in 2020. Retrieved from https://medium.com/3streams/revisiting-the-gold-standard-of-polling-new-methods-outperformed-traditional-ones-in-2020-451650a9ba5b ▪ Pennay, D., Goot, M. Neiger, D. Trewin, D. Lavrakas, P. J., Stirton, J., Hughes, P. Sheppard, J. McAllister, I. (2020). Report of the Inquiry into the Performance of the Opinion Polls at the 2019 Australian Federal Election, Association of Market and Social Research Organisations and the Statistical Society of Australia. Available from https://www.amsro.com.au/amsro-polling-inquiry-final-report/.
3	**Misinformation and (Non-)Probability Samples**. ▪ The Role of Surveys Regarding "Fake News" ▪ The Online Panel Marketing Infrastructure ▪ "Professional Respondents" ▪ The Need for Transparency ▪ Excursion: Nonprobability Samples and the Replication Crisis
	<u>Compulsory reading:</u> ▪ Damian, E., Meuleman, B., van Oorschot, W. (2019). Transparency and Replication in Cross-National Survey Research: Identification of Problems and Possible Solutions. Sociological Methods & Research, https://doi.org/10.1177/0049124119882452. ▪ Guess, A., Nyhan, B., & Reifler, J. (2018). Selective Exposure to Misinformation: Evidence from the Consumption of Fake News During the 2016 US Presidential Campaign. European Research Council. Retrieved from http://www.ask-force.org/web/Fundamentalists/Guess-Selective-Exposure-to-Misinformation-Evidence-Presidential-Campaign-2018.pdf. ▪ Zhang, C., Antoun, C., Yan, H. Y., & Conrad, F. G. (2020). Professional Respondents in Opt-in Online Panels: What Do We Really Know?. Social Science Computer Review, 38(6), 703-719. <u>Suggested reading:</u> ▪ Cornesse, C., Blom A. (2020). Response Quality in Nonprobability and Probability-Based Online Panels. Sociological Methods & Research. doi:10.1177/0049124120914940. ▪ Cornesse, C., Bosnjak, M. (2018). Is There an Association Between Survey Characteristics and Representativeness? A Meta-Analysis. Survey Research Methods 12(1), 1-13. ▪ Le Texier, T. (2019). Debunking the Stanford Prison Experiment. *American Psychologist, 74*(7), 823–839. ▪ Matthijsse, S. M., De Leeuw, E. D., & Hox, J. J. (2015). Internet Panels, Professional Respondents, and Data Quality. Methodology, 11, 81-88.
4	**The COVID-19 Pandemic and (Non-)Probability Samples** ▪ The Problem(s) ▪ The (Suggested) Solutions ▪ The Evidence ▪ Excursion: Survey Process Automation
	<u>Compulsory reading:</u> ▪ Lewis, Neil Jr. (2020) Why Coming Up With Effective Interventions To Address COVID-19 Is So Hard. Available from https://fivethirtyeight.com/features/why-coming-up-witheffective-interventions-to-address-covid-19-is-so-hard/. ▪ Schnell, R., & Smid, M. (2020). Methodological Problems and Solutions in Sampling for Epidemiological COVID-19 Research. Survey Research Methods, 14(2), 123-129. <u>Suggested reading:</u> ▪ Blom, A.G., Cornesse, C., Friedel, S., Krieger, U., Fikel, M., Rettig, T., Wenz, A., Juhl, S., Lehrer, R., Möhring, K. and Naumann, E., (2020). High-Frequency and High-Quality Survey Data Collection. Survey Research Methods, 14(2), 171-178. ▪ Klingwort, J., & Schnell, R. (2020). Critical Limitations of Digital Epidemiology. Survey Research Methods, 14(2), 95-101. ▪ Kreuter, F., Barkay, N., Bilinski, A., Bradford, A., Chiu, S., Eliat, R., Fan, J., Galili, T., Haimovich, D., Kim, B., LaRocca, S., Li, Y., Morris, K., Presser, S., Sarig, T., Salomon, J. A., Stewart, K., Stuart, E. A., & Tibshirani, R. (2020). Partnering with a Global Platform to Inform Research and Public Policy Making. Survey Research Methods, 14(2), 159-163.

5	**The Gender Data Gap and (Non-)Probability Samples** ▪ The Problem ▪ Excursion: How Algorithms Can Reproduce and Amplify Social Biases ▪ The (Suggested) Solutions ▪ The Evidence
	Compulsory reading: ▪ Perez, C. C. (2020). We Need to Close the Gender Data Gap By Including Women in Our Algorithms. https://time.com/collection-post/5764698/gender-data-gap/. ▪ Buvinic, M., & Levine, R. (2016). Closing the Gender Data Gap. *Significanc*e, 13(2), 34-37. Suggested reading: ▪ Holdcroft, A. (2007) Gender Bias in Research: How Does it Affect Evidence-Based Medicine? Journal of the Royal Society of Medicine, 100(1), 2-3. ▪ Vitale, C., Fini, M., Spoletini, I., Lainscak, M., Seferovic, P., & Rosano, G. M. (2017). Under-representation of Elderly and Women in Clinical Trials. International journal of cardiology, 232, 216-221.

Preparatory Reading:

No preparatory reading required. However, students are encouraged to start reading the required (and possibly also the suggested) reading for course session one early.

Additional Recommended Literature:

- Lohr, S. L., 2009: Sampling: Design and Analysis. 2nd edition. Cengage.
- Perez, C. C. (2019). Invisible Women: Exposing Data Bias in a World Designed for Men. Random House.
- Salganik, M. J. (2019). Bit by bit: Social Research in the Digital Age. Princeton University Press.
- Stone, D. (2020). Counting: How We Use Numbers to Decide What Matters. Liveright Publishing.