



CONSIRT

Cross-National Studies: Interdisciplinary Research and Training Program

Working Papers Series

CONSIRT Labs: Methodology of Survey Data Harmonization

Survey Data Harmonization and the Quality of Data Documentation in Cross-National Surveys

Matthew Schoene, The Ohio State University

schoene.3@osu.edu

Marta Kołczyńska, The Ohio State University and Polish Academy of
Sciences

kolczynska.1@osu.edu

October 2014

About CONSIRT Working Papers

Papers in the CONSIRT Working Papers Series are pre-publication versions of the author(s)' academic work. Working Papers posted on consirt.osu.edu are in progress, under submission, or forthcoming. Working Papers are produced on this site by the author(s) for the consumption of the international scientific community and others interested in this work. CONSIRT Working Papers are not subject to double-blind peer review. CONSIRT reviewed the Working Paper submission for general suitability, but the quality of the Working Paper itself is the responsibility of the author(s). The form and content of papers are the responsibility of individual authors. The cover sheet is standard across all papers (subject to change), but the format of the rest of the work is not standardized. Comments on papers or questions about their content should be sent directly to the author, at his or her email address.

Citation of CONSIRT Working Papers

Working Papers may be cited without seeking prior permission from the author. The proper form for citing CONSIRT Working Papers is:

Author(s). Year. "Title." *CONSIRT Working Papers Series* at consirt.osu.edu.

Once a CONSIRT Working Paper has been published elsewhere, it is customary that it be cited in its final, published version, rather than in its Working Paper version.

Copyright Statement

Copyright to CONSIRT Working Papers remains with the authors. Posting a Working Paper on consirt.osu.edu does not preclude simultaneous or subsequent publication elsewhere, including other Working Papers series. It is the current policy of CONSIRT to maintain all Working Papers posted on the CONSIRT website unless otherwise notified by the author(s). Users of this website may consume papers for their own personal use, but downloading of papers for any other activity, including but not limited to electronic reposting to the internet in any way, may only be done with the written consent of the authors. It is the author's responsibility to know if copyright has been transferred upon publication elsewhere and to ask CONSIRT to have it removed from the site, if necessary.

Survey Data Harmonization and the Quality of Data Documentation in Cross-National Surveys

Matthew Schoene, The Ohio State University

Marta Kolczyńska, The Ohio State University and Polish Academy of Sciences

Abstract

Many social phenomena bear the imprint of regional and global social, economic, and political processes, and therefore should be studied in a comparative framework. However, inequalities in survey coverage across various regions and the lack of uniform data collection and documentation standards in cross-national survey projects pose a serious limitation to comparative research. While previous efforts to advance cross-national research usually produce a new dataset, we argue that the answer to these problems is actually better integration of existing data. This paper focuses on the assessment of the quality of data documentation, an important, although often neglected, element of any data analysis that is especially crucial for data harmonization projects. We assessed the quality of surveys based on information provided in survey documentation: questionnaire pre-testing, translation method, sampling, presence of fieldwork control, as well as unit and item non-response. Finally, we discuss incorporating documentation measures into substantive analyses, as well as the potential for standardization of survey documentation and the survey process itself.

Acknowledgements

Paper prepared for the 67th Annual Conference of the World Association for Public Opinion Research (WAPOR), 4-6 September 2014, Nice. Previous versions of this paper were delivered at the Conference and Workshop on Harmonization in Warsaw, December 18-21, 2013, and at the 12th International Workshop on Comparative Survey Design and Implementation, Bethesda, MD, March 27-29, 2014. We wish to thank Kazimierz M. Słomczyński, Irina Tomescu-Dubrow, and Craig J. Jenkins for their comments and help throughout the writing process, and Carolyn Morgan for editorial support. Work on this paper was carried out within the project “New Approach to Analyses of the Relationship between Democracy and Trust: Comparing European Countries Using Quantitative and Qualitative Methodology” funded by the Mobilność Plus program of the Ministry of Science and Higher Education of the Republic of Poland (number 921/MOB/2012/0), and supported by the grant “Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective” from the (Polish) National Science Centre (number 2012/06/M/HS6/00322; dataharmonization.org). All data and calculations are available upon request.

Many social phenomena bear the imprint of regional and global social, economic, demographic and political processes, and therefore should be studied in a comparative framework. While it would be ideal to have data from all countries of interest collected according to uniformly high standards, this is rarely, if ever, the case. Cross national researchers always face two problems: first, is the area of interest well-covered by existing surveys? Second, exactly how good are the data and accompanying documentation? Of course, the data collection process is much more challenging in some contexts compared to others; this is reflected in the availability of data from certain regions. We observe a clear divide between well surveyed areas (such as the United States, Europe, Canada, Australia and Japan, etc.) and poorly surveyed regions (such as Sub-Saharan Africa or the Middle East). Even within a well sampled region like Europe, the historic East-West divide results in unequal country coverage. To further complicate matters, we have found very few empirical analyses of the quality of existing cross-national data and documentation, outside of anecdotes shared among researchers.

