

Sampling and Fieldwork Practices in Europe: Analysis of Methodological Documentation From 1,537 Surveys in Five Cross-National Projects, 1981-2017

Piotr Jabkowski ^a , Marta Kołczyńska ^b 

[a] Faculty of Sociology, Adam Mickiewicz University, Poznan, Poland. [b] Department of Socio-Political Systems, Institute of Political Studies of the Polish Academy of Science, Warsaw, Poland.

Methodology, 2020, Vol. 16(3), 186–207, <https://doi.org/10.5964/meth.2795>

Received: 2019-03-23 • Accepted: 2019-11-08 • Published (VoR): 2020-09-30

Corresponding Author: Piotr Jabkowski, Szamarzewskiego 89C, 60-568 Poznań, Poland. +48 504063762, E-mail: pjabko@amu.edu.pl

Abstract

This article addresses the comparability of sampling and fieldwork with an analysis of methodological data describing 1,537 national surveys from five major comparative cross-national survey projects in Europe carried out in the period from 1981 to 2017. We describe the variation in the quality of the survey documentation, and in the survey methodologies themselves, focusing on survey procedures with respect to: 1) sampling frames, 2) types of survey samples and sampling designs, 3) within-household selection of target persons in address-based samples, 4) fieldwork execution and 5) fieldwork outcome rates. Our results show substantial differences in sample designs and fieldwork procedures across survey projects, as well as changes within projects over time. This variation invites caution when selecting data for analysis. We conclude with recommendations regarding the use of information about the survey process to select existing survey data for comparative analyses.

Keywords

cross-national comparative surveys, sample representativeness, sampling and fieldwork procedures, survey documentation

The primary goal of cross-national comparative survey research is to develop procedures and tools, and to apply them in the process of collecting data that allow for drawing methodologically rigorous conclusions about cross-national differences regarding phenomena of interest. The broad problematic of the comparability of surveys carried out in different countries, and at different times, constitutes a major area of methodological concern pertaining to the quality of comparative research (Jowell, 1998; Harkness et al., 2010). One of the main challenges consists in ensuring that realized samples are



comparable in representing the underlying populations – despite cross-national differences in sampling frames, sampling designs, and fieldwork procedures (for a review see, e.g., Stoop, Billiet, Koch, & Fitzgerald, 2010). Comparability is tricky, because achieving it sometimes requires changes in methodology, while at other times a change in methodology may entail a loss of comparability, and the conditions that distinguish one from the other are not well understood (Lynn, Japac, & Lyberg, 2006). Cross-national survey research thus requires careful attention at all stages of the survey process, with the comparative purpose of the project in mind (Smith, 2018).

From the point of view of the concerns with sample comparability, cross-national studies can be classified into three groups, listed from the most to the least common: studies that compare (1) national samples within the same project wave, (2) national samples across two or more waves of the same project, and (3) national samples between survey projects. In the case of the first two groups the comparability of samples is typically assumed but rarely discussed or verified, while it is recognized as a primary concern in the case of studies from the third group, i.e. those combining data from multiple projects. We argue that the comparability of samples can never be taken for granted, even if a given survey project has a reputation of being of high quality and the comparability of surveys is stated explicitly as part of its mission.

At the same time comparability can be approached in different ways and met to different degrees, and it is up to the researcher – the secondary data user – to decide whether a given set of surveys is sufficiently comparable for their research problem. This article aims to facilitate such decisions by providing standardized documentation of sampling and fieldwork procedures in cross-national surveys in Europe based on the available survey documentation of 1,537 national surveys from five cross-national survey projects conducted in 43 European countries between 1981 and 2017.

The purpose of this article is two-fold. First, it identifies the types of information crucial for evaluating the representativeness and comparability of samples in cross-national surveys, and documents their availability in the documentation that accompanies survey data files. In this way it serves as a guideline for survey projects in improving their documentation, and sets reasonable expectations for secondary users. Next, it describes and discusses the variation in essential aspects of the survey process, including sampling approaches, fieldwork procedures, and survey outcomes, both between and within survey projects. To carry out a more focused analysis, we centre our attention on the survey sample as the foundation of comparability in cross-national survey research, leaving aside issues of measurement, question wording and questionnaire translation.

