

Comparative Analyses Using Multi-Source Databases: Insights from the Survey Data Recycling (SDR) Project

May 9-12, 2023

The international event *Comparative Analyses Using Multi-Source Databases: Insights from the Survey Data Recycling (SDR) Project* comprises a 1-day interdisciplinary conference on comparative survey research and ex-post survey data harmonization, followed by a 3-day workshop on statistical methods for analyzing multi-level harmonized datasets, the SDR dataset 2.0 (SDR 2) in particular. SDR 2 covers 4,402,489 respondents surveyed from 1966-2017 in 156 countries. It contains individual-level measures of socio-demographics, political attitudes and behaviors, social capital, and wellbeing, constructed via ex-post harmonization of social survey data pooled from 3,329 national surveys stemming from 23 major cross-national survey projects, including the World Values Survey, the European Social Survey, and the International Social Survey Programme, among others. SDR 2 also contains source survey quality and harmonization process metadata that we stored as control variables in the database and that are available for analyses.

Both the conference and the workshop are held at the OSU Department of Sociology, Townshend Hall, 1885 Neil Avenue, Columbus, OH, 43210. This event is a product of the SDR project (NSF 1738502), conducted in collaboration by the OSU departments of Sociology and Computer Science, the Mershon Center, and the Institute of Philosophy and Sociology at the Polish Academy of Sciences.

Conference Program

Tuesday, May 9 – Room 248, Townshend Hall

10:00 - 10:15 Welcome & Opening of the event by Kristi Williams, Chair of the OSU Department of Sociology, and Irina Tomescu-Dubrow, Kazimierz M. Slomczynski and J Craig Jenkins, PIs of the SDR Project

10:15 - 11:30 Christof Wolf, GESIS

Keynote lecture: The transformation of the GSS-model: self-completion, split questionnaire design and challenges for harmonization

Abstract

Most high-class social surveys, such as the General Social Survey in the US, collect data for a broad set of topics with questionnaires that are comparatively long and use face-to-face interviewing to collect their data. For a long time this mode of data collection has been considered the gold standard in survey research producing data of the highest quality. Two trends, partly related, have put the GSS-model under pressure: declining response rates and increasing cost. In this situation survey methodologists and practitioners have looked for alternatives. With advances in Web surveys and explorations of sampling strategies for self-completion surveys, a new model of conducting social science surveys has begun to emerge. As in many other areas the Covid-

pandemic has accelerated this development. Today many social surveys are either seriously considering or have already decided to switch to a self-completion mode with either only web interviews or a mix of web and postal interviews. The talk will explore several pertinent questions arising from such a shift in data collection, among others, questions of data quality, questionnaire length, the inclusion of panel elements or split questionnaire designs and if necessary, the imputation of missing values. Answers to these questions will be derived from a large experiment conducted in the context of the last wave of the European Values Study in Germany.

11:30 - 11:45 *Coffee break*

11:45 – 13:00 Session 1

J. Craig Jenkins, OSU; Olga Li, Graduate School for Social Research, GSSR; Przemek Powalko, IFiS; Irina Tomescu-Dubrow GSSR, IFiS & OSU; Kazimierz M. Slomczynski, IFiS & OSU; Yamei Tu, OSU; Ilona Wysmulek, IFiS.

This session summarizes the work conducted in the Survey Data Recycling (SDR) project (NSF 1738502, wp.asc.ohio-state.edu/dataharmonization). We outline the harmonization methodology and workflow developed to build the SDR 2 database for comparative analyses of democracy and political participation, social capital and wellbeing. We discuss main challenges that large-scale ex-post (retrospective) harmonization of cross-national survey data posed and solutions adopted in SDR. We conclude with an overview of SDR products, including the SDR data, the SDR Portal, and the Cotton File, which open our work to the wider community of researchers.

13:00 – 14:00 *Lunch break*

14:00 – 15:30 Session 2

Jennifer Oser, Ben-Gurion University of the Negev

Investigating Participation and Representation in the Digital Age: How Harmonized Survey Data Can Contribute to Scientific Breakthroughs

Abstract

This presentation outlines the theoretical questions of how people incorporate increasingly prevalent nonelectoral political acts into individual-level repertoires of participation; and how well represented these different types of political participators are. At a time of growing concern about unequal representation in democracies, two conflicting global trends over the last several decades highlight the importance of these questions: a clear decline in voter turnout, especially among lower status groups; and evidence of increased nonelectoral participation, especially among higher status groups. To assess how these trends in political participation affect patterns of representation, the presented theoretical framework integrates new approaches for investigating the links between individuals' participation repertoires (i.e., how individuals combine voting, protest, online activism) and objective and subjective representational outcomes. The presentation draws on a new ERC Starting Grant project on "Participation and Representation in the Digital Age." Additional information on the project can be found at this [website](#).

