Harmonization:
Newsletter on Survey Data
Harmonization in the Social Sciences

Listening and Learning

Welcome to the latest and, up to now, largest issue of *Harmonization*: Newsletter on Survey Data Harmonization in the Social Sciences. In Fall 2016, we called for newsletter content and the enthusiastic and growing survey data harmonization community of scholars, institutions, and government agencies who work on harmonizing social survey data and other projects with similar focus responded in a big way. Listening to each other, and learning from each other, this community continues to move forward the interdisciplinary field of survey data harmonization.

This issue features articles on a variety of methodological topics. **Tom Smith, of NORC at the University of Chicago**, discusses recent projects in survey data harmonization. **Claire Durand and colleagues at the University of Montreal** present their projects on analyzing trust in institutions using surveys pooled across time and countries. **Zbigniew Sawinski, long-time methodologist of the Polish Panel Survey POLPAN**, presents a schema of inter-wave harmonization of panel data. Two **graduate students at the Graduate School for Social Research of the Polish Academy of Sciences** discuss their dissertation projects on harmonizing ethnic minority status in surveys of post-Soviet nations (**Olena Oleksiyenko**) and on harmonizing corruption items in international survey projects (**Ilona Wysmulek**). We also include news from IPUMS (**Catherine A. Fitch**) and GESIS (**Kristi Winters**).

The harmonization community continues to present their research at conferences and workshops around the world. In this issue, we have reports from the International Political Science Association meeting in Poland, the QDET2 in Miami, Florida, the 3MC conference in Chicago, Illinois, and the International Social Survey Programme meeting in Lithuania.

As always, we invite all scholars interested in survey data harmonization to read our newsletter and contribute their articles and news to future editions.

In This Issue

Acknowledgements

Articles

Pre- and Post-Harmonization in Comparative Surveys: Applications and Current Initiatives

by Tom W. Smith, NORC at the University of Chicago

To achieve functional equivalence and maximize comparability across cross-national/cross-cultural surveys, pre- and/or post-harmonization are needed (also known as input and output, and ex-ante and ex-post). When designing a new cross-national study, the focus is on pre-harmonization. Researchers design items, scales, and questionnaires to be comparable across countries and languages. When dealing with already collected data, post-harmonization is used to create a cross-national file. Often, researchers need to employ both pre- and post-harmonization. For example, the International Social Survey Program (ISSP) uses pre-harmonization in the design of the 60 substantive questions developed for each annual module, but uses post-harmonization to convert national demographics to ISSP cross-national standards. Even studies like the European Social Survey (ESS) that include pre-harmonization of demographics, need a post-harmonization phase to deal with nation-specific measures such as geographic variables and education (europeansocialsurvey.org/docs/round6/survey/ESS6_appendix_a1_e02_0.pdf).

Notable advances in harmonization have appeared in recent years...

There are also post-harmonization projects that merge and/or make more comparable studies that were not originally designed for comparative purposes, such as:

- Cross-national Equivalent File: cnef.ehe.osu.edu
- Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective dataharmonization.org
- Eurostat: epp.eurostat.europa.eu
- Harmonized European Time Use Study: h5.scb.se/tus/tus
- Integrated Public Use Microdata Series International (IPUMS-I): international.ipums.org/international
- International Stratification and Mobility File: sscnet.ucla.edu/issr/da/Mobility/mobindex.html
- Luxembourg Income Study (LIS): lisproject.org
- United Nations: unstats.un.org

In addition, there are variable-level harmonization efforts rather than those designed to make major studies comparable (Hoffmeyer-Zlotnik and Wolf, 2003). One example is the International Standard Classification of Occupations (ISCO) which was developed by the International Labour Organization in 1958 and updated most recently in 2008 (ilo.org/public/english/bureau/stat/isco). It can be used to directly code occupations cross-nationally or nation-specific occupational codes that can be converted to ISCO for international comparability. For example, there is a crosswalk between ISCO08 and the US-based 2010 Standard Occupational Classification (bls.gov/soc/soccrosswalks.htm). Another example is the International Standard Classification of Education (Briceno-Rosas, 2016; Ortmann, 2016; Schneider, 2008; 2009; 2016). Currently CAMCES – Computer-Assisted Measurement and Coding of Educational Qualifications, a project at GESIS, is finalizing a tool to measure educational qualifications in surveys by accessing a large, international database of educational degrees and assigning educational codes based on detailed, national, educational qualifications.

Finally, harmonization may consist of only preparing aggregated data in the form of tables and reports (e.g. in many UN reports), while in other cases micro-level datasets are produced (e.g. LIS and IPUMS-I).

Notable advances in harmonization have appeared in recent years. Winters and Netscher (2016) at GESIS have developed a tool to better document data harmonization. Moreover, the Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective (Breustedt, 2015; "Democratic Values," 2016; Dubrow, 2015; Lillard, 2015) has advanced harmonization as a general object for development rather than just being applied for a creation of a project-specific merged file. It of course publishes the journal: Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences (consirt.osu.edu/newsletter) in which this contribution appears.

References

Briceno-Rosas, Roberto, "Asking about Education in Multinational, Multiregional, and Multicultural Contexts: Results from Cognitive Pretesting in Two Countries," Paper presented to the Second International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts (3MC 2016), Chicago, July, 2016.

Breustedt, Wiebke, "The Barometer Surveys: Insights into the Quality of the Harmonized Political Trust Items," Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences, 1: No. 2 (2015), 7-10.

Democratic Values and Protest Behavior: Data harmonization, Measurement Comparability, and Multi-level Modelling in Cross-national Perspective, "Data Harmonization: About the Project," http://dataharmonization.org/about/

Dubrow, Joshua Kjerulf, "A Brief History of Survey Data Harmonization Projects," Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences, 1: No.1 (2015), 2-4.

Hoffmeyer-Zlotnik, Juergen H. P. and Wolf, Christof, eds. Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables. New York: Kluwer Academic, 2003.

Ortmanns, Verena, "The CAMCES Database and Its Cross-national Education Coding System," Paper

presented to the Second International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts (3MC 2016), Chicago, July, 2016.

Schneider, Silke, "Computer Assisted Measurement and Coding of Educational Qualifications in Multicultural Surveys (CAMCES): A New Set of Survey Tools," Paper presented to the Second International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts (3MC 2016), Chicago, July, 2016.

Schneider, Silke L., "Confusing Credentials: The Cross-nationally Comparable Measurement of Educational Attainment," Ph.D. dissertation, University of Oxford, 2009.

Schneider, Silke L., ed., The International Standard Classification of Education: A n Evaluation of Content and Criterion Validity for 15 European Countries. Mannheim: MZES, 2008.

Winters, Kristi and Netscher, Sebastian, "Proposed Standards for Variable Harmonization Documentation and Referencing: A Case Study Using QuickCharmStats.1," PLoS One, 11.2, February 9, 2016, dx.doi.org/10.1371/journal.pone.0147795

Tom W. Smith is the Senior Fellow and Director of NORC's Center for the Study of Politics and Society, Ogburn-Stouffer Center for the Study of Social Organizations, at the University of Chicago. He is the Director of the General Social Survey (GSS) and a co-founder of the International Social Survey Program (ISSP).

Harmonize or Control? The Use of Multilevel Analysis to Analyze Trust in Institutions in the World

by Claire Durand, Isabelle Valois and Luis Patricio Peña Ibarra, Department of Sociology, University of Montreal

This article presents the current progress of a research project whose aim is to develop methods to analyze combined micro-data. The most recent papers that were presented (see references) give an insight into the work that has been accomplished to date.

This project was triggered by a preceding project where we aimed at analyzing change in support for Quebec sovereignty over time taking into account that question wordings and specific constitutional choices offered in survey questions varied over time and between surveys. We had identified close to 700 questions asked in polls over a 40-year period. In order to analyze these data, we used a multilevel model where polls were embedded within months. This allowed for analyzing the impact of question characteristics at level 1. Since time itself was at level 2, we could study change in support for sovereignty over time and the impact of events that occurred during each period controlling for question wording and constitutional choices (Yale & Durand, 2011).

This research ended with a frustration. We would have liked to be able to answer questions like whether the impact of age on support for sovereignty was fading over time. This requested combining micro-data, not just poll results. Therefore, we decided to combine data sets in order to be able to answer our research questions. However, instead of maintaining the focus on Quebec sovereignty, the focus was changed to institutional trust.

The first project is Valois's Trust in Canada which involves combining survey data over a 40-year period; the second and third project started with the objective of combining all the surveys that had questions on trust in institutions everywhere in the world. One project is Durand et al.'s Trust in the World who combined the data from all the Barometers conducted outside Europe and the Latin American Public Opinion Project (LAPOP) surveys; the other one is Peña Ibarra's Trust in Latin America, which uses a subset of these data to focus on Central and South America plus Mexico. The basic information on the three projects is presented in Table 1.

Table 1. Synthesis of the three projects			
	Trust in Canada	*Trust in the world*	*Trust in Latin America*
Period	1974-2014	1994-2014	1994-2014
Number of polls	63	634	421
Number of countries	1	98	26
Number of units			
At level 4 (survey-country)	NA	121	NA
At level 3 (survey-year/year)	63	634	421
At level 2 (respondents)	194,947	827,131	547,954
At level 1 (trust questions)	1,184,112	9,747,094	6,834,047
number of diff. institutions		60	36
number of categories of institutions	9	14	14
Distribution of variance (explained variance)			
Survey-country level		7.5% (28.9%)	
Survey-year/Year level	10.3% (61.2%)	2.6% (0%)	4.4% (27.1%)
Respondent level	22.4% (0%)	27% (0%)	27.4% (0%)
Trust level	67.2% (12.3%)	62.9% (7.9%)	68.2% (9.1%)

Valois's Trust in Canada Project

The first project is Isabelle Valois' doctoral research. Valois' first goal was to examine the hypothesis of a decline in trust over time in Canada. While it is rather easy to find data on trust in Europe or in the U.S., it is much more difficult to find similar data for Canada. As a first step, Valois tried to identify all the surveys that had questions pertaining to trust in institutions in Canada. She managed to find 63 surveys conducted from 1974 to 2014 – some of them from International survey programs like the World Values Survey (WVS), the International Social Survey Programme (ISSP) and LAPOP. Valois ran into a number of problems related to the variation in the questions pertaining to trust and to socio-demographic characteristics:

Concerning the questions on trust,

- First, there was variation between surveys and over time in the institutions on which the trust questions were asked.

- Second, the question wordings varied also. In English, some used the term "trust", others, the term "confidence" (in French, there is only one term, confiance). In addition, some questions asked about the institutions themselves, other about the leaders of these institutions – "the people in charge of" – and others about the employees, i.e., teachers instead of the school system, for

example. These differences in wordings may have an impact on the level of trust reported and therefore, should be controlled for.

- Third, the number of choices in the scales used to measure trust also varied.

At the individual level,

- Variables like age, education, and income had to be harmonized to a smaller common denominator. However, since the period is long – 40 years – categories may change meanings over time. Being less than 34 years old for example, may not have meant the same thing when this age group comprised close to 50% of the population (in the 1970's) compared to less than 30% more recently. One aim was therefore to be able to take into account the varying composition of the population.