Scope of the Analysis

We analysed the documentation of five cross-national comparative survey projects of established scientific standing, as measured by the number of publications recorded in the Web of Science Core Collection¹ database. The inventory includes: (1) all seventeen

autumn editions of the Eurobarometer (EB) from the years between 2001-2017, including three autumn rounds conducted in the years 2001-2003 in the Candidate Countries Eurobarometer (CCEB) project, (2) four editions of the European Quality of Life Survey (EQLS) from 2003, 2007, 2011 and 2016, (3) eight rounds of the European Social Survey (ESS) carried out biennially since 2002, (4) four editions of the European Values Study (EVS) from 1981, 1990, 1999 and 2008, and (5) thirty-one editions of the International Social Survey Programme (ISSP) carried out between 1985 and 2015. Each of the national survey samples in these projects aimed to be representative of the entire adult population of the given country. In total, the methodological inventory involved 64 different project editions (also called waves or rounds) and 1,537 national surveys (*project*edition*country*). The analysis was limited to surveys conducted in European countries, even if the survey project itself was covering countries outside Europe. Table 1 presents basic information about the analysed projects (for detailed information about the projects' geographical coverage see also Table S1 in [Supplementary Materials](#)).

Table 1

Cross-Country Research Projects Included in the Analysis

Project name	Time scope	Number of editions	Number of national surveys
Eurobarometer (CCEB & EB) ^a	2001-2017	17	523
European Quality of Life Survey (EQLS)	2003-2016	4	125
European Social Survey (ESS)	2002-2016	8	199
European Values Study (EVS)	1981-2008	4	112
International Social Survey Programme (ISSP)	1985-2015	31	578
Total		64	1,537

^aAutumn editions of the Standard Eurobarometer and the Candidate Countries Eurobarometer.

Method

The methodological documentation of the survey projects we reviewed takes on many different forms, from simple reports including only general information (EB and CCEB), to elaborate collections of documents with general reports and specialized reports describing in detail each national survey (ESS and EQLS, as well as the later editions of EVS and ISSP). Overall, there is little standardization in document content or formats across projects, and some variation also within projects (cf. also [Mohler, Pennell, & Hubbard,](#)

1) The queries consisted of the survey project names, returning information on the total number of publications, which amounted to 1971 positions, including: 604 for EB, 51 for EQLS, 930 for ESS, 171 for EVS, and 215 for ISSP [access on 21.07.2018].

2008). To construct our methodological database, we searched the project documentation for information on four aspects of the national surveys: (1) sampling procedures (including sampling frames, types of survey sample and procedures of within-household selection in address-based samples), (2) the sample design, (3) fieldwork procedures (including the length of fieldwork, mode of data collection, field substitution of nonresponse units, advance letters, incentives, refusal conversion, and back-checking procedures), and (4) survey outcome rates. The resulting dataset is available in [Supplementary Materials](#).

Sampling Procedures

Target Populations

While each of the analysed projects aims to collect data that are nationally representative of entire adult populations, the detailed definitions of target populations vary across projects, and even within project waves. The definition of the target population is important from two points of view. First, it identifies the population to which results from the sample can be generalized, the differences in which can be a very straightforward source of incomparability. Second, the definition of the target population influences the choice of the sampling frame, which in turn has consequences for the sampling design.

Sampling Frames

The primary characteristic differentiating national surveys is the type of sampling frame, of which the most popular types are: personal (individual) name registers, address-based registers, and household registers. Understandably, sample selection is most complicated in the case of non-register samples (Lynn, Gabler, Häder, & Laaksonen, 2007), such as area samples (Marker & Stevens, 2009). A whole separate category of registers applies to landline and mobile phone numbers, which are used solely for surveys conducted via telephone interviews. As part of the analysis, each national survey was classified into one of three categories based on the sampling frame: (1) personal / individual name register, (2) address-based or household-based register, and (3) telephone-contract-holder list. Two other types, (4) non-register samples and (5) non-probability or quota samples were coded as a separate category.