Joonghyun Kwak, Oxford University, Kazimierz M. Slomczynski, IFiS PAN & OSU

Constructing Cross-National Time-Series Index of Trust in Political Institutions: A Survey Data Recycling Framework

Abstract

Summary statistics derived from national surveys are commonly used as macro indicators in cross-national comparative analyses. However, most studies do not account for inter-survey variability arising from survey item formulations and data quality. This study proposes a novel approach for aggregating survey data into cross-national time-series indicators of trust in political institutions—parliament, legal systems, and political parties—by addressing methodological differences, biases and errors in national surveys. To obtain the best estimate of the trust index, this study constructs a measure of trust on the basis of rating and distributional scales and their product, controlling for (1) methodological differences present in the original survey questions, and (2) variations in survey data quality. Drawing on the Survey Data Recycling framework for harmonized survey data, two aggregation methods are discussed: the residual method, which captures the substantive aspect of trust after accounting for variation in question formulation and survey quality, and the prediction method, which estimates the predicted level of trust by lagged value of trust and other harmonization and survey quality control variables. This study provides a reliable cross-national time-series index of political trust that can be used by policymakers and academics in their policy-making processes and research.

Ilona Wysmulek, IFiS PAN, Nika Palaguta, YOTTA, Weronika Boruc, Mercer

Gender Items across International Survey Projects: Coding Trends and Challenges

Abstract

Despite growing recognition of gender diversity in contemporary society, survey research has been slow to adopt gender-neutral questionnaires and non-binary gender options. This study examines the extent to which survey respondents' gender is coded beyond the binary order of male and female, and explores factors that may influence this coding. Drawing on data from 3,715 national surveys conducted between 1966 and 2017 in 142 countries worldwide, pooled within the SDR project, we investigate trends and challenges in gender coding across international survey projects. Our analysis considers differences in gender coding related to standards of international projects, relationships between gender coding methods and countries, and changes in gender coding in sociological surveys over 51 years. This research contributes to discussions on gender and sex in survey research, sheds light on potential methodological changes, and identifies the advantages and challenges of using survey data for gender research.

Francesco Sarracino, STATEC Luxembourg

A new assessment of happiness findings using SDR 2 data

Abstract

SDR 2 data provide the unique opportunity to comparatively analyze the largest sample of countries over the longest period of time. This is possible thanks to a huge harmonization effort of many publicly available survey data from all over the world. The result is an individual level dataset with more than 4 million observations in more than 100 countries

observed from late 1960s till 2019. In this presentation, I will use SDR 2.0 to provide a new assessment of some well-known issues in the subjective well-being literature, such as the role of income for subjective well-being; the buffering effect of social capital against social comparisons; and the correlates of subjective well-being worldwide. Finally, I will leverage SDR 2.0 data to study how the features of subjective well-being questions affect respondents' answers.

15:30 – 15:45 *Coffee break*

15:45 – 16:30 *SDR Project Advisory Board meeting*

Preparatory session for the Multilevel Models Workshop

16:30 – 18:00 Brief introduction to R

Topics: Getting started with R, RStudio interface, basic R objects, and importing data into R.

Suggested readings:

Navarro 2019: Chapter 3 & 4

Workshop Program¹

Location: IPR Conference Room, Townshend Hall

The Workshop features the SDR database 2.0 (SDR 2) as a key empirical resource to discuss methodological considerations in analyzing multi-source databases constructed via ex-post harmonization. Lectures and practical sessions are run by Dr. Malgorzata Mikucka, Mannheim University, and Dr. Michal Kotnarowski, Institute of Philosophy and Sociology, Polish Academy of Sciences. The workshop is addressed mainly to PhD students and post-doctoral scholars

Wednesday, May 10

10:00 -11:30

Lecture 1. Introduction to multilevel modeling

When and why use MLMs? Basic concepts: levels, units, and hierarchical structures. Notation. Repetition of regression analysis. MLMs vs. traditional regression; problems with aggregation and disaggregation.