While there is not much choice but to use the smallest common denominator when it comes to socio-demographic characteristics, the situation is different with the trust questions. The decision was taken to enter the information as it was found in the original data files and to introduce variables that inform on the type of question wording and the number of anchors in the scales used, a solution that had been used in the Change in Support for Sovereignty project. This meant putting all the answers on a "largest common denominator" scale, i.e., a 7-point scale. The 4-point scales were recoded into 1, 3, 5, 7, the 3-point scales into 2, 5, 6, etc. and each survey was coded on the type of scale originally used.

The second major problem to deal with was whether it was necessary to "harmonize" right away the different institutions. New institutions kept appearing with any new survey entered in the data base. The solution to that problem was not found right away. The second project allowed to solve it and the solution was applied a posteriori to Valois's Trust in Canada project.

The combined data base of 63 surveys for this project comprises 194,947 respondents with a total of 1,184,112 measures, i.e., an average of 6.07 items answered per respondent. These measures pertain to trust in up to 49 institutions, grouped into nine categories (see Table 2 for a description of the institutions and categories in which they are grouped). The pooled data were analyzed with a 3-level longitudinal multilevel model of repeated measures. Polls are at the highest level. It is at that level that the impact of time and of events can be tested. Respondents are at level 2. The possible impact of age, sex, and region is analyzed at this level. Answers to questions on trust are level 1. This allows for analysing the impact of the object of trust – the different institutions – on the level of trust. Valois (2016a, 2016b) shows that, although there was no decline in overall trust during the period, trust in specific institutions did change. Trust in religion, for example, declined during the period while trust in the army started increasing after 9/11. Valois also shows that trust in the political institutions in Canada declined much during the period of intense constitutional debate that preceded the Quebec referendum on sovereignty in 1995. Afterwards, it increased back to its pre-crisis level and even higher.

Table 2: Categories and institutions included in the various projects

Trust in the World & Trust in Latin America

Institutions of the civil society

Media	Media; Newspapers; Radio; Television; *Governmental Broadcasting; Independent Broadcasting; National Broadcasting; Govermental Newpapers; Independent Newspapers;*
Religion	Church; Catholic Church; Evangelical/Protestant Church; *Religious Leaders*
Trade unions	Trade Unions
NGO- Civil Society	Non-governmental Organizations; Indigenous Movements; *Institutions of the civil society*

Institutions of the administration & the judiciary system

Army/police	Armed Forces, Police, *Police station*
Public administr.	Public Administration; *Civil Service; Public Educational System; Public Health System*
Judiciary System	Attorney Office; Consitutional Court; Corruption Commission; Judicairy System; Ombudsman Office; Public Ministry; Superior Court; Supreme Court

Institutions of the economy

Financial inst.	Banks; Stock Exchange
Entreprises	Large Domestic Companies; Private Enterprises; State enterprises; Multinational Companies; Small Businesses; Local markets;

Institutions of the political system

Political parties	Political Parties; *Ruling Political Parties; Opposition Political Parties*
Government	Congress/Parliament; Government; Local Government; National Legislature; Regional Governement Body; *Local Governement Body; Local Government elected persons*
State/President	Executive Power; President; State; *Our Country*
Elections	Elections; National Electoral Commission
Intn'l organizations	United Nations; *International Monetary Fund; World Bank; World Trade Organization*

* The institutions that are absent in Latin America are in italics.

Trust in Canada

Institutions of the civil society

Media	Press; People in charge of the Press; Journalists; News media; Media; Press and Media; Mass Media; Canadian Media
Religion	Church; People in charge of the Church
Trade unions	Unions; People in charge of the Unions

Institutions of the administration & Judiciary

Army	Armed Forces; Canadian Armed Force; People in charge of the army
Police	Police; Police Force; RCMP; People in charge of the Police
Education System	Public Shcool; Teachers; Education System; People in charge of the Education System
Judiciary System	Supreme Court; Judges & Lawyers; Justice System; People in charge of the Justice System

Institutions of the economy

Economy	Large Corporations; Big Business; People in charge of Big Business; People in charge of Business; Business & Industry; Gobal Companies; Canadian Corporations; Private Enterprise; Entrepreneurs; Banks; People in charge of the Banks

Institutions of the political system

Political System	House of Commons; Parliament; Politicians; Political Parties; Goverment; People in charge of Government National Goverment; Federal Government; People in charge of the Federal Government; Provincial Government People in charge of the Provincial Governement

The Durand, Peña Ibarra and Charest Project

Following the problems regarding the multiplication of the number of institutions encountered in the Valois project, it became obvious that it was necessary to develop and test methods that could

allow for an analysis of combined survey results that take into account the substantial variation in the institutions on which trust questions were asked.

We proceeded to combine the data sets of all the Barometers, giving precedence to the Barometers conducted outside Europe. We combined the Latino Barometro, the Asian and East Asian Barometers, the Africa Barometer and the Arab Barometer. In doing so, we added a new variable each time a new institution appeared in a data file. For example, in the Latino Barometro, there was a question on trust in the Church, but in the Arab Barometer, a similar question pertained to trust in religious leaders. This information was kept as two questions. In this way, all the original information was kept in our combined data base. We then decided to add the data from LAPOP that had questions similar to the Latino Barometro but used 7-anchor scales. Finally, we withdrew the data from the United States and Canada present in LAPOP and the data from Spain present in the Latino Barometro so as to restrict ourselves to the regions for which we had data from many countries.

> *The combined file finally comprises 635 surveys from four regions of the world, conducted from 1994 to 2014.*

This combined file finally comprises 635 surveys from four regions of the world, i.e., Latin America, Asia, Africa and the Middle East and North Africa (MENA), conducted from 1994 to 2014. These surveys were conducted in 98 countries. Since surveys were conducted over 20 years (for the longest stretch), it gives us 560 country-years. There are 827,131 respondents in the combined data file and answers to 9,747,094 questions, an average of 11.8 per respondent, pertaining to trust in 60 different institutions.

- At the first level, there are the different answers given by respondents to the trust questions. These answers are embedded within each respondent so that the fact that a single respondent answered the different questions is controlled for. Each of these answers is coded for the institution on which the trust question is asked. Table 2 shows how the 60 different institutions were categorized into 14 groupings for analysis. The advantage of grouping institutions a posteriori instead doing so from the outset is triple: First, we can check whether there are differences between the different institutions that are grouped together a priori; we can modify the groupings as needed; and finally, if new institutions appear when we combine new data files, they can be either integrated in the current categories or new categories can be introduced.

- At the second level, there are respondents. There is not that much common information on the respondents present in all the data files. Age or age categories and sex are present in all the surveys. Education is present in all but seven of them. The harmonized categories are: no formal education, primary school, secondary school/high school, professional/technical school and university/graduate school. Measures of occupation are present for 77% of the respondents, with the following harmonized categories: employed, including self-employed and salaried, out of

work, retired, homemaker, student and other.

- At the third level, there is time. Since it is possible that a curvilinear relationship is present, two variables were computed, time (centered on the mean) and time squared. Having a separate level for time has very interesting advantages. It allows for checking hypotheses regarding a differential change in trust over time for different institutional categories or whether differences according to age for example decrease over time.

- Finally at the fourth level, there is each country. The 98 countries are categorized according to region of the world, i.e. Latin America, Asia, Africa and MENA. In addition, at that level, we control for the impact of using a 7-point scale in LAPOP instead of a 4-point scale. This requires us to keep the Latino Barometro and LAPOP files in separate units for the same countries. Therefore, we have 121 units at that level.

A first presentation of this analysis can be found in Durand, Charest, Peña Ibarra & Valois (2016a) and Durand, Charest & Peña Ibarra (2016a, 2016b) for 3-level and 4-level models. This project showed that political parties get the lowest level of trust in all regions of the World and that trust is generally lower in Latin American than elsewhere. It also showed that there was a decrease in trust in some institutions in the MENA countries following the invasion of Iraq. However, on average, the level of trust did not change much over time. Finally, it allowed for determining that the difference in average trust according to age groups was fading over time.

The next steps of this project are the following:

- We would like to integrate variables that characterize the different countries over time. These can be economic or socio-political indicators. We face the problem of finding indicators that are available and comparable for all countries and periods.

- Second, we would like to integrate the data files from the World Values Survey and the International Social Survey Programme on religion in 1998 and 2008 in which there are questions pertaining to trust.

- Finally, we will go on combine new data files from all the projects as they become available.

The Peña Ibarra Project

Luis Peña Ibarra was the main research assistant who combined all the files for the international project. For his doctoral research, he decided to focus on a subset of that project, i.e., South and Central America, and therefore on the combination of the Latino Barometro and LAPOP projects. Focussing on a specific region has two advantages. First, the trust questions have more comparability. For example, trust in Church means quite the same thing in all South and Central America. In addition, a number of socio-demographic variables are also more comparable. The second advantage is that it makes it easier to find macro level indicators that characterize each country over time.

The combined file pertains to 26 countries with data from 1994 to 2014. The restricted number of countries makes it impossible to use a 4-level model. The highest level comprises 421

units at the level of country-survey-year. It comprises 547,954 respondents with answers on a total of 6,834,047 questions of trust in institutions, an average of 12.5 questions per respondent. Table 2 shows the 36 institutions that are present in this data file.

Peña Ibarra et al. (2016) could show that trust is stable on average because there are compensation mechanisms, i.e., when trust in the Church decreases, trust in the media increases. At the country-year level, he tested whether belonging to a common economic market -- like the Mercado común del sur (MERCOSUR), the Mercado común centroamericano (MCCA) and the Comunidad económica del Caribe (CARICOM) -- and whether the ideological orientation of the government – from right to left – were related to trust. He could show that when left-wing parties form the government, the level of trust is, on average, 0.14 points higher on a seven-point scale.

Analysis

The lower part of Table 1 presents the distribution of the variance at the different levels for the three projects and, in parenthesis, the variance that has been explained by the independent variables entered. It is interesting to notice the similarity in the distribution of variance between projects. The variance at the trust level – between institutions within individuals – varies from 63% to 68%. It is lower when there are more countries, which is expected since the many different contexts take some variance. Between 8% and 12% of that variance is explained by the different institutions and the variation in trust according to region and time for these institutions. The fact that two-thirds of the variation is between institutions means that analysing only one measure of trust or a mean of different measures would not have been appropriate.

Between 22% and 27% of the variance is found between individual respondents. At this stage, the variables used to explain this variance are age and sex (and region in Canada). Sex is significant only in the case of Canada. Region is also significant in Canada. Age is significant in all three files. However, the contribution of these variables is not substantial enough to explain a significant amount of the variance.