Types of Survey Samples

The projects we analysed exhibit substantial variation in sample types, with the ultimate choice determined by diverse factors, such as the availability of quality sampling frames, resources, local traditions and social science infrastructure, preferences or requirements of the given projects, and other conditions (Heeringa & O'Muircheartaigh, 2010). The primary distinction is between probability samples, where each respondent has a known and non-zero probability of being selected into the sample, and non-probability samples (Baker et al., 2013; Lohr, 2008), such as quota samples (Hubbard, Lin, Zahs, & Hu, 2016).

Each national survey in our inventory was placed in one of six categories, differentiating between (1) non-probability quota samples, and probability samples, of one of the five varieties: (2) simple random sample or stratified random sample with stratum allocations proportional to population size, (3) multistage individual sample, (4) multistage address-based or household-based sample, (5) random route sample, and (6) unspecified multistage address-based or household-based sample, following a classification proposed by Kohler (2007).

It is worth noting that the random route category comprises different types of samples with different selection rules. For example, in Round 4 of the ESS in France, households were selected via a random route procedure within each PSU in advance of fieldwork (ESS, 2008, p. 113). In Round 5 of the ESS in Austria, a random route sub-sample complements the sub-sample drawn from a telephone register to correct for imperfect coverage of the register (ESS, 2010, p. 26). In the 1999 Edition of the ISSP in Russia, random route procedures were used to select households within secondary sampling units, but the documentation does not make it clear whether household enumeration was separate from household selection (ISSP, 1999, p. 55). Some random route variants yield better quality samples than others, but they do not closely approximate probability sampling (Bauer, 2014, 2016).

Within-Household Selection of Target Respondents in Address Samples

Within-household selection of target respondents is necessary in the case of non-register samples or address and household-based samples. The survey literature describes over a dozen procedures for the within-household selection of the respondent (Gaziano, 2005; Koch, 2018). In our analysis for each national survey, we coded the type of within-household selection procedure according to the following key: (1) probabilistic: Kish grid, (2) probabilistic: non-Kish grid, (3) quasi-probabilistic: last-birthday, (4) quasi-probabilistic: next-birthday, (5) quasi-probabilistic: closest-birthday, (6) quasi-probabilistic: birthday method unspecified, (7) non-probabilistic: any.

Sample Design

Detailed information about the sample design, in particular about stratification and clustering, is necessary for understanding how exactly the sampling was carried out. Additionally, we searched for information about non-probabilistic selection at any stage – as a potential threat to sample quality, and about the size of the design effect, which can be used to compare the effective sample size between samples drawn following different designs (cf. Lynn et al., 2007).

A review of the available survey documentation showed that full information about all elements of sample design is practically present only in the ESS. Hence, while we include the presence of information on sample design in the assessment of the survey

documentation quality, we abstain from examining the differences in this regard across projects in our main analysis.

Fieldwork Procedures

Among fieldwork procedures, we first focus on the interview mode, distinguishing between Paper-And-Pencil Interviewing (PAPI), Computer-Assisted Personal Interviewing (CAPI), Face-to-Face Interviewing (where the distinction between PAPI and CAPI was not possible), Computer-Assisted Telephone Interviewing (CATI), Computer-Assisted Web Interviews (CAWI), Postal Surveys and Drop-off surveys. Next, we record information about the length of fieldwork (based on the dates of the beginning and end of fieldwork) as a proxy for the overall fieldwork effort. Related to the quality of the sample, we note whether substitutions were allowed. Finally, we record whether the surveys used response-enhancing procedures, such as incentives, advance letters, or refusal conversion, and whether they employed back-checks as a way of verifying the quality of fieldwork.