We argue that an answer to these data challenges is the better integration of existing surveys. Most survey programs are regional in scope, therefore allowing only comparative work within the specific regions; this presents a serious limitation to cross-national research. In other words, because of the regional specificity, it is easy to compare Germany to France, but much harder to compare Germany to Brazil. To address this limitation, we argue in favor of ex-post harmonization of multiple datasets. In this paper, we report on a harmonization project that addresses the above limitations to cross-national research: "Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective." This project takes data from 22 cross-national survey projects from 1983-2012 to produce an integrated dataset of political attitudes and participation, supplemented by a linked dataset of country-level indicators.

Before we can start harmonizing data we must better understand the country coverage and quality of the data documentation of the to-be harmonized datasets that form the building blocks of our harmonized data file. First, we discuss the benefits to survey data harmonization, and then examine the scope of data coverage offered by performing such work. We then turn our attention to quality of data documentation. Survey documentation and its availability often vary substantially from project to project; some painstakingly document every stage of the process while others offering little more than a short description of the sampling frame. Trained researchers are well aware of how errors can damage the dataset and bias results, so there is a real demand to see the process, not just the final product. Unfortunately, these documents are sometimes not available. We searched the publically available documentation for evidence of high quality survey implementation. Specifically, we focus on response rates, questionnaire pre-testing, translation methods and presence of fieldwork control. Additionally, given the substantive interests that motivate this project, we examine non-response to selected questions. We evaluate the extant documentation quality of each survey program relative to one another.

Harmonization and Survey Data Coverage

Harmonization is already an established concept in the social sciences. However, we must differentiate between two separate harmonization strategies: *ex-ante* and *ex-post* (Granda et al. 2010). All cross-national survey programs, by definition, are harmonized *ex ante*, or before completing any kind of fieldwork. The exact extent of this harmonization varies, but research teams in each setting use the same questionnaire (albeit often with some country-specific items), and try to ensure comparability among surveys by using uniform data collection standards. Examples of *ex ante* harmonization include the European Social Survey, the European Values Study, World Values Survey and Latinobarómetro, among many others. These survey programs collect data for the

specific purpose of making comparisons across different geographical and cultural units, most often countries.

In contrast, our project uses *ex post* harmonization. That is, we take existing data that were not collected with comparison in mind and manipulate it to make them comparable (Granda et al. 2010). Examples of similar projects include the Cross-National Equivalent File¹ (Burkhauser, Butrica, Daly and Lillard 2001) and studies of voting behavior over time performed by Schnell and Kohler (1995).

Ex-post harmonization offers several advantages to researchers. A harmonized dataset has a larger sample size and can offer more comprehensive country coverage. It allows for researchers to compare geographical units that may not be paired together in any currently existing dataset. Furthermore, harmonization adds more variability to the dataset and possibly even new variables.

Despite all of the potential advantages, harmonization projects are not common in the social sciences. This is likely because the challenges facing data harmonizers are substantial. Harmonization also comes at a cost, and we would be remiss not to acknowledge the problems endemic to this process. While many cross-national datasets contain similar or even identical questions, we must consider cultural contexts and language differences in judging which variables are appropriate for harmonization. We try to achieve comparability through proper selection, weighting and standardization of input variables. Additionally, every component survey contains its own errors and quality issues. These errors tend compound one another, so if they are not addressed then they will produce substantially biased results in the master data file.

¹ For more information on the Cross-National Equivalent File, see their website: <http://cnef.ehe.osu.edu>. Accessed 2 August 2014.

Data Scope and Coverage

Cross national researchers often choose their data with two considerations in mind. First, does the dataset cover the geographical area and timeframe of interest? Second, is the data of high enough quality to warrant further analysis? As part of the Harmonia project, we focus on variables related to protest and political behavior (dataharmonization.org). The Harmonia team is led by Professor Kazimierz M. Slomczynski of The Ohio State University and the Polish Academy of Sciences, and Director of Cross-national Studies: Interdisciplinary Research and Training (CONSIRT). They examined many different cross-national survey programs either included or eliminated them on the basis of their content, as well as the following criteria: Since the aim of the project is among other the development of cross-national survey methodology, survey programs chosen for this analysis must sample at least two countries². Second, because of the academic nature of the project, they must be in the public domain. Third, the individual must be the unit of analysis in order to allow for substantive analyses within the scope of the project, i.e. political attitudes and engagement. We display our chosen surveys, waves and timeframe covered in Table 1.

Insert Table 1 Here

The harmonization dataset that we examined includes 22 different survey programs and 1,721 country waves within 132 different countries/territories.³ To show the reach of the data across the regions of the world, Figure 1 displays a map of the world with the total number of

² While Harmonia allowed for two-country surveys in theory, in practice, the vast majority of survey programs are multinational, multicultural, and multiregional (3MC).