At the survey-year level, there are expected differences between projects. In Canada, there is data for only one country but over a 40-year period with surveys that come from very diverse sources. This is where the variance between surveys is highest at 10.3%. In Latin America, there are 26 countries surveyed over 20 years. There is no substantial variation over time. The variance at that level is 4.4% and 27% of it is explained, mostly by the ideological orientation of the government and by the type of scale used (7-anchor vs 4-anchor). The Trust in the World project is the only one with four levels. Most of the variance is between countries, but not between years. This confirms that trust is, on average, quite stable over time unless a specific event has an impact on trust in a specific organization or unless trust a specific organization is declining (Church in Latin America and in Canada, for example). At the country level, there is 7.5% of the variance, which is not much. This variance can be explained by region and by the scale used to measure trust. In general, using a 7-anchor scale gives an estimation close to half a point higher.

Discussion and Conclusion

In this project, we decided that it was better to "go-around" the problem of harmonization of data instead of grouping items of trust in institutions that focus on similar institutions in large categories or keeping only items that were common across all surveys. It became obvious from the beginning, in the Valois Trust in Canada project, that harmonizing the trust questions meant losing much information. In addition, every new file brought new problems to solve. 'Going around' meant finding ways to analyse the data as is and control a posteriori for the fact that different scales and question wordings were used and that different institutions were surveyed.

We would like to improve our understanding of the variance that is found at the different levels. For example, we could not yet explain much variance between countries and at the individual level. We will need to find indicators that are available for all countries and year and are related to trust. At the individual level, unfortunately, there is not much information that we can add. However, analysis performed with education, though it forces us to withdraw seven surveys, show that this is the major new indicator to add into the analyses.

Finally, we would like to let researchers have access to these files. However, legal problems may restrict our capacity to do so. We depend on the cooperation from the various projects in finding ways to let researchers access the harmonized data files. We will work towards solving this problem in order to be able to share our work.

References

Durand, C., Peña Ibarra, L. & A.S. Charest (2016a). L'évolution de la confiance dans les institutions, une analyse transnationale, multiniveaux et longitudinale. 9ème Colloque francophone sur les sondages, Gatineau, Québec, 11-14 octobre, 2016. Text : 3-level analysis: http://paperssondages16.sfds.asso.fr/submission_41.pdf; Diaporama: 4-level analysis : http://sondages2016.sfds.asso.fr/wp-content/uploads/2016/10/Session05-Durand.pdf

Durand, C., Charest, A.S., Valois, I. & L.P. Pena Ibarra (2016a). A Transnational Analysis of Change in Trust over Time. WAPOR 69th Annual Conference, Austin, Texas, May 10-12, 2016. http://www.mapageweb.umontreal.ca/durandc/Recherche/Publications/Combining_data/Transnational_analysis.pdf

Durand, C. Charest, A.S., Valois, I. & L.P. Pena Ibarra (2016b). The use of longitudinal multilevel models to understand change in trust in institutions over time and between countries, Presented at the ISA-RC33 Conference, September 11-16, 2016.

Durand, C. et I. Valois (2015). Pour une analyse historique des données d'enquête: Le cas de la confiance au Canada, Conférence organisée par la Société Statistique de Montréal (SSM) et la Faculté des Sciences de l'Éducation de l'Université de Montréal, 17 Novembre 2015.

Durand, C., I. Valois & F. Yale (2015). Historical Analysis of Survey Data and Survey Results: The Incredible Possibilities Afforded by Longitudinal Multilevel Analysis Using Time at a Higher level, European Survey Research Association 6th conference, Reykjavik, Iceland, July 13-17, 2015. http://www.mapageweb.umontreal.ca/durandc/Recherche/Publications/Combining_data/Historical_analysis.pdf

Peña Ibarra, L.P., Durand, C., Charest, A.S. & I. Valois (2016). ¿Crisis de confianza en las instituciones en América Latina? VII Congreso WAPOR Latino América, Monterrey, México, 12-14 octubre, 2016.

Valois, I. (2016a). La confiance envers les institutions au Canada est-elle en diminution? Une analyse longitudinale Text: http://paperssondages16.sfds.asso.fr/submission_70.pdf; Diaporama: http://sondages2016.sfds.asso.fr/wp-content/uploads/2016/10/session_5_Valois.pptx

Valois, I. (2016b). Is institutional confidence in Canada declining? A longitudinal analysis, WAPOR Political Trust in Contemporary Representative Democracies, Barcelona, Spain, November 24-25, 2016.

Yale, F. & C. Durand (2011). What did Quebeckers want? Evolution of Support for Sovereignty1976-2008. American Review of Canadian Studies, 42 (3), p. 242-258.

Professor Claire Durand, Isabelle Valois and Luis Patricio Peña Ibarra are at the Department of Sociology, University of Montreal. The presentations from this article are available on ResearchGate or on Claire Durand's website http://www.mapageweb.umontreal.ca/durandc

Inconsistent Responses: Hard Adjustment and Soft Adjustment as Solutions for Inter-Wave Harmonization of Panel Surveys

by Zbigniew Sawiński, Institute of Philosophy and Sociology, Polish Academy of Sciences

When respondents are asked the same questions in different waves of a panel survey they may not give the same answers, although the subject of the question could not have changed between waves (e.g. father's occupation when respondent was 14 years). This short article presents two approaches to resolving this type of inconsistencies in panel data. The first approach, which we call 'hard adjustment', replaces the original response with a different value. The second approach, 'soft adjustment', provides a recommended target variable, based on a set of assumptions and selection rules, as a suggested replacement to the original variable. I briefly discuss advantages and disadvantages of both approaches with illustrations from the Polish Panel Survey POLPAN (POLPAN.org).

POLPAN data

POLPAN is one of the longest-running panel studies in the world. It started in 1988, before the fall of communism in Poland, and the next waves continued in 1993, 1998, 2003, 2008 and 2013, with preparations underway for 2018 (for an overview, see Slomczynski, Tomescu-Dubrow and Dubrow 2015). Although POLPAN has a panel design, it is also meant to enable analyses of each single wave separately, as cross-sectional surveys. For this reason, some items were repeated in different waves, which in some cases resulted in inconsistent answers.

Hard adjustment

In hard adjustment, the research team identifies inconsistent responses and replaces the original response with a value that can be considered 'true' – or at least substantially closer to the true value. An example of this approach is the practice of correcting answers to questions about year of birth and place of residence on the basis of information taken from the records of the sampling frame. When inconsistent responses cannot be reconciled with the external data source, the adjustment is based on comparing responses across POLPAN waves. This involves taking into account other facts from the respondent's biography related to the subject of the question. Because hard adjustment changes the original data, it must be used with caution, and only when the consistency of responses is necessary for anyone to use the "hard adjusted" variable. Examples of such variables are presented in Table 3, with the percentage of adjustments that we considered necessary.

Table 3. Percentage of Hard Adjusted Cases in POLPAN 1988 – 2013

Variables	1988	1993	1998	2003	2008	2013	Average
Father's year of birth	8	NA	NA	24	24	4	15
Father's education (when R. was 14)	19	NA	NA	19	30	NA	23
Marital status / year of (current) marriage	6	NA	12	10	9	8	9
Respondent's education (level)*	10	12	10	11	10	6	10
Respondent's education (field)*	NA	NA	NA	29	11	10	17

NA: questions were not asked in the respective waves.
* Inconsistencies in reported education do not pertain to the situation of continuing schooling through the life course.
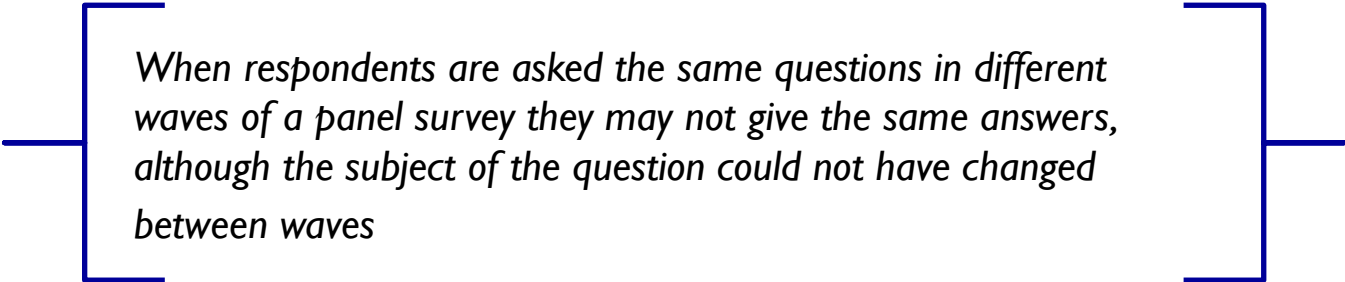
A general requirement in data harmonization is strict documentation of the process of data adjustment (Granda 2015). In practice, the procedure of hard adjustment can be difficult to document strictly. The example of the variable "respondent's education" illustrates this. POLPAN required adjustments if, in a given wave, the education claimed by the respondent turned out to be lower than what they reported it to be in a previous wave. In determining which response is most likely the accurate one, there are two main criteria. One criterion is to take into account other characteristics of the respondent, e.g. the qualification requirements of the job. A second criterion is to refer to the research team's knowledge about the kind of errors in responses on education items that emerge during the survey interview. In such situations, the documentation preserves information about the original education variable that was changed, the code used for transforming it, the criteria the research team considered when making the decision; however, if several criteria jointly influenced this decision, documenting to the extent to which one criterion vs. another drove the harmonization process is difficult.

In hard adjustment, the original responses are replaced by corrected counterparts. Since the original values are preserved, scholars interested in using them, for instance in methodological research, can request the original variables and the survey administrators provide them as attachments to the harmonized data. Additionally, the variables which underwent hard adjustments are flagged by supplementing their descriptive labels with the words "[partly adjusted]."

Soft adjustment

In panel surveys, some inconsistent answers cannot be so resolutely resolved, and forcing a hard adjustment would lead to bias. In these situations, the project's research team can propose a solution to the inconsistency, but only as a recommendation, or "soft adjustment." This soft adjustment can take the form of a "target variable" that does not replace the original variable, but instead supplements it. For data users to determine whether the soft adjustment proposed by the research team is helpful, the procedure for creating the target variable must be fully transparent.

The problem can be illustrated by survey items about work performed in different points of one's occupational career. Certain research questions, such as those about intragenerational occupational mobility, require the sequencing of respondents' careers, which is usually done using occupational codes. For example, to assess whether occupational change (i.e. mobility) occurs, scholars compare respondents' occupations (measured through occupational codes) through time. Selecting occupational codes is not as straightforward as it may seem at first glance. For one, it can happen that panelists report on their occupation differently in one wave than in another, which may result in that person having two different occupational codes although they did not change jobs. It is also possible that panelists have more than one job at a time, and hence they are assigned more than one occupational code.

> *When respondents are asked the same questions in different waves of a panel survey they may not give the same answers, although the subject of the question could not have changed between waves*

In POLPAN, we developed an algorithm that allows the user to select a single occupation when two or more occupations were coded (Sawiński 2016). This algorithm makes it possible to reconstruct the careers of all respondents according to the same rules. These rules were chosen by testing different solutions and they provide a recommendation based on knowledge about the labor market in Poland and the mechanisms that cause errors in surveys. For example, in the case of a person who is running a small farm and at the same time is employed in the factory, we decided that the job outside agriculture is more important for general social standing and should be treated as the main job. However, when data users come to the conclusion that job careers should be reconstructed in a different way, they can always do it by themselves, as the full set of original occupational codes is available in dataset.