The presence or absence of these fieldwork procedures can influence the quality of the data and the survey outcome rates, as shown by methodological studies discussing the mode effect (Bethlehem, Cobben, & Schouten, 2011), the negative correlation between the length of fieldwork and the fraction of nonresponse units (Vandenplas, Loosveldt, & Beullens, 2015), as well as the positive impact on the measurement quality of: advance letters (von der Lippe, Schmich, Lange, & Koch, 2011), incentives (Grauenhorst, Blohm, & Koch, 2016), refusal conversion (Stoop et al., 2010), back-checking procedures (Kohler, 2007) and the negative impact of fieldwork substitutions on nonresponse bias (Elliot, 1993).

Survey Outcome Rates

Reporting on the effects of fieldwork consists of establishing the number of respondents, the number of non-respondents, and the evaluation of several standard survey outcome rates, e.g. the *response rate*, *cooperation rate*, *refusal rate* and *contact rate*. If the survey outcome rates are to be compared, they need to be calculated according to the same definitions (Smith, 2002). We relied on the standard of the American Association for Public Opinion Research (AAPOR, 2016). For each survey, we attempted to establish the number of: (1) respondents, i.e. of complete and partial interviews, (2) non-contacts, (3) refusals and break-offs, (4) cases to which other reasons for nonresponse apply, (5) persons with unknown eligibility, (6) not eligible units. Based on this information, we calculated survey outcome rates based on the AAPOR (2016) standard.

Response rates are now commonly provided by survey projects and increasingly calculated following standard procedures. While response rates are associated with sample representativeness (Corney & Bosnjak, 2018), they remain “second best quality indica-

tors” (Mohler, 2019; cf. Couper & de Leeuw 2003), substituting for nonresponse bias, which is notoriously difficult to quantify and thus not addressed in the documentation of most of the analysed survey projects. The exception is the ESS, which in recent rounds documents nonresponse bias based on neighbourhood characteristics as well as population register data on age and gender (Beullens et al., 2016, pp. 70-73; Wuyts & Loosveldt, 2019, pp. 86-99).

Results

Quality Assessment of Methodological Documentation

Before presenting cross-project differences in survey practices based on survey documentation, we assess the information content of the project documentation itself. Whether the survey documentation contains all the essential information about the key stages of the survey process has a direct bearing on survey usability, which is an important – if underappreciated – dimension of survey quality (Biemer, 2010). Our evaluation draws from an earlier study of survey documentation quality that included fewer and more general indicators, also covering questionnaire development (Kołczyńska & Schoene, 2018).

With a focus on sample design and representativeness, we constructed indicators in four areas: (1) sampling procedures, (2) sampling design, (3) fieldwork procedures, and (4) reported survey outcome rates, each with between two and six pieces of information that are crucial for forming an opinion about a survey’s quality. In the area of sampling, we looked for information that would allow us to judge the representativeness of the drawn sample, referring to the target population, the sampling frame, the type of the sample, and within-household selection procedures. Regarding sampling design, we searched for information that would allow us to reconstruct the exact sampling process, i.e., on the presence or absence of stratification, clustering, whether the sample was non-probabilistic, and on the design effect. Among fieldwork procedures, we focused on the survey mode, information on whether substitutions were allowed, as well as on procedures aimed at increasing survey response and data integrity: whether the survey employed incentives, advance letters, refusal conversion techniques, and back-checking or fieldwork control. Finally, pertaining to survey outcomes, we collected information about the response rate and information breaking down the drawn sample depending on eligibility, contact, and response, as necessary for the calculation of different outcome rates (for full schema see Table S2 in [Supplementary Materials](#)).

We score each national survey on whether the documentation contains information on the selected aspect of methodology (coded 1) or not (coded 0). The proportion of positive scores in each area constitutes the value in that area, ranging from 0 if the

project documentation did not contain that kind of information, to 1 if all pieces of information we searched for were included.

According to the documentation quality scores shown in Table 2, the best-documented project – in all editions – is the ESS, which kept a consistently high level of reporting on the methodology of national surveys, and the EQLS. The quality of survey documentation conducted as part of EVS and ISSP was much lower and much more differentiated. EB and CCEB have the weakest methodological documentation. The documentation for both these projects is only available for entire editions (*project*edition*), and not for individual national survey samples, which explains the lack of variability in documentation quality within editions. The documentation content of EB and CCEB changed beginning with the 2004 edition when an additional piece of information on within-household selection procedures was included. In order to show this, we represent EB and CCEB in Table 2 with two rows.