³ There are certain areas within cross-national surveys whose statehood is questioned, or that for a variety of reasons do not qualify as countries but are distinct territories nonetheless. For the purposes of our analysis, we simply treat these territories as equal to countries. Examples include Kosovo, Belgian Flanders, Palestine and others.

survey*country*waves per region, along with the sample size, respectively. The final, harmonized master file will likely include around 2.3 million respondents.⁴

Insert Figure 1 here

Figures 2 and 3 explore country coverage in more detail, presenting a gradient map of the world and Europe (where over a million residents from our sample reside), respectively. The color of each country corresponds to the number of survey*waves in that nation. These maps demonstrate how unequal survey coverage is across different regions of the world, and for countries within regions. Western and Central Europe are the most commonly surveyed regions of the world, followed by North and South America, Asia and the Middle East. Meanwhile most of Sub-Saharan Africa remains completely under-surveyed.

Insert Figure 2 here

Insert Figure 3 here

We use these maps to answer our first question with a definitive yes: there are more than enough data to justify a harmonized dataset of the world, but a more interesting question remains. Exactly how good are these data?

Data Harmonization and Survey Data Quality

There is an impressive body of literature on survey quality (for an overview see e.g. Harkness et al. 2010; Groves et al. 2010). Survey methodologists have examined response scales, translation

⁴ This number does not include the Asian Barometer, wave 2, for Hong Kong, as we were not able to locate information about the sample size in survey documentation. The total number used in this paper is a sum of sample sizes according to available survey documentation. The number of cases in published, publically available datasets is slightly lower (n= 2,291,074).

methodology, cross-national equivalence, mode effects, and many other issues. This work has greatly advanced our understanding of survey error, but there is comparatively little attention paid to evaluating the quality of survey data for the purposes of secondary analysis. This is crucial for the harmonization project for two reasons. First, we must assume that *ex ante* harmonization has already occurred. Put another way, we must trust that each survey program we use as the building blocks of our master file has *already* been harmonized in a satisfactory way and made comparable across the individual countries where the survey was implemented. There are ways to evaluate the strength of cross-national comparability within a single survey*wave, but they are not always readily apparent to the researcher. Secondly, we know that all surveys contain errors and as such we need to identify these sources of error prior to data manipulation.

How should we evaluate survey quality? This is a unique marketplace. Since the data we use are publicly available, there is no price to serve as a mark of quality.⁵ In addition, competition between surveys is limited, especially in regions where there is little geographic overlap between survey programs. In Europe, the ESS and the EVS sample generally the same region, so researchers have a choice. There are also other European surveys one can consider, like the Eurobarometers, the European Quality of Live Surveys, or Life in Transition surveys. In contrast, Latin America has much fewer options, and places like Africa or the Middle East (as figures 1 and 2 demonstrate) have few survey programs. Nevertheless, quality assessment remains an important part of harmonization, even if there are no other options from which to choose.

Survey documentation serves as a window into data quality, but is not quality itself. As previously mentioned, survey data collection is a complex and multi-stage process that requires many investigators to make many decisions. Sometimes, one option is clearly better than others (e.g.

⁵ And, the assumption that higher price equals better quality is already extremely tenuous.

probability sampling is generally more desirable than non-probability sampling). In other cases, it is a more difficult decision to make (e.g. the type of response scales, or survey mode). Additionally, on some occasions not all options are available (e.g. in a country that lacks a general population sampling frame available to researchers, random sampling is not feasible). Some of these decisions may lead to data collection bias, or result in limited comparability of data collected using other designs. In recognition of these problems, WAPOR has suggested strict documentation rules for survey teams, but these rules can be difficult to follow in practice (Inglehart 1997). During fieldwork, all efforts tend to be focused on the data collection itself, not on documenting the process (Mohler and Uher 2003). As a result, fully documented surveys are quite rare (Mohler, Pennell, and Hubbard 2008). Even when surveys are comprehensively documented, researchers have noted inconsistencies in documentation both the WVS and ISSP (Scholz and Heller 2009, Smith et al. 2011).

Documentation of the data collection process, even when spotty, is the best tool we have to understand how and why decisions were made in the field. It simply is not feasible to track down and contact the respondents who took part in a given survey. To find the documentation, we scoured online text at the websites of respective survey programs: study descriptions, method reports, sampling reports, technical reports, codebooks, and even random places online.⁶ Then, we constructed documentation quality indicators to search for within these places. De Leeuw, Hox and Dillman (2008) offer a good starting place for evaluating data quality, suggesting researchers focus on *cornerstones of data quality*: coverage, sampling, non-response and measurement. We reviewed the available documentation with these cornerstones in mind and used them to collect relevant indicators of documentation quality.

⁶ As a general rule, older surveys tend to have the poorest documentation, although this should not imply that newer surveys always have more in-depth documentation. Some very recent surveys, such as the Eurobarometers, do not provide any methodological information at all.