Increasing the quality of the data elicited from the respondent

In survey projects, including panel studies, the main focus should be to collect the highest-level quality data possible. In POLPAN, the usefulness of this long-praised strategy is reflected by the

item on respondent's father year of birth (Table 3).

Across POLPAN's waves that inquired about father's year of birth, a given proportion of panelists provide inconsistent information. However, the number of inconsistencies can be considerably reduced when the question is asked in the context of other facts from the respondent's family life. In 2013, the question was asked, for the first time, alongside the question about mother's year of birth, which presumably forced the respondents to think whether their parents were of the same age or whether one parent was older. This made it easier for respondents to figure out the year of birth for their father. Once the context of the item was expanded, only four percent of responses concerning father's year of birth were identified as inconsistent with the responses given in previous waves.

Further Considerations

To prevent data inconsistencies in panel studies, so-called "dependent interviewing" was introduced. Panelists are presented with answers they provided in previous interviews and then are asked whether in the meantime any change occurred (Jäckle, Laurie and Uhrig 2007). This technique improves the consistency of panel data, but it can also introduce another type of distortion, as respondents may be reluctant to provide detailed explanations when they can avoid that. After hearing the description of a job they gave during the interviewer's previous visit, some respondents may be inclined to say that nothing significant changed in their work, although in reality such changes did occur; if these changes would have been reported, the respondent's occupation would have been classified differently (i.e. that person would have received a different occupational code than in the previous wave). In some panel studies, after the introduction of dependent interviewing, the number of panelists changing jobs decreased three times (Perales 2014). Given (a) the need to preserve the continuity of panel data, and (b) the possibility that dependent interviewing could introduce response bias, the question is whether such radical methodological changes should be implemented in long-running panels like POLPAN.

The alternative is to limit the number of errors that stem from survey field-work. In questions about occupations, for example, it appears that occupations have unequal chances to be coded correctly (Conrad, Couper and Sakshaug 2016). In general, research teams flawlessly code jobs such as "barber" or "cleaner." However, in the case of occupations that can be easily confused with each other, such as manager or assistant, there is a need to ask further questions on the details of work, qualification requirements, the number of subordinates, or the type of institution in which the respondent is employed. CAPI software easily allows researchers to add extra questions when certain words or phrases appear in respondent's answer that require deeper explanation. When there is a need for clarification of the data collected in previous waves of the panel, additional questions can be implemented in advance. Possibly, in the future the role of hard and soft adjustments will dwindle, as survey participants themselves will contribute to solving inconsistencies or settling doubts related to their response history.

References

Conrad, F. G., Couper, M. P. and Sakshaug, J. W. (2016). Classifying open-ended reports: Factors affecting the reliability of occupation codes. Journal of Official Statistics 32(1): 75-92.

Granda, P. (2015). Preserving the relationships between harmonized survey datasets and their sources. Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences 1(2): 2-4.

Jäckle, A., Laurie, H. and Uhrig, S. C. N. (2007). The Introduction of Dependent Interviewing on the British Household Panel Survey. ISER Working Paper 2007-07. Colchester: University of Essex.

Perales, F. (2014). How wrong were we? Dependent interviewing, self-reports and measurement error in occupational mobility in panel surveys. Longitudinal and Life Course Studies 4(3): 299-316.

Sawiński, Z. (2016). Occupational careers and job interruptions: On methodological issues of constructing long trajectories. International Journal of Sociology 46(4): 244-263.

Słomczyński, K. M., Tomescu-Dubrow, I. and Dubrow, J. K. (2015) Changes in social structure, class, and stratification: The Polish Panel Survey (POLPAN). Ask Research & Methods 24(1): 19-37.

Zbigniew Sawiński is Professor at the Institute of Philosophy and Sociology of the Polish Academy of Sciences (IFiS PAN), and Professor at the Educational Research Institute in Warsaw (IBE). His research focuses on social stratification, educational inequality, and the methodology of cross-national research. Sawiński is part of POLPAN's research team since the project originated, in the late 1980s. He also belongs to the research teams involved in the European Social Survey and PISA. E-mail: zsawins@ifispan.waw.pl

Harmonization of Ethnic Minority Status in International Survey Projects: The Case of the Russian-speaking Minority in Former-Soviet States

by Olena Oleksiyenko, Graduate School for Social Research, Polish Academy of Sciences

This article focuses on issues of harmonizing information on ethnic minority status as part of a larger project on patterns of electoral and non-electoral political participation in post-soviet states. Specifically, I am interested in differences in political participation between a given country's Russian-speaking minority and the majority population in Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Tajikistan, Uzbekistan and Ukraine.

There is no single international survey project that adequately covers all the former Soviet republics since the Soviet Union's collapse, to current times. Even projects with the broadest country coverage, such as Life in Transition, do not allow for meaningful over-time comparisons. Hence, I selected, for purpose of ex-post harmonization, international projects that measure peoples' electoral and non-electoral participation and ethnic identification in any of the post-soviet countries. Table 4 presents the list of the international survey projects I included, which taken together, span the period 1993- 2015.

Table 4. Quality of Survey Weights by Survey Project

International Survey Project	Sample Size	Time Span	Country coverage
Caucasus Barometer 2009	5 761	2009	AM, AZ, GE
Caucasus Barometer 2010	6 012	2010	AM, AZ, GE
Caucasus Barometer 2011	6 133	2011	AM, AZ, GE
Caucasus Barometer 2012	6 715	2012	AM, AZ, GE
Caucasus Barometer 2013	5 953	2013	AM, AZ, GE
Consolidation of Democracy in Central and Eastern Europe (wave 2)	5 304	1998-2001	BY, EE, LV, LT, UA
European Social Survey (round 2)	4 020	2004-2005	EE,UA
European Social Survey (round 4)	3 506	2009	EE,UA
European Social Survey (round 6)	2 380	2012	EE
European Social Survey (round 7)	4 301	2014-2015	EE
European Values Study (wave 3)	2 018	1999	EE,LV
European Values Study (wave 4)	13 587	2008	AM,AZ,BY,EE,GE,LT, LV,MD,UA
Life in Transition (wave 2)	14 149	2010	AM,AZ,BY,EE,GE,KG,KZ,LT,LV,MD,TJ,UA,UZ
New Baltics Barometer (wave 1)	6 136	1993	EE, LV,LT
New Baltics Barometer (wave 2)	3 339	1995	EE, LV,LT
New Baltics Barometer (wave 3)	2 923	1997	EE, LV,LT
New Baltics Barometer (wave 4)	3 126	2000	EE, LV,LT
New Baltics Barometer (wave 5)	3 068	2001	EE, LV,LT
New Baltics Barometer (wave 6)	3 009	2004	EE, LV,LT
World Values Survey (wave 2)	15 127	1996-1997	AM,AZ,BY,EE,GE,LT, LV,MD,UA
World Values Survey (wave 4)	2 051	2002-2003	KG,MD
World Values Survey (wave 5)	3 546	2008-2009	GE, MD, UA
World Values Survey (wave 6)	12 372	2011-2014	AM,AZ,BY,EE,GE,KG, KZ,UA,UZ
TOTAL	134 536	1993-2015	

Cross-national comparisons of ethnic groups are not as straightforward as it may seem, since in many cases the underlying concept of "minority group" is different in each state. The literature proposes different approaches to increase comparability of the concept. The "absolutist" approach suggests that only one marker of minority status should be taken into account, e.g. citizenship or language. The advantage of such a solution is conceptual clarity, but one can argue that the complexity of the minority status cannot be precisely studied with only one indicator. An alternative is the "relativist" approach to harmonization of items on minority status. This involves cross-classification of different ethnic referents to obtain a single, cross-nationally equivalent score on "ethnic minority status" (Lambert 2005). The problem with the "relativist" approach is the low availability of the same markers across all surveys.

In using survey data to identify Russian minority, there is no clear and simple solution. The literature suggests that, in the former Soviet republics, the Russian-speaking minority consists of not ethnic Russians per se, but specifically people who speak the Russian language and who also have a social identity that is opposed to full integration (Hagendoorn et al 2001). The logic is as follows: During the Soviet period, ethnic Russians were the most mobile group among all nationalities and they settled in the different Soviet republics, in a move that some perceived as an "empire expansion

to the peripheries" (Tishkov et al. 2005). After the Soviet Union fell, the identity of the Russian-speakers living in the former Soviet states seems context-dependent: Russians often identified simultaneously with the country of residence, with Russia, and with the former Soviet Union (Hagendoorn et al 2001). Although national identity (ethnic minority) and language spoken (linguistic minority) are mixed and both are ideologically loaded, the literature suggests that using the term "Russian speaker" is more accurate to describe this minority group than using the term "Russian" (e.g. Laitin 1998).

> *There is no single international survey project that adequately covers all the former Soviet republics from the USSR's collapse to the present.*

Thus, I settled on language as a primary marker for Russian ethnic minority status. However, this solution does not completely solve the problem of concept equivalence. Table 5 presents the variety of survey items concerning ethic identity available in the survey projects I selected for harmonization. All questions can be divided into three main categories: ethnic group (Caucasus Barometer), native language (Consolidation of Democracy in Central and Eastern Europe and Life in Transition), and language used at home (the rest of the surveys). In the European Values Survey the only available information about the language respondent preferred to use is the language of the interview, which can be the native language, the language used at home, or none of these. The issue of using the ethnic group indicator in the case of the Caucasus Barometer is also questionable and needs further investigation.

Table 5. Items Concerning Ethnic Minority Status in International Survey Projects*

International Survey project	Item Wording	Identification of Ethnic Minority Status		
		Language used at home	Ethnic group	Native language
Caucasus Barometer	There are a number of ethnic groups living in this country. Which ethnic group do you consider yourself a part of?		X	
Consolidation of Democracy in Central and Eastern Europe	In what language did/do you communicate with your mother?			X
European Social Survey	What language or languages do you speak most often at home?	X		
European Values Study	Language of the interview	X		X
Life in Transition	What is your mother tongue?			X
World Values Survey	What language do you normally speak at home?	X		
New Baltics Barometer	What language did you speak at home when you were a child?	X		

* For multi-wave surveys, only one item is presented if the wording did not change from wave to wave.

It could be said that harmonizing (Russian) minority status involves a series of tradeoffs between increasing concept equivalence and increasing country and time coverage. To lessen such compromises as much as possible, I intend to construct a set of control variables that will account for inter-project differences in question wording (Slomczynski, Tomescu-Dubrow and Jenkins 2016, eds.). These control indicators can be used in substantive analyses, to partial out item-related methodological variability. A complementary step would be to compare the proportion of Russian-speakers obtained via language/ethnic group identification in any given country with official statistical records on the size of the Russian minority in that country. In any case, I have found that to harmonize even a limited range of ethnic minorities, it is necessary to delve deep into history and the qualitative and quantitative methodological studies about the minorities of interest.