Table 2

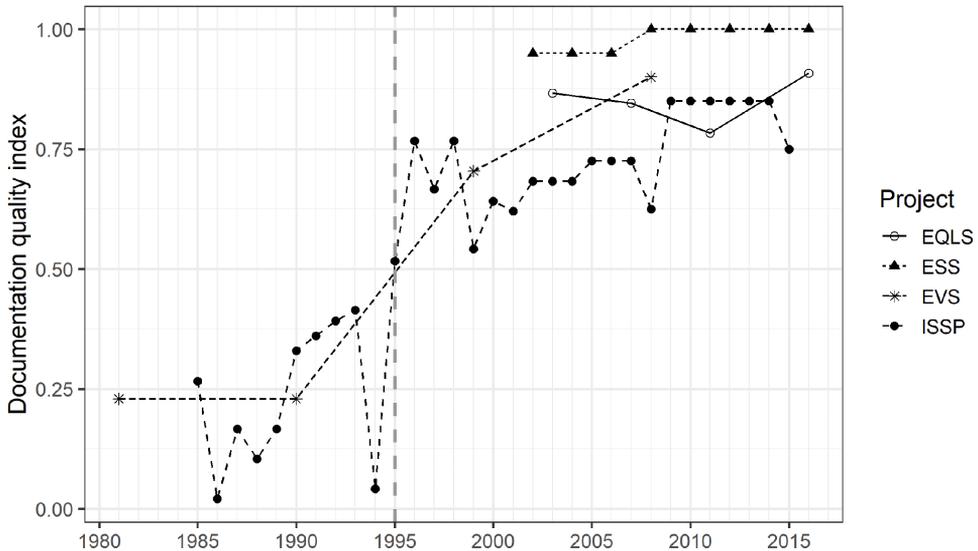
Comparison of the Quality of Methodological Documentation of the Cross-Country Projects

Project abbreviation	Number of national surveys	Survey documentation quality ^a				Overall description quality ^a
		Sampling	Sample design	Fieldwork	Outcomes	
CCEB & EB (2001-2003)	84	0.750 (0.00)	0.800 (0.00)	0.333 (0.00)	0.000 (0.00)	0.471 (0.00)
EB (2004-2017)	439	1.000 (0.00)	0.800 (0.00)	0.333 (0.00)	0.000 (0.00)	0.533 (0.00)
EQLS	125	0.938 (0.11)	0.800 (0.00)	0.796 (0.07)	0.860 (0.23)	0.848 (0.05)
ESS	199	0.999 (0.02)	0.924 (0.12)	0.995 (0.03)	1.000 (0.00)	0.979 (0.03)
EVS	112	0.848 (0.21)	0.336 (0.29)	0.557 (0.36)	0.586 (0.46)	0.581 (0.28)
ISSP	578	0.841 (0.31)	0.303 (0.20)	0.710 (0.33)	0.566 (0.36)	0.605 (0.25)

Note. CCEB = Candidate Countries Eurobarometer; EB = Eurobarometer; EQLS = European Quality of Life Survey; ESS = European Social Survey; EVS = European Values Study; ISSP = International Social Survey Programme.

^aMean value of national indicators (*project*edition*country*). Values in parentheses represent standard deviations (*SD*).

Figure 1 presents the median documentation quality for each project edition, showing the clear superiority of ESS compared to all the other projects. It should be noted that in the comparatively weakest-described second round of ESS from 2004, the quality of documentation was still substantially higher than that of the best-described edition from among all the other projects. It should also be pointed out that the later editions of the two longest-lasting projects, EVS and ISSP, saw substantial improvement in the quality of documentation in the late 1990s, following the seminal report on the consistencies and differences between national surveys in ISSP Round 1995 (Park & Jowell, 1997), which set the standard for future survey documentation efforts of ESS and EQLS.