11:30 – 11:45

Coffee break

11:45 – 13:00

Lecture 2. Multilevel models, continuous Dependent Variable, DV (Part 1)

Random intercept models. Random intercept random slope models. Model fit assessment.

¹ For details, including readings, see the MLMs Lectures and Practical sessions' syllabi.

- 13:10 – 14:00 *Lunch break*
- 14:00 – 15:00 **Practical session 1.** Features of the SDR 2 dataset. Preparing SDR 2 dataset for analyses.
- 15:00 -15:15 *Coffee break*
- 15:15– 16:30 **Practical session 2.** Multilevel models, continuous DV

Thursday, May 11

- 10:00 -11:30 **Lecture 3.** Multilevel models, continuous DV (Part 2)
Estimation techniques. Sample size and power. Robustness of MLMs. Influential cases.
- 11:30 – 11:45 *Coffee break*
- 11:45 – 13:00 **Lecture 4.** Cross-level interactions. 3-level models. Cross-classified models.
Between-group and a within-group variation; centering. Cross-level interactions. 3-level models; MLMs for repeated cross-national data. Cross-classified models.
- 13:00– 14:00 *Lunch break*
- 14:00 – 15:15 **Practical session 3.** Cross-level interactions.
- 15:15 -15:30 *Coffee break*
- 15:30 – 16:30 **Practical session 4.** 3 -level models.

Friday, May 12

- 10:00 -11:30 **Lecture 5.** Multi-level modeling for dichotomous DV
Logistic regression: a brief repetition. MLMs for dichotomous dependent variables.
- 11:30 – 11:45 *Coffee break*

11:45 – 13:00	Practical session 5. Cross-classified models.
13:00 – 14:00	<i>Lunch break</i>
14:00 – 15:15	Practical session 6. Multi-level modelling for dichotomous DV.
15:15 - 15:30	Closing discussions

Brief Biographies of Presenters and Instructors

J. Craig Jenkins

J. Craig Jenkins is Academy Professor of Sociology, Ohio State University and writes on protest and conflict, methods of event data analysis, the harmonization of survey data, the political economy of high technology development, and problems of environmental sustainability. He was elected Fellow of the American Association for the Advancement of Science, the Robin Williams Distinguished Contribution Award, a Fulbright Fellow to Norway, and has held numerous elective offices in the American Sociological Association. Most recently he edited *The American Behavioral Scientist* special issue on “Innovative Methods in the Study of Protest.”

Michał Kotnarowski

Michał Kotnarowski is an Assistant Professor at the Institute of Philosophy and Sociology, the Polish Academy of Sciences, a lecturer at GSSR, and the National Coordinator of the European Social Survey (ESS) in Poland. Michał has extensive teaching experience in quantitative methods for the social sciences, including as instructor of the Generalized Linear Models course at the 2018-2020 Winter Schools organized by the European Consortium for Political Research, ECPR. His work on voting behavior, comparative politics and political methodology has been published in, among others, *Party Politics*, *Electoral Studies*, *Communist and Post-Communist Studies*, *Acta Politica*, and the *International Journal of Sociology*

Joonghyun Kwak

Joonghyun Kwak is a postdoctoral researcher in the Department of Education at the University of Oxford. He is a quantitatively oriented, comparative sociologist with a focus on globalization, international migration, social and educational inequalities, and survey research methods. His current research project examines the impact of international student mobility on their home countries, particularly in terms of economic inequality, poverty, and the development of democracy.

Olga Li

Olga Li is a PhD candidate at the Graduate School for Social Research and is a member of the research unit on Comparative Analyses of Social Inequality at the Institute of Philosophy and Sociology, Polish Academy of Sciences. For her PhD thesis, she is conducting a quantitative research on political participation in authoritarian regimes.

Malgorzata Mikucka

Malgorzata Mikucka works as a researcher at Mannheim University, where her primary research interests revolve around studying social inequalities, as well as social determinants of health and well-being. Her research has been published, among others, on *Journal of Marriage and Family*, *World Development*, *Journal of Happiness Studies*, and *European Journal of Population*. With a strong quantitative background, she has been teaching multilevel models since 2015. Malgorzata earned her PhD from GSSR in 2010 and presently holds the positions of treasurer and newsletter editor for the European Society for Health and Medical Sociology.

Jennifer Oser

Jennifer Oser is an associate professor in the Department of Politics and Government at Ben-Gurion University in Israel. She completed her Ph.D. at the Hebrew University of Jerusalem in Israel and has conducted research as a Visiting Research Fellow at Harvard University and at the University of Pennsylvania. Her research agenda focuses on the relationships between public opinion, political participation, and policy outcomes.