References

Hagendoorn L., Linssen H., Tumanov S. (eds.) 2001. Intergroup Relations in States of the Former Soviet Union: The Perception of Russians, London: Phycology Press.

Laitin D.D. 1998. Identity in formation: The Russian-speaking populations in the near abroad. Ithaca: Cornell University Press.

Lambert P.S. 2005. "Ethnicity and the Comparative Analysis of Contemporary Survey Data" [in:] Hoffmeyer-Zlotnik J.H.P, Harkness J.A.(eds), Methodological Aspects in Cross-National Research, ZUMA, Mannheim

Tishkov V., Zayinchkovskaya Z., Vitkovskaya G. 2005. Migration in the countries of the Former Soviet Union, paper prepared for the Policy Analysis and Research Program of the Global Commission on International Migration, Global Commission on International Migration.

Slomczynski, Kazimierz M, Irina Tomescu-Dubrow, and J. Craig Jenkins, with Marta Kołczyńska, Przemek Powałko, Ilona Wysmułek, Olena Oleksiyenko, Marcin W. Zieliński, and Joshua Dubrow. 2016. Democratic Values and Protest Behavior in Cross-National Perspective. Harmonization of Data from International Survey Projects. Technical Report. Warsaw: IFiS Publishers.

Olena Oleksiyenko is a PhD candidate at the Graduate School for Social Research at IFiS PAN. Currently she works as a research assistant in the Data Harmonization Project and the Polish Panel Survey (POLPAN) 1988-2013 project conducted at IFiS PAN.
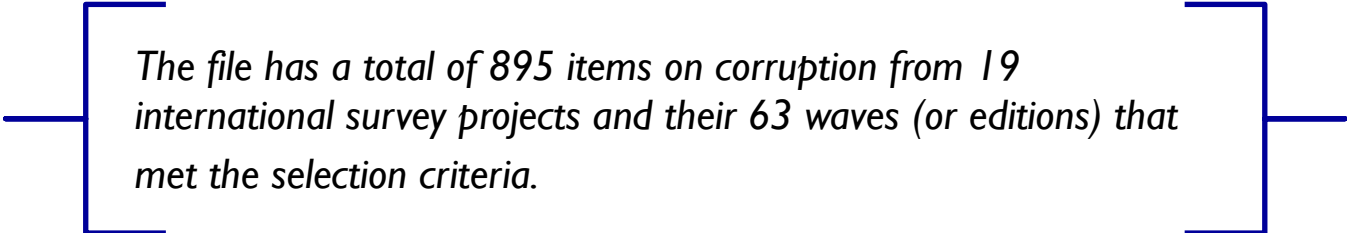
Creating a 'Common File' of Source Variables for Ex-post Harmonization of International Surveys featuring Corruption Items

by Ilona Wysmułek, Graduate School for Social Research, Polish Academy of Sciences

In corruption research, surveys are among the major sources of our knowledge about the subject (Heath, Richards and de Graaf 2016; Karalashvili, Kraay and Murrell 2015). However, there are several methodological challenges to studying cross-national trends in corruption with public

opinion data. Corruption, given its secretive nature, is a phenomenon that is hard to capture in the interview situation. Some respondents are reluctant to answer sensitive questions and some may understand the concept differently than intended by researchers (Azfar and Murrell 2009; Bertrand and Mullainathan 2001). Moreover, international survey projects dealing with corruption continue to face challenges of unequal country representation. Estimation of rare event determinants also remains problematic, given that reported corruption instances are, for most modern democracies, highly infrequent.

To overcome some of these methodological problems, I apply ex-post harmonization of cross-national survey data in corruption research. In my dissertation project, I study corruption perception and individual corruption experience of giving informal payments (as a bribe or a gift) in public schools in Europe. I use cross-national survey data on corruption in public schools in Europe combined with country-level indicators, for example from the World Bank Education Statistics and OECD's Education at a Glance. I follow the Survey Data Recycling (SDR) framework developed by the research team of Kazimierz M. Slomczynski, which provides a blueprint for ex-post survey data harmonization and for integrating surveys and other data sources (please see https://dataharmonization.org/related-projects/corruption-in-education-systems/ for more detailed information about the project).

The file has a total of 895 items on corruption from 19 international survey projects and their 63 waves (or editions) that met the selection criteria.

In this article, I present an overview of existing survey data suitable for research on corruption, and their documentation. The overview of available data is important in the process of conducting research, similar to summarizing relevant literature; however, scholars rarely discuss it explicitly. This information is essential for creating a 'common file' with source variables of interest – itself a key step in ex-post harmonization. The growing availability of survey data during the last 20 years offers rich topic coverage and multiple research opportunities, but also demands knowing where data sources are located, and what issues on corruption they cover. I intend for this article to help others find these sources.

By and large, the leading role in providing public opinion data for corruption analysis is played by two international organizations: Transparency International, with the Global Corruption Barometer survey project based on an adult population sample, and the World Bank, with the World Bank Enterprise Survey based on a firm-level sample (Holmes 2015). Additionally, there are a number of high-quality international public opinion surveys that cover the topic of corruption, along with other items on government perceptions, democratic values, and institutional experiences, however research on corruption rarely uses them.

In reviewing questionnaires and codebooks of international public opinion surveys in search of items on corruption, I started with the pool of 22 international projects in the Data Harmonization Project (dataharmonization.org). Based on the review of literature and consultations with experts, I also went through the collections of data archiving institutions, such as the Inter-university Consortium for Political and Social Research (ICPSR), GESIS Data Archive for the Social Sciences, and ROPER Public Opinion Research Archive. From the pool of possible data sources, I selected the 17 listed in Table 1, based on the criteria listed below.

Selection Criteria

I pooled the datasets and their documentation (master codebooks and questionnaires for project waves) for all relevant survey project waves meeting following criteria:

- Surveys are cross-national and feature European countries;

- Their samples are intended to be representative of the adult population of a given country;

- They are academically driven and available free of charge for non-commercial use (in public domain or upon request);

- Their documentation is written in English;

- They contain at least one item dealing with corruption.

Using Cygwin command-line environment for automatic search, the input files (codebooks, questionnaires and SPSS dictionaries) were checked for lines containing a match with such key words (and their grammar variations) as 'corrupt', 'bribe', 'gift', 'tip', 'favor' ('favour'), 'compensation', 'reward', 'payment', 'present', 'tie', 'connection' and 'network'. If the program detected a matching key word, I explored the neighboring questions and response categories, to get information on the contextual meaning of the key word and to document the availability of the filtering questions (e.g. contact with an institution) and follow-up questions (e.g. amount of bribe). Key words refer to a broad definition of corruption as 'the misuse of public position for private gains' (see Heath, Richards and de Graaf 2016 for corruption definition overview) and capture items that ask generally about corruption and about its specific types, such as for example bribe-giving or nepotism.

Characteristics of Selected Datasets

Table 6 presents basic information on survey projects that met my selection criteria. The table shows the full name of the project, its abbreviation, the number of waves in which the corruption item appeared, year coverage, and availability of documentation sources for survey projects. It also shows the total number of items on corruption available in the survey project.

Table 6. International Surveys featuring Corruption Items: Basic Characteristics and Documentation

Survey name (abbreviation)	Number of waves	Years*	Master Documentation			Number of corruption items
			#Code Books	#Questio-nnaires	#SPSS dictionaries	
Special surveys						
Eurobarometer Corruption Themed 64.3, 68.2, 72.2, 76.1, 79.1 (EB_corr)	5	2005-2013	5	-	5	283
Global Corruption Barometer (GCB)	8	2003-2013	8	-	1	349
International Crime Victims Survey (ICVS)**	4	1992-2005	-	4	1	108
Life in Transition Survey (LITS)	2	2006-2010	-	2	2	43
Large general surveys						
European Social Survey (ESS)**	2	2004-2010	-	2	2	5
European Values Study (EVS)**	3	1990-2008	3	-	1	4
International Social Survey Programme (ISSP)	3	2004-2009	-	3	3	7
World Values Survey (WVS)**	4	1989-2005	4	-	1	5
Smaller surveys: general						
Asia Europe Survey (ASES)	1	2000	1	-	1	3
Comparative Study of Electoral Systems (CSES)	1	2001	1	-	1	1
European Quality of Government Survey (QoG)	2	2010-2013	-	2	2	20
General Eurobarometer (EB)	7	1997-2012	7	-	7	12
International Social Justice Project (ISJP)**	2	1991-1996	2	-	1	4
Pew Global Attitudes Project (PEW)	4	2002-2012	-	4	4	9
Smaller surveys: regional						
Candidate Countries Eurobarometer (CCEB)	2	2003	2	-	2	5
Caucasus Barometer (CB)	4	2009-2012	-	4	4	10
Consolidation of Democracy in Central and Eastern Europe (CDCEE)**	2	1990-1998	1	-	1	11
New Baltic Barometer (NBB)**	6	1993-2004	1	-	1	14
Values and Political Change in Postcommunist Europe (VPCPCE)**	1	1993	-	5	5	2
TOTAL	**63**	**1989-2013**	**35**	**26**	**45**	**895**

* Note that in situations when survey projects do not specify the year of survey wave but do give year brackets (for example WVS 2005-2007), I recorded the year when the questionnaire was first launched (WVS 2005).

** The number of documentation sources and data waves was different in two cases: (1) when the merged data set and documentation for all survey waves was used (see CDCEE and NBB) (2) when the datasets and their documentation were available on the country level only but not on the project wave level (see VPCPCE). In all cases if both master codebook and master questionnaire were available, priority was given to the master questionnaire.

I divided surveys into three categories:

- Survey projects with a block of corruption items ("special surveys"): all selected survey projects and their waves that contain the block of items on corruption (more than ten items per survey

wave);

- Large general survey projects with some corruption items: large survey projects in terms of country and year coverage with less than ten items on corruption;

- Smaller survey projects with some corruption items: smaller survey projects covering minimum three countries and with less than ten items on corruption per survey wave, subdivided into other general survey projects and other regional survey projects.

For each item on corruption, I gathered information on: survey name, survey wave, year of survey wave, name of the variable, question wording, response categories, comments (including mainly information on filtering items and waves repeating the item).

A total of 895 items on corruption are distributed in the 19 international survey projects and their 63 waves (or editions) that met the selection criteria. Generally, the number of corruption items is not equally distributed across different surveys and survey waves. Some selected survey waves contained the block of corruption items, whereas other waves of the same survey project may contain one or two items.