Quality Indicators in Survey Documentation

Among quality indicators we consider response rates, probably the single most widely used indicator of data quality. Although increasing the response rate does not necessarily improve the accuracy of results (Jones and Lang 1980), excessively low response rates can make the results unreliable.

While AAPOR defines response rate as full interviews divided by full and partial interviews plus non-interviews (for any reason) plus non-eligible cases, this definition assumes a probability sample. Hence, if a survey – for any reason – used a different sampling method (like random-route or quota sampling), the response rate figure indicated in survey documentation refers to something much different. Additionally, some survey programs allow for replacement in the case of a refusal, while others do not. With this, it is safe to say that calculating the response rate can be quite tricky in practice. As written by Johnson and Owens (2003: 130), “when a ‘response rate’ is given with no definition, it can mean anything”, and response rates without such definitions are not comparable.

To further complicate matters, we have found four additional definitions of response rates, four definitions of cooperation rates, three definitions of refusal rates and three definitions of contact rates (AAPOR 2011). It is surprising that basic indicators of survey quality have no internationally accepted definitions. Despite the many definitions of a response rate, it is still widely considered to be an important element of data quality. Hence, while acknowledging all the problems, we still make use of response rates as a documentation quality indicator in our study.

Next, we identify a survey’s translation methodology. If the translation process contains errors, then questions may lose their comparability and the survey instrument will be greatly harmed. Translation methods are a hotly debated topic in the survey methodology field (see e.g. Harkness et al. 2004, 2010). Our understanding of how translation influences the meaning of a question has

advanced greatly. Reflecting the importance of this issue, the number of survey*waves documenting some kind of translation methodology increases as the surveys become more recent. Back translation was popular during the 1990s, until its effectiveness came under scrutiny from the survey community. Other surveys use expert translation teams or focus groups to evaluate the translation. For the purposes of this paper, we searched documentation for evidence that the survey team had evaluated the translation in any way, such as back-translation, translation by professional translator, and team translation.⁷

Next, we searched the documentation for evidence that the questionnaire was pretested. Pilot studies are used to help researchers preemptively identify potential problems, and thereby correct unanticipated errors in the survey instrument, such as ambiguity or unintended connotations of certain terms in different cultures. While the specifics of a pretest differ from country to country, in much the same way as we searched for evidence of translation evaluation, we do the same for pretests.

The last element of survey quality gathered is documentation of fieldwork control. High quality surveys usually perform some kind of fieldwork control, and typically consists of a personal visit or phone call to back-check the previously collected data. Regardless of the method of control (phone call or personal visit), such fieldwork control is generally beneficial for a survey program (DeMaio et al. 1998) because it provides a quality check of the survey process, and improves interviewer performance. We search the documentation for evidence that this occurred, and identify the country*survey*waves that implement these quality assurance procedures.

⁷ We do not penalize surveys that use the language in which the survey questionnaire was originally written. So, if a questionnaire is originally written in English, and then implemented in the United Kingdom, then this country*wave is not penalized for its lack of a translation check. Such evaluations would be meaningless.

Finally, the third component of this study is an analysis of non-response to specific question items of substantive interest to the Harmonia project. We chose trust in parliament as the most popular item on political attitudes, and analyzed item non-response to questions about trust or confidence to the national parliament in all available surveys. Item non-response occurs when respondents who agreed to participate in the survey do not provide a valid answer to a question. In such cases, the variable is assigned a special value, usually “don’t know” or “no answer,” and is typically excluded from any analysis. An especially high non-response rate for a particular variable is a red flag to researchers. We then compare this question’s response rate to the quality scores established by the other quality indicators.

Survey Quality: Results

Response Rates

Given the importance of response rates in surveys, it is surprising that this information is often absent from the survey’s documentation and thus publicly unavailable to researchers. This information, or at least the components used to calculate it, no doubt exists, but in many circumstances it proved impossible to obtain from publically available resources. Table 2 presents the total number of survey*waves by survey program (column 2), the number of country*waves for which we were able to find a response rate (column 3), and the share of country*waves with response rates in each survey (column 4).

Insert Table 2 Here

Only one survey of the 22 provides information on response rates for all country*waves: the European Social Survey. The EQLS, Afrobarometer and ISSP all score highly, with between 90 and 98 percent of country*waves documenting a response rate. In contrast to these programs, many

surveys provide only spotty information about response rates. Examples of this include the Caucasus Barometer and the Arab Barometer. In the extreme, however, are the 9 survey programs that do not provide such information at all: Americas Barometer, ASES, CDCEE, CNEP, Eurobarometer, Latinobarómetro , PA2, PA8NS, VPCPCE.

Table 3 expands on Table 2's response rate information by including basic descriptive statistics of available response rates by survey program.

Insert Table 3 Here

There is broad variation in this documentation quality indicator, with the documentation reporting response rates ranging from 10% to 100%. This further justifies the need for publishing not only response rate figures themselves, but also the assumed definition of response rates and sampling methodology.