Figure 1*Differences in the Overall Documentation Quality in Cross-Country Projects Over Time*

On the contrary, EB and CCEB – not shown in the graph due to lack of changes over time – clearly stand out as not only exhibiting no improvement and consistently withholding information on outcome rates but also as not providing documentation on the level of national surveys. Similar patterns – both in terms of cross-project differences and within-project changes – were found in [Kolczyńska and Schoene \(2018\)](#).

Cross-Project Differences in Target Populations

Cross-national survey projects typically have project-wide definitions of target populations that apply to most national surveys. According to general definitions, target populations in EB and CCEB include the resident populations aged 15 and over ([European Commission, 2017](#)). EQLS targeted “all people aged 18 and over whose usual place of residence is in the territory of the countries included in the survey” ([EQLS, 2013](#), p. 3). ESS surveys “all persons aged 15 and over resident within private households, regardless of their nationality, citizenship, language or legal status” ([ESS, 2017](#), p. 6). In EVS the definition is practically the same as for ESS; the only difference is the lower age limit of 18 ([EVS, 2016](#)). ISSP included “persons aged 18 years and older” ([ISSP, 2017](#), p. VII).

These standard definitions are not uniformly enforced. ISSP and – to a lesser extent – EVS allow certain deviations in the target populations. In some surveys in ISSP the lower age cut-off is as low as 14, while in others it is as high as 21; if the surveys have an upper

age cut-off, it ranges from 65 to 94. EVS conforms to the project-wide minimum age of 18, but a few surveys employ an upper age cut-off at 74, 75, or 80.

Apart from age cut-offs and the exclusion of institutionalized populations, the other potential restrictions of target population definitions are in terms of geographic coverage. These geographic exclusions typically pertain to sparsely populated or distant territories that are unlikely to have a substantial influence on the results of cross-national analyses. However, they may also include regions where conducting fieldwork would be associated with increased risk, e.g., due to security concerns.

Cross-Project Differences of Sampling Procedures and Sample Types

Table 3 shows distributions of sampling frames, types of survey samples, and procedures used to select respondents within households in the analysed survey projects. The presented comparisons show significant cross-project variation mainly resulting from the different sampling strategies employed by projects. For example, EB and CCEB use area samples, where households are chosen primarily using *random route* procedures, and target respondents are selected using one of the different birthday rules. EQLS relies on either personal or address-based registers, but area samples using *random route* protocols (for household selection) are also common. EVS is characterised by a large proportion of non-probability quota samples in the early waves of the project. ISSP relies primarily on probabilistic samples based on address or – less frequently – individual registers, with a small proportion of quota samples and multistage samples of an undetermined type. ESS uses probabilistic samples exclusively. EQLS, ISSP and ESS more often use the birthday rules than the *Kish grid* for the within-household selection of respondents. In EVS, both methods are used equally.

Table 3

Types of Sampling Frames, Fieldwork Procedures, and Methods of Within-Household Selection of Target Respondents in Cross-Country Comparative Projects

Compared aspect	CCEB	EB	EQLS	ESS	EVS	ISSP
Sampling frame						
[1] personal/individual	-	-	17.0%	46.2%	17.3%	38.2%
[2] address-based	-	-	27.7%	45.2%	26.0%	47.5%
[3] telephone	-	-	-	-	-	0.4%
[4] non-register/area	100%	100%	55.3%	8.6%	18.3%	11.9%
[5] non-random	-	-	-	-	38.4%	2.0%
No data/total number of surveys	0/39	0/484	31/125	0/199	8/112	81/578

