

GESIS Summer School in Survey Methodology 2022

Syllabus for course: “(Non-)Probability Samples in the Social Sciences

Lecturers: Dr. Carina Cornesse Dr. Olga Maslovskaya
E-mail: CCornesse@diw.de om206@soton.ac.uk
Homepage: https://www.diw.de/sixcms/detail.php?id=diw_01.c.832939.en <https://www.southampton.ac.uk/people/5wzzpw/doctor-olga-maslovskaya>

Date: 08-12 August 2022

Time: 10:00-13:00 | 14:00-17:00

Venue: Online via Zoom

About the Lecturers:

Dr. Carina Cornesse is a Senior Researcher at the German Socio-Economic Panel (SOEP) and the Principal Investigator (PI) of the German Social Cohesion Panel (SCP) at the German Institute for Economic Research (DIW Berlin). She is also the co-PI of the research project “Recruiting Probability-Based Online Panels: Cost-Efficiency and Data Quality” at the Mannheim Centre for European Social Research at the University of Mannheim, a research associate at the Data Centre Social Cohesion of the Research Institute Social Cohesion located at the University of Bremen, and an associate editor at the Journal of Survey Statistics and Methodology (JSSAM). Carina is a survey methods researcher and her research focuses on the benefits and limitations of probability-based and nonprobability survey samples, the recruitment and maintenance of panel studies, the application of mixed-mode data collection designs, and on linking survey data to data from other sources.

Dr. Olga Maslovskaya is an Associate Professor in Survey Research and Social Statistics at the University of Southampton, UK. She is a Survey Methodologist and is interested in all aspects of surveys, her special interests are the areas of survey data collection, data quality, and (non)-probability surveys. Olga is a Principal Investigator (PI) and the lead of the Survey Data Collection Network (SDC-Net) in the UK which investigates impact of Covid-19 pandemic on survey data collection. She was previously a PI and the lead of GenPopWeb2 international network of academic and non-academic partners. GenPopWeb2 network addressed issues associated with transitioning to online data collection. Olga is also a Co-Investigator on ESRC-funded project “The Generations and Gender Survey (GGS) in the UK: Investigating demographic changes in the family and advancing online survey methodology“. She leads the workpackage which is responsible for the design of the survey and also for methodological experiments which will be implemented in the first wave of the survey in the UK. Olga has an extensive experience of analysing wide range of large-scale datasets and of employing various advanced statistical methods in different research contexts.

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, <https://doi.org/10.1093/jssam/smz041>.

- Cornesse, Carina, Annelies Blom (2020). Response Quality in Nonprobability and Probability-based Online Panels. *Sociological Methods & Research*. doi:10.1177/0049124120914940.
- Cornesse, C., Krieger, U., Sohnius, M. L., Fikel, M., Friedel, S., Rettig, T., ... & Blom, A. G. (2021). From German Internet Panel to Mannheim Corona Study: Adaptable probability-based online panel infrastructures during the pandemic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. <https://doi.org/10.1111/rssa.12749>.
- Maslovskaya, O., & Lugtig, P. (2022). [Representativeness in six waves of CROss-National online survey \(CRONOS\) panel](https://doi.org/10.1111/rssa.12801). *Journal of the Royal Statistical Society. Series A (General)*, 183(3): 851-871. <https://doi.org/10.1111/rssa.12801>
- Sturgis, P., Maslovskaya, O., Durrant, G., & Brunton-Smith, I. (2021). [The interviewer contribution to variability in response times in face-to-face interview surveys](https://doi.org/10.1093/jssam/smaa009). *Journal of Survey Statistics and Methodology*, 9(4), 701-721. <https://doi.org/10.1093/jssam/smaa009>
- Maslovskaya, O., Durrant, G., Smith, P. W. F., Hanson, T., & Villar, A. (2019). [What are the characteristics of respondents using different devices in mixed-device online surveys? Evidence from six UK surveys](https://doi.org/10.1111/insr.12311). *International Statistical Review*, 326-346. <https://doi.org/10.1111/insr.12311>

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 election polling debacles, mispredictions during the pandemic (e.g., willingness to get vaccinated), 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., Total Survey Error framework adaptations, fit-for-purpose designs, causal inference logic), including hands-on recommendations and exercises on how to design your own (hypothetical) research study.