Survey process quality

Next, we present indicators that describe quality control throughout the survey process: documentation of the questionnaire translation method, pretesting, and fieldwork control. Each one of these quality indicators are dummy variables: 1 if a method was used, 0 if not. We display the quality controls used by each survey program in Table 4. Columns 3-5 contain information about the share of country*waves with a documented translation method, pretest, and fieldwork control, respectively. We use a simple method to combine these elements: the sum of the binary indicators for the three elements, results of which are presented in the last column.

Insert Table 4 Here

Again, the European Social Survey sets the pace. The ESS publishes information about all three aspects of the survey process for all country*waves.⁸ The EQLS and ASB are not far behind, and the ranking closes with six survey programs whose documentation contains no information on any of the three elements. All in all, it seems that pretesting is somewhat more problematic for most survey programs than translation and fieldwork control, with only 37% of all 1721 country*waves implementing a pilot study, compared to 46% and 44% of survey*waves documenting the translation method and fieldwork control, respectively.

Item Non-Response

Finally, we examined non-response to the variable, trust in parliament. In the 22 survey programs analyzed in this project, we found around 30 different value labels that could suggest non-response: “not sure,” “can't choose,” “don't know,” “do not understand the question,” “haven't thought much about it,” “decline to answer,” “break off,” “interviewer error,” “not asked in survey,” “not applicable,” “legal skip,” “no answer,” “missing,” “missing; unknown.” Some of these labels are very ambiguous and could refer to the midpoint of a Likert scale, which is not a true refusal to answer the question. For the purpose of this analysis, when unsure whether a value corresponded to non-response (e.g. “haven't thought much about it,” “not sure,” or “can't choose”), we checked if the option was part of the response scale presented to the respondent, (e.g. as the mid-point), outside of the scale, or not presented at all. The two latter situations qualified as non-response.

⁸ It is worth adding that ESS provides much more than just an indication that a given quality assurance procedure was implemented, and typically gives all details and data a user would need, if not more

Table 5 displays our findings on item non-response within the 22 surveys examined. It presents mean item non-response (as defined above) to questionnaire items about trust/confidence in the national parliament, by survey program.⁹

Insert Table 5 Here

According to the data, these averages range from a low of 3% in the ESS to a high of 10.5% in CDCEE. Some of these differences may result from specific cultural or political characteristics of the location at the time of survey, especially since many of the survey programs included in this analysis are region-specific. Also, results from CDCEE and ESS are not readily comparable for two reasons. First, the surveys were carried out in different years, with CDCEE's two waves occurring in the early 1990s and again at the turn of the century, while the ESS collects data biannually since 2002. Secondly, the surveys were implemented in different countries, and therefore drastically different political and social circumstances. The ESS covers mainly Western and Central Europe while CDCEE focused on post-communist countries. Nevertheless, this item non-response variation raises serious questions about the quality of the interviews and the training and oversight of the interviewers. This information is often absent from the available documentation, however. When high rates of non-response are observed, especially for theoretically important variables, researchers must give special care to the assessment of non-response bias and its consequences for research results.

Conclusions

Ultimately, we seek to make two contributions with this examination of survey data documentation.

First, we have made the case that there should be more attention paid to the documentation of the

⁹ The wording of questions on trust/confidence in the parliament naturally differs from survey to survey, and so do the length and type of response scales. The analysis of this variation and its consequences on comparability, although certainly important in cross-national studies, is however beyond the scope of this paper.

data collection process. Researchers choose datasets among competing options based on the questions that survey allows them to answer. This, of course, will likely remain the primary concern. Nevertheless, researchers need to better understand the method by which their data was collected, and can then make decisions accordingly. Among researchers, there is little consensus about why some datasets are better than others, and no accepted metric by which to compare and evaluate surveys. We offer these basic indicators of data documentation quality as a window into data quality. Response rates, translations checks, fieldwork control and pilot studies universally accepted as important to multinational, multicultural, and multiregional surveys. Therefore, the amount of widely used datasets that do not provide publicly available documentation on whether or not these quality controls were implemented is concerning. Survey researchers should, at minimum, document these four indicators and make that information available to researchers. Many of the input surveys we aim to harmonize are regularly scheduled events. The ESS, easily the highest scoring survey in our sample, is collected every other year. Most other regional barometers are carried out on a similar schedule. It would not be especially difficult for survey teams to provide complete documentation of the data collection process, and the benefits to researchers would be great.

Secondly, concerning future harmonization work, our team has found that the harmonization process can be quite tricky when we cannot find proper information about how variables were collected and where the potential errors are. We believe that harmonization can advance cross-national research by increasing the timeframe and geographical scope of existing datasets. In order to produce the best possible product, however, we need to be aware of potential errors and account for them during the harmonization process. There is no way to do that without examining documentation quality.