Compared aspect	CCEB	EB	EQLS	ESS	EVS	ISSP
Type of survey sample						
[1] non-probabilistic: quota	-	-	-	-	48.0%	2.4%
[2] simple random	-	-	4.0%	20.2%	6.9%	17.2%
[3] stratified random PPS	-	-	4.0%	26.3%	10.8%	19.6%
[4] multistage: address-based	-	-	26.4%	44.9%	14.7%	49.4%
[5] multistage: random route	100%	100%	65.6%	8.6%	15.7%	8.6%
[6] multistage: nondescript	-	-	-	-	3.9%	2.6%
No data/total number of surveys	0/39	0/484	0/125	1/199	6/112	102/578
Within-household selection procedure						
[1] probabilistic: Kish grid	-	-	34.1%	35.5%	40.5%	47.3%
[2] probabilistic: other	-	-	-	-	-	1.4%
[3] quasi-probabilistic: birthday	-	100%	65.9%	64.5%	38.1%	44.9%
[4] non-probabilistic: any	-	-	-	-	21.4%	6.4%
No data/total number of surveys	39/39	45/484	28/116	0/107	52/94	112/402

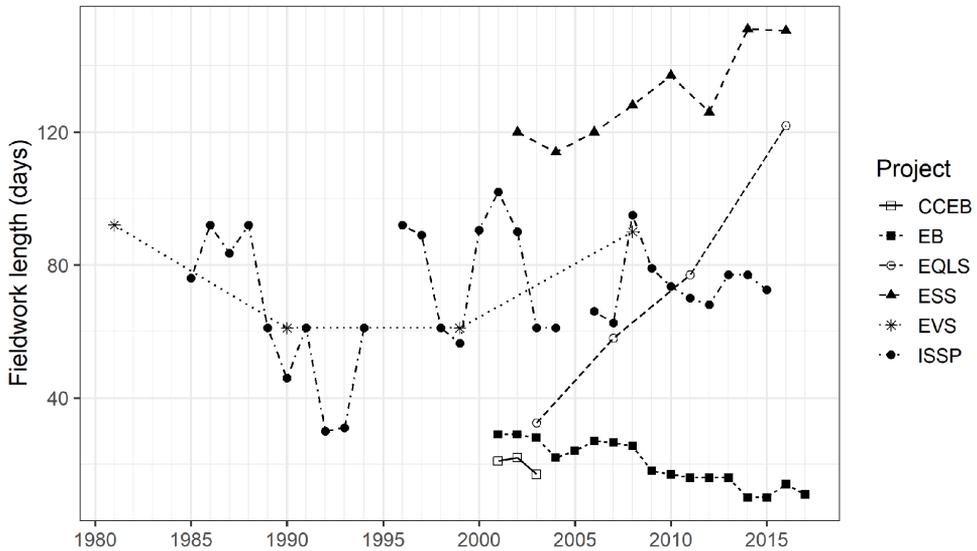
Note. CCEB = Candidate Countries Eurobarometer; EB = Eurobarometer; EQLS = European Quality of Life Survey; ESS = European Social Survey; EVS = European Values Study; ISSP = International Social Survey Programme.

Cross-Project Differences in Fieldwork Length and Fieldwork Procedures

The next part of the analysis deals with differences in fieldwork length and fieldwork procedures. Figure 2 indicates significant cross-project differences in the length of fieldwork. EB and CCEB have by far the shortest fieldwork periods, with the median in all the analysed editions not exceeding 29 days, and in the editions conducted after 2008 dropping below 20 days. Other projects have considerably longer fieldwork periods than EB and CCEB. Specifically, the last edition of EQLS and all the editions of ESS, saw a median length of fieldwork of between 114 and 151 days (between 16 and 22 weeks). In EVS median fieldwork length ranged between 60 and 92 days (between 9 and 13 weeks), and between 30 and 102 days (between 4 and 15 weeks) in ISSP. It is worth noting that the specification of ESS requires that fieldwork not be shorter than four weeks to avoid a high rate of non-contacts, while the maximum recommended fieldwork length is four months (Koch & Blohm, 2006).

Figure 2

Median Length of Fieldwork in Days in Different Editions of Survey Projects



The length of fieldwork is not without consequences for survey outcome rates. One of the simplest methods for improving the response rate is by extending the time devoted to fieldwork (Stoop et al., 2010). Additionally, longer fieldwork results in broadening the scope of available procedures that can increase the chance of conducting interviews with people who spend less time at home and with reluctant respondents. Recognizing this, the ESS, as of Round 10, decided to extend the minimum length of the fieldwork from four to six weeks (European Social Survey, 2019, p. 5). At the same time, extended duration may be a symptom of difficulty during fieldwork execution, including an insufficient number of interviewers or low interviewer motivation.