My study revealed that there are several public opinion survey projects that contain a large block of corruption items, and a great amount of survey projects with several corruption-related questions per wave. Despite all the information available, there are rarely attempts to harmonize it and to verify our knowledge using multiple sources of data (see Table 6). Interest in the subject is growing rapidly, especially since 2003, which opens new possibilities on corruption research (see Figure 1).
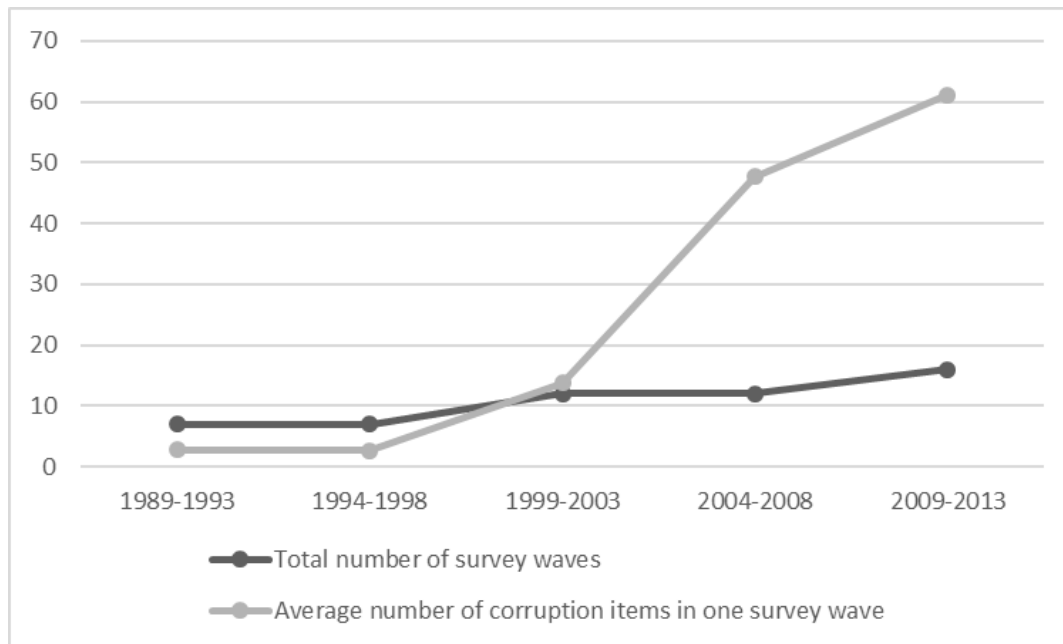

Figure 1. Number of survey project waves and the average number of corruption variables per survey project wave by year.

References

Azfar, Omar & Peter Murrell. (2009) Identifying Reticent Respondents: Assessing the Quality of Survey Data on Corruption and Values. Economic Development & Cultural Change 57(2): 387-411.

Bertrand, M. & S. Mullainathan (2001) Do People Mean What They Say? Implications for Subjective Survey Data. American Economic Review 91: 67-72.

Karalashvili, N., Kraay, A., Murrell, P. (2015) Doing the Survey Two-Step: The Effects of Reticence on Estimates of Corruption in Two-Stage Survey Questions. World Bank Policy Research Working Paper 7276.

Seligson, Mitchell A. (2006). The Measurement and Impact of Corruption Victimization: Survey Evidence from Latin America. World Development 34(2): 381-404.

van Kesteren, J. N. (2007) Integrated Database from the Crime Victim Surveys (ICVS) 1989-2005, Tilburg, INTERVICT.

Heath, Anthony F., Lindsay Richards, and Nan Dirk de Graaf. (2016) Explaining Corruption in the Developed World: The Potential of Sociological Approaches. Annual Review of Sociology 42, no. 1: 51–79.

Ilona Wysmułek is a research assistant at the Research Group on Comparative Analysis of Social Inequality and a PhD candidate in sociology at the Institute of Philosophy and Sociology, Polish Academy of Sciences. Her dissertation research is supported by the Mobility Grant of the Ministry of Science and Higher Education of Poland funding a half-year visiting fellowship (winter-spring 2016) at the Mershon Center for International Security Studies, The Ohio State University and the University of Chicago (National Opinion Research Center) (1292/MOB/IV/2015/0).

Conference presentations

Conference Presentation of "Effects of Inequality on Attending a Political Demonstration: Harmonized Survey Data on 142 Countries, 1966-2014" at IPSA in Poznan, Poland

by Joshua Dubrow, Irina Tomescu-Dubrow, Kazimierz M. Slomczynski, Marta Kołczyńska, and J. Craig Jenkins

'Attending a political demonstration' is among the most common and widely used items in cross-national social science surveys, past and present. Recent research explores whether and how characteristics of the countries that people live in influence this meaningful democratic action. As members of the Harmonization Project team, of the Polish Academy of Sciences and The Ohio State University, we presented a paper on this topic at the 24th World Congress of the International Political Science Association (IPSA), in Poznań, Poland, July 23-28 2016, in the session, "Patterns of Political Participation in a Global Comparison," convened by Christian W. Haerpfer. In this

presentation, we focused on the impact of economic inequality on 'attending a presentation,' when other important features, degree of democracy and economic development especially, are accounted for. Based on grievance theory, resource theory, relative power theory and opportunity structure theory, the hypothesized expectations for macro variables are as follows:

Table 7. Hypothesized Relationship between Macro Variables and 'Attending a Demonstration'

Theory	Hypothesized Relationship		
	Level of democracy	Economic development	Economic inequality
Grievance	-	-	+
Resource	+	+	-
Relative power	+	Interaction with inequality	-
Opportunity structure	+	+	Interaction with development

So far, survey-based studies are mostly region-specific, with a great emphasis on Europe, and have produced mixed results regarding the relevance of macro characteristics, economic inequality included. Most studies found that, for disadvantaged groups, macro-level economic inequality depresses electoral and non-electoral political participation (Dubrow et al 2008; Solt 2015). Some found that it can slightly increase political participation (Karakoc 2013) whereas others found no relationship (Dalton 2010).

We extended the scope of extant analyses by using the Harmonization dataset, developed in the project "Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability and Multi-Level Modeling in Cross-National Perspective" financed by the Polish National Science Centre (2012/06/M/HS6/00322). This dataset contains harmonized individual-level measures of political behavior, social attitudes and socio-demographics constructed with information pooled from 22 international projects that cover a total of 142 countries and territories surveyed at various points from 1966 to 2014; it also contains macro-level characteristics from non-survey sources, and two types of methodological control indicators: control variables for the quality of the source surveys, and controls of the harmonization process (see Tomescu-Dubrow & Slomczynski 2016; Powalko & Kolczynska 2016).

The outcome of interest in our analyses was 'attending a political demonstration.' While 1,127 surveys from the 22 international projects in the Harmonization dataset asked about this political act, the wording of the questions varied. We identified 31 versions of how respondents were asked about attending demonstrations, with the most important difference pertaining to the frame of time in the question, i.e. whether participation occurred with the last 12 months or one year; the last two years, 3, 4, 5, 8, 10 years, or ever (Table 8).

We analyzed two forms of the dependent variable. In one set of regression models, the harmonized dependent variable measured whether respondents participated in demonstrations within the last 12 months or one year (yes = 1, else =0). In the second set of regression models, the harmonized dependent variable captured respondents' participation in demonstrations in the last 8, 10 years, or ever (yes = 1, else = 0).

Table 8. The time frame of respondent behavior: years in the survey item

	1	2-3	4-5	8-10	Ever
Number of Surveys	334	62	42	25	664
Number of Respondents	525,857	81,989	56,126	29,220	832,108

At the individual level, our models contained the following harmonized variables: Gender, Age, Rural locality (yes = 1, otherwise = 0), Education, ISCED, and Interest in politics (scale 0-4). Country-year variables included the Freedom House Democracy Index (civil liberties & political rights, added & reversed; scale 0-12), GDP per capita PPP US $ (World Bank) and the Gini Index (SWIID, Standardized World Income Inequality Database, Solt 2015). In some models we also included country-level variables: regions or micro-regions.

We estimated two types of multilevel logistic regression models: (a) three-level models, where people are nested in country-years and country-years are nested in countries; and (b) two-level models, where people are nested in countries and we controlled for time.

The effects of individual-level determinants conformed to the theorizing developed since the 1960s, irrespective of whether the dependent variable was demonstrations in the last year, or having demonstrated ever, and irrespective of the modeling choice (3-level or 2-level). Regarding macro-variables, however, we did not find a clear pattern of effects that would support the major theories invoked. Under the more restrictive models (3-levels), the effects of the Freedom House Index, GDP per capita, and the Gini Index are usually insignificant. Under the more relaxed models (people nested in countries and controlling for time) the results vary across world regions, and are different for different formulations of the dependent variable.

At the moment, we conclude that methodological artifacts, including formulation of the dependent variable, what regions are included, and variations in model restrictions, impacts the empirical relationship between economic inequality and other macro-level factors on 'attending a demonstration,' as it is currently formulated in the empirical literature. Moreover, there does not seem to be a clear worldwide pattern of which theory – grievance, resource, relative power, and opportunity structure – best explains this relationship. Clearing away methodological artifacts should be the first priority of future research on this topic.

References

Dalton, Russell, Alix Van Sickle, and Steven Weldon. 2010. "The Individual–Institutional Nexus of Protest Behaviour." British Journal of Political Science 40(1): 51 – 73.

Dubrow, Joshua Kjerulf, Kazimierz M. Slomczynski and Irina Tomescu-Dubrow. 2008. "Effects of Democracy and Inequality on Soft Political-Protest in Europe: Exploring the European Social Survey Data." International Journal of Sociology 38(3): 36 – 51.

Karakoc, Ekrem. 2013. "Economic inequality and its asymmetric effect on civic engagement: evidence from post-communist countries." European Political Science Review 5(2): 197 – 223.

Powalko, Przemek and Marta Kolczynska. 2016. "Working with Data in the Cross-national Survey Harmonization Project: Outline of Programming Decisions." International Journal of Sociology 46(1): 73 – 80.

Solt, Frederick. 2015. "Economic Inequality and Nonviolent Protest." Social Science Quarterly 96(5): 1314 – 1327.

Tomescu-Dubrow, Irina and Kazimierz M. Slomczynski. 2016. "Harmonization of Cross-national Survey Projects on Political Behavior." International Journal of Sociology 46(1): 58 – 72.

Conference Presentation of "Optimizing Questionnaire Design in Cross-National and Cross-Cultural Surveys" at QDET2 in Miami, Florida

by Tom W. Smith, NORC at the University of Chicago

Tom W. Smith, of NORC at the University of Chicago, presented the invited paper, "Optimizing Questionnaire Design in Cross-National and Cross-Cultural Surveys," at the Second International Conference on Questionnaire Design, Development, Evaluation, and Testing (QDET2 - amstat.org/meetings/qdet2) November 9–13, 2016 in Miami Florida.

In his presentation, Professor Smith argued that questionnaire development in cross-national and cross-cultural surveys starts with the same challenges that exist for single monolingual and mono-cultural surveys. The individual measures need to be reliable and valid and need to function well among the target population (e.g., understandable, answerable, consistently understood across respondents). Furthermore, when multi-item measures are to be used, the scales must measure well the concepts of interest (e.g., high Cronbach's alpha, strong and appropriate factor loadings). But then for cross-national/cultural surveys, the process becomes more complicated and challenging. The individual measures and scales need to achieve a high level of functional equivalence across surveys and minimize comparison error. Achieving this is not easy. First, one needs to optimize wording comparability by applying rigorous translation procedures such as the TRAPD model (Translation, Review, Adjudication, Pretesting, and Documentation) and the analysis of results for differences due to measurement variation resulting from errant or at least suboptimal translations. Second, one needs to consider whether there are structural or cultural differences (besides language) that undermine the comparability of items. Third, when multi-item scales are involved, one needs to test for equivalence (e.g., by using IRT or CFA procedures). Then if, as is often the case, equivalence is not established, one needs to identify the source of the nonequivalence. There are numerous possible reasons for this such as (1) poorly operating scales, even within countries; (2) poor translations; (3) structurally or culturally based differences in specific items; or (4) true and meaningful differences in the configuration of values and attitudes related to a concept due to substantive variation in how societies assess the concept. The source for the non-equivalency must be ascertained, since the steps that need to be taken to deal with it depend on its source and nature.