References

- AAPOR (The American Association for Public Opinion Research). 2011. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition.* AAPOR.
- Burkhauser, Richard V., Barbara A. Butrica, Mary C. Daley and Dean R. Lillard. 2001. "The Cross-National Equivilant File: A Product of Cross-National Research." Pp. 354-376 in *Einer Dynamischen Gesellschaft*, eds. Becker et al. Frankfurt, Germany: Campus Verlag.
- De Leeuw, Edith D., Joop J. Hox, and Don A. Dillman. *International Handbook of Survey Methodology.* New York, NY: Taylor & Francis Group.
- DeMaio, Theresa J., Jennifer Rothgeb, and Jennifer Hess. 1998. "Improving Survey Quality Through Pretesting." Statistical Research Division Working Papers in Survey Methodology, no.98/03/ Washington, DC: U.S. Bureau of the Census. Available at: <https://www.census.gov/srd/papers/pdf/sm98-03.pdf> accessed 2 August 2014.
- Granda, Peter, Christof Wolf and Reto Hadorn. 2010. "Harmonizing Survey Data." Ch. 17 in *Survey Methods in Multinational, Multiregional, and Multicultural Contexts.* Ed. Harkness et al. Hoboken, NJ: John Wiley & Sons.
- Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2010. *Survey Methodology.* 2nd edition. Hoboken, NJ: Wiley.
- Harkness, Janet A., Beth-Ellen Pennell, and Alisú Schoua-Glusberg. 2004. Questionnaire translation and assessment. Pp. 453-472 in Stanley Presser, Jennifer Rothgeb, Mick Couper, Judith Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer (Eds.), *Methods for testing and evaluating survey questionnaires.* Hoboken, NJ: John Wiley & Sons.
- Harkness, Janet A., Ana Villar, and Brad Edwards. 2010. "Translation, adaptation and design." Pp. 117-139 in Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts.* Hoboken, NJ: John Wiley & Sons.
- Inglehart, Ronald. 1997. *World Values Survey 1995-1997.* Ann Arbor: Inter-university Consortium for Political and Social Research.
- Johnson, Timothy and Linda Owens. 2003. "Survey response rate reporting in the professional literature." Paper presented at the 58th Annual Meeting of the American Association for Public Opinion Research, Nashville, May 2003. Available at: http://www.srl.uic.edu/publist/Conference/rr_reporting.pdf, accessed 29 July 2014.
- Jones, Wesley H. and James R. Lang. 1980. "Sample composition bias and response bias in a mail survey: A comparison of inducement methods." *Journal of Marketing Research* 17:69-76.
- Medrano, J.D. (2009) World values survey 2005 codebook. Retrieved May 14, 2011, from <http://www.worldvaluessurvey.org/WVSDocumentationWV5.jsp>, accessed 10 August 2014.
- Mohler, Peter and Rolf Uher. 2003. "Documenting Comparative Surveys for Secondary Analysis." Ch. 20 in *Cross-Cultural Survey Methods*, ed. Harkness et al. Hoboken, NJ: John Wiley & Sons.

- Mohler, Peter Ph., Beth-Ellen Pennell and Frost Hubbard. 2008. "Survey Documentation: Towards Professional Knowledge Management in Sample Surveys." Ch. 21 in *The International Handbook of Survey Methodology*, eds. Edith de Leeuw, Joop Hox and Don A. Dillman. New York, NY: Taylor & Francis Group.
- Schnell, Ranier and Ulrich Kohler. 1995. "Empirical Test of an Individualization Hypothesis Using Party Preference Data from 1953 to 1992." *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 47, 1995, S. 634-657.
- Scholz, Evi and Marleen Heller. 2009. "ISSP Study Monitoring 2007: Report to the ISSP General Assembly on monitoring work undertaken for the ISSP." *GESIS-Technical Reports 2009*.
- Smith, Shawna N., Stephen D. Fisher, and Anthony Heath. 2011. "Opportunities and challenges in the expansion of cross-national survey research." *International Journal of Social Research Methodology* 14(6): 485-502.

Tables

Table 1: Surveys Selected for Harmonization in the Harmonia Project: Waves and Years

Survey	Acronym	Years	Waves	Country*Waves
Afrobarometer	AFB	1999-2008	4	66
Americas Barometer	AMB	2004-2012	5	92
Arab Barometer	ARB	2006-2011	2	16
Asia Europe Survey	ASES	2001	1	30
Asian Barometer	ASB	2001-2010	3	18
Caucasus Barometer	CB	2009-2012	4	12
Comparative National Elections Project	CNEP	2004-2006	1	27
Consolidation of Democracy in Central and Eastern Europe	CDCEE	1990-2001	2	8
Eurobarometer	EB	1983-2012	7	152
European Quality of Life Survey	EQLS	2003-2012	3	93
European Social Survey	ESS	2002-2012	6	147
European Values Study	EVS	1981-2008	4	128
International Social Justice Project	ISJP	1991-1996	2	21
International Social Survey Programme	ISSP	1985-2011	13	363
Latinobarómetro	LB	1995-2010	15	260
Life in Transition Surveys	LITS	2006-2010	2	64
New Baltic Barometer	NBB	1993-2004	6	18
Political Action - An Eight Nation Study	PA8NS	1973-1976	1	3
Political Action II	PA2	1979-1981	1	8
Political Participation and Equality in 7 Nations	PPE7N	1966-1971	1	8
Values and Political Change in Postcommunist Europe	VPCPCE	1993	1	5
World Values Survey	WVS	1980-2008	5	182
Total			89	1721

Table 2: Availability of response rate information by survey program.