Furthermore, we will discuss the benefits and challenges of different approaches to probability and nonprobability sampling (e.g., simple vs. stratified random sampling, snowball vs. respondent-driven sampling, social media and river sampling) as well as recent insights on sample weighting and data integration techniques (e.g., propensity score weighting and blended calibration). All material and discussions will be hands-on and intuitive.

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.

Day-to-day Schedule and Literature:

Day	Topic(s)
1	Introduction: (Non-)Probability Samples in the Social Sciences <ul style="list-style-type: none"> - The Debate - Accumulated Evidence - Common Bias Measures - Common Sampling Approaches
	<u>Compulsory reading (have to be read before class):</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. <i>Journal of Survey Statistics and Methodology</i> , 8(1), 4-36.

	<p><u>Suggested reading (suggested, yet do not have to be read before class):</u></p> <p>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. <i>Public Opinion Quarterly</i> 81(S1), pp.213-239.</p> <p>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. <i>Public Opinion Quarterly</i>, 82(4), 707-744.</p> <p>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. <i>Public opinion quarterly</i>, 75(4), 709-747.</p>
2	<p>Theory & Practice</p> <ul style="list-style-type: none"> - (Adaptations of) The Total Survey Error Framework - “Fitness-for-purpose” & Causal Inference Thinking - Assumptions Underlying Inference from Nonprobability Samples - “Professional Respondents”
	<p><u>Compulsory reading:</u></p> <p>Dever, J. A., Amaya, A., Srivastav, A., Lu, P. J., Roycroft, J., Stanley, M., ... & Williams, W. W. (2020). Fit For Purpose in Action: Design, Implementation, and Evaluation of the National Internet Flu Survey. <i>Journal of Survey Statistics and Methodology</i>, https://doi.org/10.1093/jssam/smz050.</p> <p>Mercer, A. W., Kreuter, F., Keeter, S., & Stuart, E. A. (2017). Theory and Practice in Nonprobability Surveys: Parallels Between Causal Inference and Survey Inference. <i>Public Opinion Quarterly</i>, 81(S1), 250-271.</p> <p><u>Suggested reading:</u></p> <p>Cornesse, C., Bosnjak, M. (2018). Is There an Association Between Survey Characteristics and Representativeness? A Meta-Analysis. <i>Survey Research Methods</i> 12(1), 1-13.</p> <p>Elliott M. R., Valliant R. (2017), Inference for Nonprobability Samples, <i>Statistical Science</i>, 32, 249–264.</p> <p>Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. <i>Public opinion quarterly</i>, 74(5), 849-879.</p> <p>Kohler, U. (2019). Possible Uses of Nonprobability Sampling for the Social Sciences. <i>Survey Methods: Insights from the Field</i>. https://surveyinsights.org/?p=10981.</p> <p>Matthijsse, S. M., De Leeuw, E. D., & Hox, J. J. (2015). Internet Panels, Professional Respondents, and Data Quality. <i>Methodology</i>, 11, 81-88.</p> <p>Unangst et al. (2020). A process for decomposing total survey error in probability and nonprobability surveys: A case study comparing health statistics in US Internet panels. <i>Journal of Survey Statistics and Methodology</i>, 8(1), 62-88.</p>

	Zhang, C., Antoun, C., Yan, H. Y., & Conrad, F. G. (2020). Professional Respondents in Opt-in Online Panels: What Do We Really Know?. <i>Social Science Computer Review</i> , 38(6), 703-719.
3	<p>Election (Mis-)Predictions & (Non-)Probability Samples.</p> <ul style="list-style-type: none"> - The Problem & Role of the Media - (Suggested) Solutions & Empirical Evidence - Online Panels - Survey Weighting Approaches
	<p><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. <i>Journal of the Royal Statistical Society: Series A (Statistics in Society)</i>, 181(3), 757-781.</p> <p>Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting Elections with Non-Representative Polls. <i>International Journal of Forecasting</i>, 31(3), 980-991.</p> <p><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.</p> <p>Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. <i>Journal of official statistics</i>, 22(2), 329-349.</p> <p>Mercer, A., Lau, A., Kennedy, C. (2018). For Weighting Online Opt-In Samples, What Matters Most? https://www.pewresearch.org/methods/2018/01/26/how-different-weighting-methods-work/</p> <p>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/.</p> <p>Valliant, R. (2020). Comparing alternatives for estimation from nonprobability samples. <i>Journal of Survey Statistics and Methodology</i>, 8(2), 231-263.</p>
4	<p>The COVID-19 Pandemic & (Non-)Probability Samples</p> <ul style="list-style-type: none"> - The Problem - Suggested Solutions - Empirical Evidence - Data Integration Approaches
	<u>Compulsory reading:</u>