Przemek Powalko

Przemek Powalko is a Data Scientist at the Institute of Philosophy and Sociology PAN. He has been with the SDR project from the beginning, including in the Harmonization Project that prepared the first version of the SDR database, currently archived at Harvard Dataverse.

Francesco Sarracino

Francesco Sarracino is head of the research unit on subjective well-being and entrepreneurship at STATEC, the national institute of statistics of Luxembourg. His work aims at identifying policies to make economic growth compatible with people's well-being and to pursue socially and environmentally sustainable development. He holds a PhD in development economics from the University of Firenze (Italy). He is a member of the World Wellbeing Panel, and of the board of directors of the International Society for Quality of Life Studies. His works have been published in journals such as *Oxford Economic Papers*, *World Development*, *Ecological Economics*, *Journal of Happiness Studies*, *Social Indicators Research* and *Journal of Business Ethics*.

Kazimierz M. Slomczynski

Kazimierz M. Slomczynski is a professor of Sociology at the Institute of Philosophy and Sociology PAN, and Academy Professor of Sociology, OSU. He directs the *Cross-National Studies: Interdisciplinary Research and Training* program (CONSIRT) of OSU and PAN. Maciek is co-PI of the SDR Project on ex-post harmonization of cross-national survey data. He initiated the Polish Panel Survey, POLPAN, a unique longitudinal study on the transformation of the Polish stratification system after 1989. Maciek currently oversees, as Senior Advisor, the design and implementation of POLPAN's 2023 wave.

Irina Tomescu-Dubrow

Irina Tomescu-Dubrow is a professor of sociology at the Institute of Philosophy and Sociology, the Polish Academy of Sciences, PAN, and manages the *Cross-national Studies: Interdisciplinary Research and Training* program of OSU and PAN. Irina is PI of the project "Structures and Futures: The Polish Panel Survey, POLPAN 1988-2023" (polpan.org), funded by Poland's National Science Centre, and co-PI of the SDR Project on ex-post harmonization

of cross-national survey data (NSF# 1738502). She now directs the Graduate School for Social Research, GSSR (gssr.edu.pl), housed at IFiS PAN.

Yamei Tu

Yamei Tu received a BS degree in Software Engineering from East China Normal University. She is a Ph.D. student in the GRaphics & Visualization sTudY (GRAVITY) Research Group under Computer Science and Engineering at the Ohio State University. Her research interests are Visualization, Text Analysis, and Machine Learning.

Christof Wolf

Christof Wolf is President of GESIS - Leibniz Institute for the Social Sciences and Professor of Sociology at the University of Mannheim. He studied sociology, economics, economic and social history, and statistics at the University of Hamburg. In 1996 he received his doctorate in sociology from the University of Cologne and in 2003 he received the *venia legendi* for sociology. His research interests include issues related to the production of social science research data and survey methodology. Christof contributed to various survey programs and is currently one of the PIs of the German Family Demographic Panel FReDA and the German part of the International Social Survey Programme ISSP.

Ilona Wysmulek

Ilona Wysmulek is an assistant professor of Sociology in the research group on Comparative Analysis of Social Inequality at the Institute of Philosophy and Sociology, PAN. Ilona has been working in the SDR project since 2013, supervising ex-post harmonization work on over 3,000 national surveys of 23 international projects and leading international workshops on the [Harmonization Workflow in SDR](#) (Wysmulek, 2019). She co-edited (with I. Tomescu-Dubrow and J. Kwak) the 2022 Special Issue “Ex-post harmonization of cross-national survey data: advances in methodological and substantive inquiries.” *Quality and Quantity* <https://doi.org/10.1007/s11135-021-01187-7>.

Acknowledgements

This event is organized by CONSIRT (consirt.osu.edu) at The Ohio State University and the Polish Academy of Science, as part of, and with funding from, the Survey Data Recycling project (NSF 1738502, <https://wp.asc.ohio-state.edu/dataharmonization/>). We thank the OSU Department of Sociology for organizational support.

About CONSIRT

The CONSIRT Program was established by The Ohio State University and the Polish Academy of Sciences. Its main objective is to stimulate and facilitate research, graduate training, and undergraduate education. CONSIRT provides an opportunity for preparing and administering complex research projects with a cross-national perspective. For more information, please visit our website, consirt.osu.edu.