Survey	Total country*waves	with response rates	with rr % total
ESS	147	147	100
EQLS	93	91	98
AFB	66	61	92
ISSP	363	326	90
ISJP	21	16	76
NBB	18	13	72
EVS	128	74	58
PPE7N	8	4	50
WVS	182	89	49
ASB	30	14	47
LITS	64	29	45
CB	12	3	25
ARB	16	1	6
Total	1721	868	50
No information on response rates: AMB, ASES, CDCEE, CNEP, EB, LB, PA2, PA8NS, VPCPCE.			

Table 3: Reported Response Rates by Survey Program: Descriptive Statistics.

Survey	Mean	Std.dev.	Min.	Max.
ESS	0.621	0.100	0.310	0.800
EQLS	0.530	0.178	0.140	0.910
AFB	0.792	0.131	0.464	0.991
ISSP	0.575	0.172	0.103	0.975
ISJP	0.724	0.099	0.520	0.910
NBB	0.572	0.157	0.340	0.910
EVS	0.597	0.176	0.230	0.950
PPE7N	0.828	0.085	0.760	0.950
WVS	0.681	0.222	0.100	0.980
ASB	0.679	0.218	0.276	1.000
LITS	0.745	0.116	0.510	0.890
CB	0.750	0.030	0.720	0.780
ARB	0.954	-	0.954	0.954
Total	0.618	0.179	0.100	1.000
No information on response rates: AMB, ASES, CDCEE, CNEP, EB, LB, PA2, PA8NS, VPCPCE.				

Table 4: Data Documentation Quality Indicators by Survey Program

Survey	Total Wave*Countries	Translation Method	Pretest	fieldwork control	t+pt+fc
ESS	147	1.00	1.00	1.00	3.00
EQLS	93	1.00	0.99	0.70	2.69
ASB	30	0.97	0.50	0.97	2.43
ISSP	362	0.70	0.32	0.73	1.75
EVS	128	0.46	0.46	0.63	1.55
ASES	18	1.00	0.11	0.00	1.11
WVS	182	0.23	0.31	0.50	1.04
LITS	64	0.00	1.00	0.00	1.00
VPCPCE	5	1.00	0.00	0.00	1.00
AFB	66	0.30	0.30	0.30	0.91
EB	152	0.77	0.00	0.00	0.77
AMB	92	0.00	0.65	0.00	0.65
ISJP	21	0.24	0.10	0.19	0.52
LB	260	0.00	0.00	0.22	0.22
NBB	18	0.00	0.00	0.22	0.22
ARB	16	0.00	0.00	0.00	0.00
CB	12	0.00	0.00	0.00	0.00
CDCEE	27	0.00	0.00	0.00	0.00
CNEP	8	0.00	0.00	0.00	0.00
PA2	3	0.00	0.00	0.00	0.00
PA8NS	8	0.00	0.00	0.00	0.00
PPE7N	8	0.00	0.00	0.00	0.00
Total	1721	0.46	0.37	0.44	1.27

Table 5: Item Non-Response to Questions on Trust in Parliament by Survey Program.

Survey	% non-response
ESS	3.17
EQLS	3.41
EVS	4.16
LB	4.55
AMB	4.63
WVS	4.99
VPCPCE	5.19
ISSP	5.21
ARB	5.71
LITS	6.11
EB	6.33
CNEP	6.46
ASB	6.75
NBB	7.09
AFB	7.60
ASES	8.48
CB	8.49
CDCEE	10.54
Total	5.04%