Professor Smith discussed the operational and methodological techniques that have been developed to deal with this challenge. These include (a) best practices to carry out and evaluate translations; (b)

design-to-archiving metadata protocols to maximize the careful development and documentation of the surveys (e.g., the Questionnaire Design Documentation Tool and Translation Management Tool); (c) coordination across data collection to fully implement and enforce input harmonization from the conceptual level down to the nitty-gritty of coding, cleaning, and other details of surveys; (d) items development aids such as SPQ2 and QUAID; and (e) scale evaluation procedures such as IRT, CFA, and Bayesian approaches.

The author provided detailed examples of specific innovative approaches used by major international surveys, including the European Social Survey (ESS), International Social Survey Program (ISSP), and the Survey of Health, Ageing, and Retirement in Europe (SHARE).

Conference Session on "Harmonization, Data Documentation and Dissemination" at the 2nd 3MC Conference

by the Harmonization Project team

The Harmonization Project team presented their work at the Second International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts (3MC 2016), which took place in Chicago, IL, from 25 to 29 July, 2016. The special session "Harmonization, Data Documentation and Dissemination" consisted of four invited papers preceded by a general introduction to the Project, and was chaired by Peter Granda from the University of Michigan. The four presentations were devoted to different aspects of survey data quality based on a collections of 1721 national surveys from 22 survey projects analyzed in the Harmonization Project, funded by the (Polish) National Science Centre (grant number 2012/06/M/HS6/00322). Although data quality was discussed primarily from the point of view of survey data harmonization, some of the theoretical considerations and proposed solutions are applicable to any research using survey data.

The first paper, by Marcin W. Zieliński and Przemek Powałko, focused on statistical weights as a source of information about the quality of the sample. Weights are a common tool for improving the representativeness of survey data, and their availability in survey datasets has been increasing in the last decades. However, weights are often not well described and documented, and sometimes have technical errors that can influence the results of statistical analyses. The paper proposed four indicators of the quality of weights, and concluded with a discussion of possible approaches to harmonizing weights in the Harmonization Project.

The second paper, by Olena Oleksiyenko, Ilona Wysmułek and Anastas Vangeli, dealt with the issue of processing errors, or the discrepancies between the data and the documentation. Processing errors occur after the data is collected, during data coding and recoding, and generation of survey documentation, such as codebooks. Applying a new typology of processing errors, the paper discussed the prevalence of processing errors in cross-national surveys, and their severity and potential risks for researchers. The observed trend is that the quality of data processing is decreasing over time. The proposed typology can serve as a check list of potential problems at the stage of data processing.

In the next presentation, Marta Kołczyńska and Matthew Schoene talked about data quality as reflected in survey documentation. Based on a review of available descriptions of the key stages of the survey process – sampling, questionnaire translation, pretesting, response, and post-hoc fieldwork control – the authors created a set of quality indicators that they used to evaluate survey quality. Results showed that, despite major differences across surveys, the quality of surveys as reflected in their documentation tends to improve over time, both within projects from wave to wave, and overall.

Finally, the last paper by Marta Kołczyńska and Kazimierz M. Slomczynski, stepped away from direct evaluations of data quality to focus on item-specific metadata that capture methodological differences between equivalent items in different surveys that may affect the harmonized (target) variable. The authors identify a number of features of questionnaire items that need to be taken into account in the process of variable harmonization, including item wording, the type of response scale, and item non-response. They discussed ways of recording these differences as harmonization control variables using the example of two target variables: trust in the national parliament and participation in demonstrations.

Presentation slides are available on the dataharmonization.org website.

List of Presentations:

Przemek Powałko. "Introduction to the Harmonization Project"

Marcin W. Zieliński and Przemek Powałko. "The past, present and future of statistical weights in cross-national Surveys: Implications for survey data harmonization."

Olena Oleksiyenko, Ilona Wysmułek, and Anastas Vangeli. "Identification of processing errors in cross-national surveys."

Marta Kołczyńska and Matthew Schoene. "Survey data harmonization and the quality of data documentation in cross-national surveys."

Marta Kołczyńska and Kazimierz M. Slomczynski. "Item-specific metadata in ex-post harmonization of international survey projects."

Conference Papers on Harmonization Presented at International Social Survey Programme General Assembly in Lithuania

by the Harmonization Project team

The Harmonization Project team, of the Polish Academy of Sciences and The Ohio State University, presented two papers at the International Social Survey Programme (ISSP) General Assembly on May 1, 2016 in Kaunas, Lithuania. The first paper was by Olena Oleksiyenko and Ilona Wysmułek,

"Quality of data processing in International Social Survey Programme (1985-2013): evaluation and potential threats." They discussed results of the data processing quality analyses of International Social Survey Programme (1985-2013). The presentation consisted of examples of processing errors encountered in the master documentation of ISSP. Additionally, since ISSP is included in the Data Harmonization Project, the data processing quality index for ISSP was presented in comparison with other well-known survey projects, such as European Social Survey and World Values Survey. In line with the definitions of the data processing quality, which refer to the consistency and transparency of the data, the authors concentrated on the consistency of the data and documentation for the selected socio-demographic, behavioral and attitudinal variables. They selected key demographic variables commonly used by researchers and two substantive variables. The selected target variables were gender, age, birth year, education levels, schooling years, trust in parliament, and participation in demonstrations. As a result of the analyses, they created a typology of processing errors that can be used at the stage of data production to avoid inconsistencies and increase the quality of the released data in future.

The second paper was by Marcin W. Zieliński, "Diversity of statistical weights in the International Social Survey Programme 1985-2013: Methodological implications for survey data comparability and cross-national analysis: trade-off between standardization and keeping local context." The author focused on two dimensions of data weighting in ISSP. The first one was diversity of types of weights in ISSP. In general there are two types of weights, each of them having different aims and different impact on the data: design and post-stratification. ISSP provides only one variable containing weights, which means that, in some cases, national teams have to combine design with post-stratification factors. The second deals with diversity of the quality of existing weights in ISSP. In this part, the author focused on the assessment of the formal quality of provided weights discussing some of their properties and impact on the data. The author made some recommendations for national ISSP members and Methodology Committee in order to improve the quality of weighting procedures in the future work on methodological growth of the ISSP project.

News

Use It for Good: IPUMS and Data Harmonization

by Catherine A. Fitch, Associate Director, Minnesota Population Center & Co-Director, Minnesota Census Research Data Center

In collaboration with 105 national statistical agencies, nine national archives, and three genealogical organizations, IPUMS has created the world's largest accessible database of census microdata. IPUMS includes almost a billion records from U.S. censuses from 1790 to the present and over a billion records

from international censuses of over 100 countries. We have also harmonized survey data with over 30,000 integrated variables and 150 million records, including the Current Population Survey, the American Community Survey, the National Health Interview Survey, the Demographic and Health Surveys, and an expanding collection of labor force, health, and education surveys. In total, IPUMS currently disseminates integrated microdata describing 1.4 billion individuals drawn from over 750 censuses and surveys.

In addition to census and survey microdata, IPUMS integrates and disseminates the nation's most comprehensive database of area-level census data and electronic boundaries describing census geography from 1790 to the present. The National Historical Geographic Information System (NHGIS) includes 366 billion data points and 28 million map polygons describing U.S. Census geographic units. As part of the TerraPop project, we now archive and disseminate a third class of data: raster data derived from satellite imagery, climate models, and other sources.

Our signature activity is harmonizing variable codes and documentation to be fully consistent across datasets. This work rests on an extensive technical infrastructure developed over more than two decades, including the first structured metadata system for integrating disparate datasets. By using a data warehousing approach, we extract, transform, and load data from heterogeneous sources into a single view schema so data from different sources become compatible. The large-scale data integration from IPUMS makes thousands of population datasets interoperable. We have created software for consistency checking, automated data cleaning and editing, sampling, disclosure control, database harmonization, metadata creation, and parsing. Our data projects exploit machine-learning technology for automated string classification and record linkage and employ parallel processing to manipulate large datasets in our high-performance computing environment.

IPUMS designed the first web-based data dissemination system that automatically pools datasets from multiple sources and rectangularizes hierarchical files for ease of analysis. Today, our fourth-generation dissemination systems provide powerful tools for data discovery, metadata browsing, data manipulation, and online analysis. Our current data access software, based on entirely new architecture, was developed using the Ruby on Rails web application framework. Users can download data for analysis using their own statistical software or mapping application or analyze data online using our web-based data analysis tool.

Metadata is at the core of IPUMS harmonization and dissemination. The IPUMS metadata specification has five major components: (1) source data dictionaries documenting the raw data files from data producers; (2) variable translation tables, providing most of the variable-level information required to create the database, including IPUMS-format variable labels, value labels, and codes, as well as dataset-specific information on universe, location of source variable, and all information required to harmonize codes across datasets; (3) variable descriptions, containing information about the meaning of each variable and comparability across datasets; (4) control files identifying the symbolic location of each piece of data, metadata, and software needed by the system to control numerous options for the creation and display of each dataset and variable; and (5) ancillary documentation such as enumeration instructions and forms, sample designs, and other material related to the particular census or sample. We deliver customized metadata for each data extract in Data Documentation Initiative (DDI Codebook 2.5) format, the leading metadata specification for the social and behavioral sciences.

Without IPUMS, much public-use microdata would be inaccessible to researchers. There would

be little machine-processable metadata describing the data; rather, most microdata documentation would be locked away in unstandardized PDF codebooks readable only by humans. The available datasets would not be interoperable across time or space, forcing each investigator to expend redundant effort developing ad hoc crosswalks for particular research projects.

Data sharing is central to our mission; effective dissemination is essential if the data are to be widely used to produce new discoveries. To connect with new and experienced data users, we provide workshops locally in Minneapolis and nationally at selected conferences. We also reach out to users with data release information and novel uses of the data through our social media channels (Facebook and Twitter). We recently launched a blog, Use It for Good (blog.popdata.org), which explores more detailed research using our data projects.

CharmStats Pro: A New Harmonization Software For Large-Scale Research

by Kristi Winters, GESIS

GESIS – Leibniz Institute for the Social Sciences has just released a new version of CharmStats, the open-source and free software package for variable harmonization.
CharmStats Pro was designed to meet the needs of large scale studies and research teams. The software reduces the time and effort spent harmonizing and recoding variables for documentation and for use in statistical analysis. It was designed for researchers who want to quickly and easily create variable harmonization recoding syntax in SPSS, Stata, SAS and Mplus. CharmStats Pro can be used by solo researchers or by research teams to centralize, document and process the harmonization documentation process because it allows multiple users to work on the same projects on a shared database.