	<p>Beaumont, J. F., & Rao, J. N. K. (2021). Pitfalls of making inferences from non-probability samples: Can data integration through probability samples provide remedies? <i>Surv. Stat</i>, 83, 11-22.</p> <p>Bradley, V. C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X. L., & Flaxman, S. (2021). Unrepresentative big surveys significantly overestimated US vaccine uptake. <i>Nature</i>, 600(7890), 695-700.</p> <p><u>Suggested reading:</u> Burton, J., Lynn, P., & Benzeval, M. (2020). How Understanding Society: The UK Household Longitudinal Study adapted to the COVID-19 pandemic. <i>Survey Research Methods</i>, 14(2), 235-239.</p> <p>Cornesse, C., Krieger, U., Sohnius, M. L., Fikel, M., Friedel, S., Rettig, T., ... & Blom, A. G. (2021). From German Internet Panel to Mannheim Corona Study: Adaptable probability-based online panel infrastructures during the pandemic. <i>Journal of the Royal Statistical Society: Series A (Statistics in Society)</i>.</p> <p>Fahimi M., Barlas F. M., Thomas R. K., Buttermore N. (2015). Scientific Surveys Based on Incomplete Sampling Frames and High Rates of Nonresponse, <i>Survey Practice</i>, 8, 1–15.</p> <p>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. <i>Survey Research Methods</i>, 14(2), 159-163.</p> <p>Rinken, S., Domínguez-Álvarez, J. A., Trujillo, M., Lafuente, R., Sotomayor, R., & Serrano-del-Rosal, R. (2020, June). Combined mobile-phone and social-media sampling for web survey on social effects of COVID-19 in Spain. <i>Survey Research Methods</i> (Vol. 14, No. 2, pp. 165-170).</p> <p>Schnell, R., & Smid, M. (2020). Methodological Problems and Solutions in Sampling for Epidemiological COVID-19 Research. <i>Survey Research Methods</i>, 14(2), 123-129.</p> <p>Wiśniowski, A., Sakshaug, J. W., Perez Ruiz, D. A., & Blom, A. G. (2020). Integrating probability and nonprobability samples for survey inference. <i>Journal of Survey Statistics and Methodology</i>, 8(1), 120-147.</p>
5	<p>Conclusion & Future Research</p> <ul style="list-style-type: none"> - The State of the Art & Recent Developments - Beyond Social Surveys - Transparency and Ethics <p><u>Suggested reading:</u> Amaya, A., Biemer, P. P., & Kinyon, D. (2020). Total error in a big data world: Adapting the TSE framework to big data. <i>Journal of Survey Statistics and Methodology</i>, 8(1), 89-119.</p>

	<p>Bajorek, J. P. (2019). <i>Voice recognition still has significant race and gender biases</i>. Harvard Business Review. https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases.</p> <p>Carnahan, T., & McFarland, S. (2007). Revisiting the Stanford Prison Experiment: Could participant self-selection have led to the cruelty? <i>Personality and Social Psychology Bulletin</i>, 33(5), 603–614. https://doi.org/10.1177/0146167206292689</p> <p>Damian, E., Meuleman, B., van Oorschot, W. (2019). Transparency and Replication in Cross-National Survey Research: Identification of Problems and Possible Solutions. <i>Sociological Methods & Research</i>, https://doi.org/10.1177/0049124119882452.</p> <p>Gutierrez, M. (2021). Algorithmic gender bias and audiovisual data: A research agenda. <i>International Journal of Communication</i>, 15, 439-461.</p> <p>Vitale, C., Fini, M., Spoletini, I., Lainscak, M., Seferovic, P., & Rosano, G. M. (2017). Underrepresentation of Elderly and Women in Clinical Trials. <i>International journal of cardiology</i>, 232, 216-221.</p>
--	---

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:

Heeringa, S. G., West, B. T., & Berglund, P. A. (2017). *Applied survey data analysis*. CRC press.

Lohr, S. L., 2009: *Sampling: Design and Analysis*. 2nd edition. Cengage.

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples* (Vol. 1). New York: Springer.