Figures

Figure 1: Survey*Country*Waves and Sample Sizes by Region, Harmonized Dataset

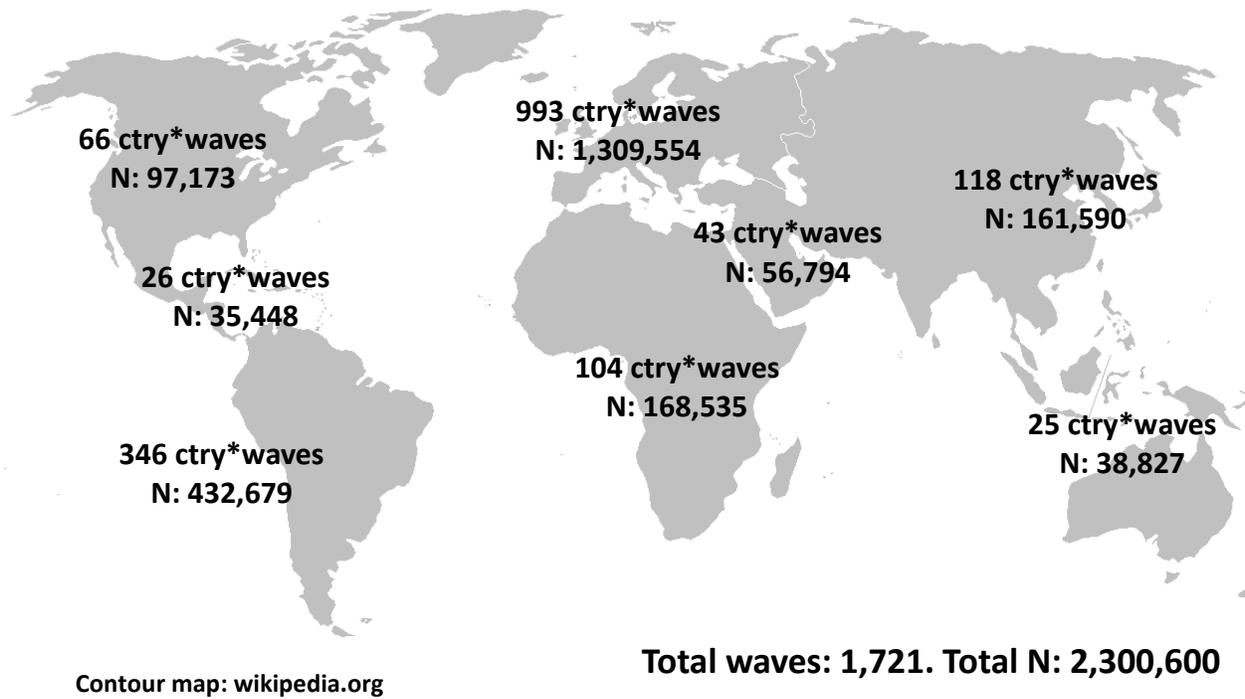


Figure 2: Gradient Map of Survey Waves by Country in Harmonization Project

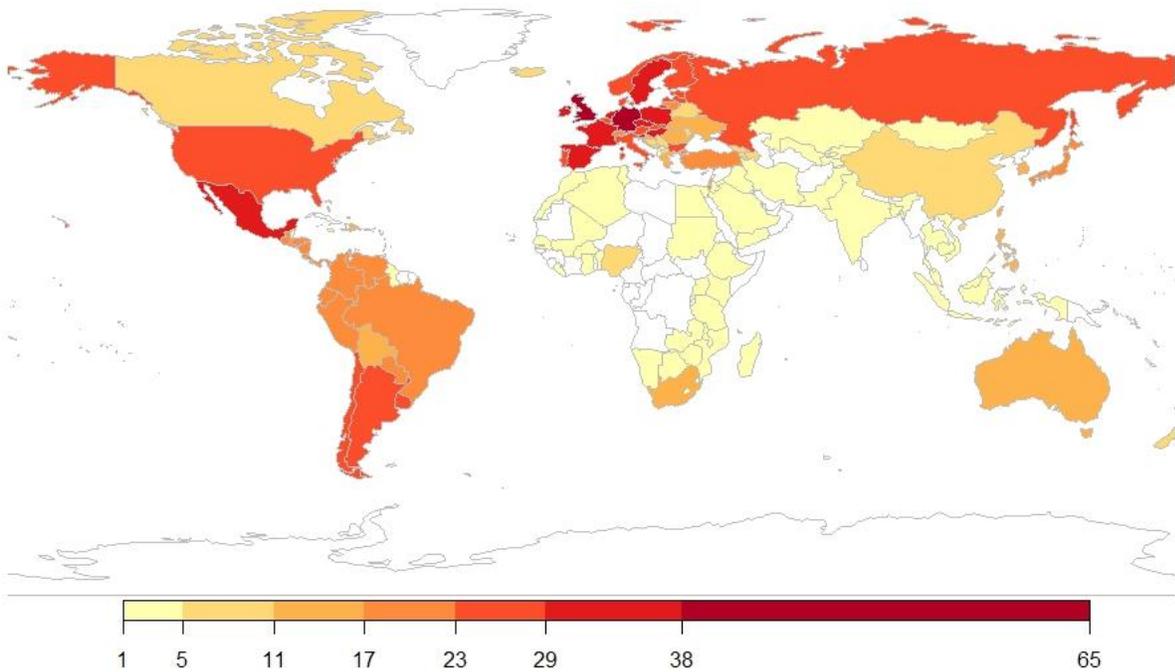


Figure 3: Gradient Map of Survey Waves by Country in Harmonization Project: Europe