To facilitate a cooperative digital environment, CharmStats Pro includes a Communications Suite featuring an internal email system and task manager to help team leaders organize their work. The MessageManager centralizes all the relevant communications and tasks associated with creating, editing and publishing harmonization projects within the software. It works similar to an eMail client and is designed to allow users to send messages to other accounts within the database. CharmStats Pro also includes a connected TaskManager feature to help users organize their work. Rather than team members communicating with an external email service and maintaining their 'to do' lists there, all relevant communications can take place within CharmStats Pro and are maintained on the software's database. This ability to centralize task assignments, questions and advice between team members, and convert messages into actionable tasks enhances a team's ability to document its own work in a way is unlike other software packages.

CharmStats Pro streamlines the variable harmonization process in several ways. It allows for quick importation of variable level metadata while facilitating the inclusion of question level and study level metadata. New to CharmStats Pro is importing variable and measurement metadata using DDI 3.1 xml files. All the relevant metadata can be included in your work and project reports without any additional work.

Bibliographic information on the literature relevant to your variable harmonization can now be

added to your CharmStats database. Reference can be connected at each step in the workflow and to variables and measurements. Stored references can be searched and re-imported into other projects, meaning that once a member of a team enters the literature information all other members will be able to find it and use it themselves.

Comments are a main source of documenting coding decisions in CharmStats. They allow researchers to include important notes, comment on the problems they faced or provide a rationale for coding decisions. The improved CharmStats Pro comment editor allows for more specific, and therefore finer granulated, documentation. The Comment feature in CharmStats Pro allows for more than one comment per object, unlike in QuickCharmStats.

There is a new, alternate way to document data recoding syntax that allows researchers who do not use mapping-based variable harmonization to document their syntax codes. In addition to the standard report CS Pro offers the Advanced Report feature that generates codebooks: the creation of a mass report to display information from multiple projects simultaneously.

We have also made it easier to manage variables in the database. It is possible to generate an overview of all the variables and measurements contained within the CharmStats Pro database or contained within a specific harmonization project. For users who want to compare variables before using them we have created a new Compare Variables function. Want to check variable, question and study level information? The Compare Metadata window allows you to review the metadata for each of the variables you are comparing.

We have provided more ways to document the work done within CharmStats to facilitate team planning and progress reports.It gives research teams more flexibility to organize harmonization projects in creating project folders and the ability to move projects between folders. Within the CharmStats Pro database an activity log records and stores any changes to projects. This list is updated, processed and a report is created each time a user logs into the program. Any changes to variables or measurements and their metadata (question and study level information) are documented in the Database reports. When a user logs in, the system checks if any information has been added or changed since the last time the user connected to the CharmStats Pro database.

The report feature creates .html files based on the information in your project. Use this feature to save and display documentation and bibliographical information, notes containing the coding explanations, as well as the harmonization syntax. Share it with others at anytime, anywhere, or to post it online as a reference. The graph feature provides a visual image of the harmonization of the source to target variable. Graphs can be saved as .jpegs for use in presentations or included in the documentation. A new feature allows users to insert their definitions and comments into the graph display area and a "Prepare Report" button has been added. This feature creates connections between your graphs and project reports by saving a graph for use in a report. Change the colors in the graph to match your institution, university or research team logo.

If you would like to learn more about the ways CharmStats Pro can help reduce the workload of variable harmonization, please get in contact with Dr. Kristi Winters at GESIS (kristi.winters@gesis.org) or download CharmStats Pro at http://www.gesis.org/en/services/data-analysis/data-harmonization/

New publications

Democratic Values and Protest Behavior: Harmonization of Data from International Survey Projects

The Harmonization Project has released their first book:

"Democratic Values and Protest Behavior: Harmonization of Data from International Survey Projects" by Kazimierz M. Słomczyński, Irina Tomescu-Dubrow, and J. Craig Jenkins, with Marta Kołczyńska, Przemek Powałko, Ilona Wysmułek, Olena Oleksiyanko, Marcin W. Zieliński and Joshua K. Dubrow. 2016. Warsaw: IFiS Publishers.
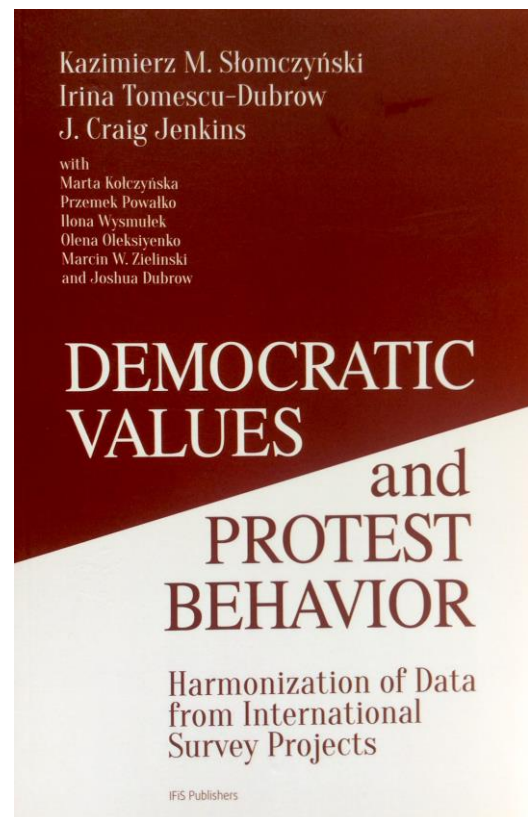
This book is available on dataharmonization.org free to download and read.

About the Book

Across the world, mass political protest has shaped the course of modern history. Building on decades of theory, we hypothesize that the extent and intensity of political protest is a function of micro-level democratic values and socio-demographics, country-level economic development and democratic practices, and the discrepancy (i.e. cross-level interaction) between a country's democratic practices and peoples' trust in key democratic institutions – that is, political parties, the justice system, and parliament.

This book is a Technical Report on the logic of, and methodology for, creating a multi-year multi-country database needed for comparative research on political protest. It concerns both the selection and ex-post harmonization of survey information and the manner in which the multilevel structured data can be used in substantive analyses.

The database we created contains information on more than two million people from 142 countries or territories, interviewed between the 1960s and 2013. It stores individual-level variables from 1,721 national surveys stemming from 22 well-known international survey projects, including the European Social Survey, the International Social Survey Programme, and the World Values Survey. We constructed comparable measures of peoples' participation in demonstrations and signing petitions, their democratic values and socio-demographic characteristics. We complemented the harmonized individual-level data with macro-level measures of democracy, economic performance, and income inequality

gathered from external sources. In the process, we pulled together three strands of survey methodology – on data quality, ex-post harmonization, and multilevel modeling.

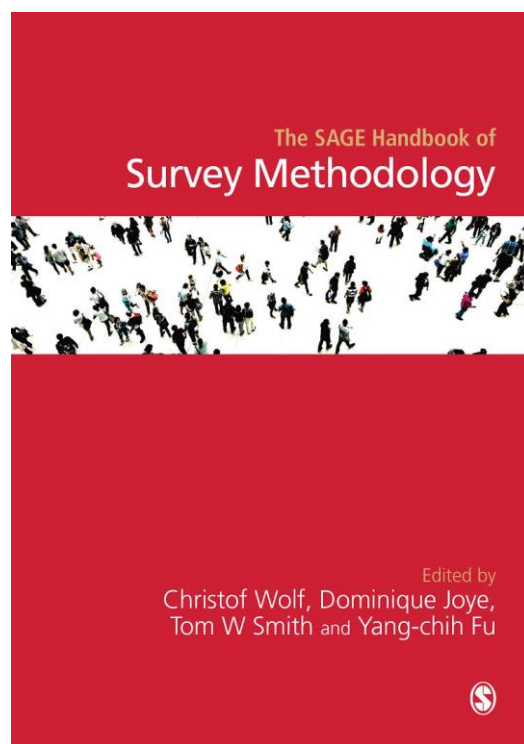# SAGE Handbook of Survey Methodology: An Essential Tool for Scholars within the Social Sciences

London, UK (2016). Exploring both the increasingly scientific endeavor of surveys and their growing complexity, ***The SAGE Handbook of Survey Methodology*** (740 pages; ISBN: 9781446282663) provides a comprehensive guide to conducting survey research studies.

As different data collection modes and information sources are combined and Survey Methodology becomes a more structured field of research, the Handbook provides an essential, up-to-date tool for researchers. With a team of international experts, the Handbook looks at local and national specificities, as well as problems of cross-national, comparative survey research. The chapters are organized into nine sections, each of which represents a stage in the survey life-cycle:

- Basic Principles
- Surveys and Societies
- Planning a Survey
- Measurement
- Sampling
- Data Collection
- Preparing Data for Use
- Assessing and Improving Data Quality

As the editors, Christof Wolf, Dominique Joye, Tom W. Smith & Yang-chih Fu, explain:

"There are a lot of reasons to publish a new handbook of survey methodology. Above all, the field of survey methodology is changing quickly in the era of the Internet and globalization. Survey methodology could also be seen as a bridge between disciplines, resting on the shared methodological preoccupations between specialists of very different fields. These are some of the challenges we are addressing here."

Harmonization would like to hear from you!

We created this *Newsletter* to share news and help build a growing community of those who are interested in harmonizing social survey data. We invite you to contribute to this Newsletter. Here's how:

1. Send us content!

- Send us your announcements (100 words max.), conference and workshop summaries (500 words max.), and new publications (250 words max.) that center on survey data harmonization in the social sciences;
- Send us your short research notes and articles (500 – 1000 words) on survey data harmonization in the social sciences. We are especially interested in advancing the methodology of survey data harmonization. If we have any questions or comments about your items, we will work with you to shape them for this *Newsletter*.

    Send it to: Joshua Kjerulf Dubrow, dubrow.2@osu.edu.

2. Tell your colleagues!

To help build a community, this *Newsletter* is open access. We encourage you to share it in an email, blog or social media (Facebook, Twitter, Google+, and so on).

Support

This newsletter is a production of Cross-national Studies: Interdisciplinary Research and Training Program, of The Ohio State University (OSU) and the Polish Academy of Sciences (PAN). The catalyst for the newsletter is our ongoing project, "Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling" (hereafter, Harmonization Project). Financed by the Polish National Science Centre in the framework of the Harmonia grant competition (2012/06/M/HS6/00322), the Harmonization Project joins the Institute of Philosophy and Sociology PAN and the OSU Mershon Center for International Security Studies in creating comparable measurements of political protest, social values, and demographics using information from well-known international survey projects. The team includes: Kazimierz M. Slomczynski (PI), J. Craig Jenkins (PI), Irina Tomescu-Dubrow, Joshua Kjerulf Dubrow, Przemek Powałko, Marcin W. Zieliński, and research assistants: Marta Kołczyńska, Matthew Schoene, Ilona Wysmułek, Olena Oleksiyenko, Anastas Vangeli, and Anna Franczak. For more information, please visit dataharmonization.org.

Copyright Information