An examination of the variation in fieldwork procedures (Table 4) provides more context for the short fieldwork periods in EB and CCEB. Both projects, in all their national surveys, allow fieldwork substitutions (cf. Kohler, 2007). Allowing substitutions significantly reduces the effort required to obtain a successful interview, as the interviewers do not need to make repeated attempts to contact the particular individual selected for the survey. Instead, after a failed interview attempt, they can choose a substitute respondent. Such practices in no way improve the quality of the survey sample, but they certainly shorten the length of fieldwork (Chapman, 2003; Vehovar, 1999). In EQLS and ESS, substitutions of unavailable respondents were prohibited in all the national surveys and in the ESS they were treated as interview falsification (cf. Stoop, Koch, Halbherr,

Loosveldt, & Fitzgerald, 2016). Substitutions were allowed in some national surveys in EVS and ISSP.

Table 4

Differences in Fieldwork Procedures in Cross-Country Comparative Studies

Compared aspect	CCEB	EB	EQLS	ESS	EVS	ISSP
Mode of data collection						
[1] PAPI	-	-	40.8%	42.2%	85.6%	52.9%
[2] CAPI	-	-	59.2%	57.8%	12.6%	46.7%
[3] F2F (PAPI or CAPI)	100%	100%	-	-	-	-
[4] CATI	-	-	-	-	0.9%	4.4%
[5] CAWI	-	-	-	-	0.9%	4.3%
[6] Postal-Survey	-	-	-	-	0.9%	21.7%
[7] Drop-off survey	-	-	-	-	-	18.0%
No data / total number of surveys	0/39	0/484	0/126	0/199	1/112	38/540
Fieldwork substitutions						
[1] Allowed by protocol	100%	100%	0%	0%	17.9%	15.4%
No data / number of surveys	0/39	0/484	0/125	0/199	42/112	123/578
Advance letters						
[1] Present	-	-	100%	83.4%	25.9%	26.3%
No data / number of surveys	39/39	484/484	0/125	0/199	71/112	255/578
Incentives						
[1] Present	-	-	12.8%	63.8%	17.0%	13.5%
No data / number of surveys	39/39	484/484	92/125	1/199	21/112	252/578
Refusal conversion						
[1] Present	-	-	51.2%	74.9%	20.5%	24.4%
No data / number of surveys	39/39	484/484	33/125	4/199	71/112	192/578
Back-checking procedures						
[1] Present	-	-	77.6%	96.5%	55.4%	58.0%
No data / number of surveys	39/39	484/484	28/125	1/199	42/112	147/578

Note. PAPI = Paper-And-Pencil Interviewing; CAPI = Computer-Assisted Personal Interviewing; CATI = Computer-Assisted Telephone Interviewing; CAWI = Computer-Assisted Web Interviews; CCEB = Candidate Countries Eurobarometer; EB = Eurobarometer; EQLS = European Quality of Life Survey; ESS = European Social Survey; EVS = European Values Study; ISSP = International Social Survey Programme.

As shown in Table 4, by far the most common mode of data collection in all the cross-country projects was the face-to-face interview. CAPI was more common in the younger projects, ESS and EQLS, while PAPI dominates the ISSP and – to a lesser extent – also EVS. In EB and CCEB, the documentation does not allow us to establish whether PAPI or CAPI was employed. In a small percentage of national surveys in EVS also CATI,

CAWI and postal surveys were used. ISSP occasionally employed multiple modes, in most cases consisting of PAPI or CAPI supplemented by postal or drop-off surveys. It is worth mentioning that ISSP questionnaires are often administered with another survey to reduce the cost of data collection. For example, the 2011 wave of ISSP in Germany was fielded with the German General Survey 2012 ALLBUS (ISSP, 2013, p. 38)².

An analysis of the use of performance-enhancing procedures provides information about the effort survey projects put into data collection. Advance letters were employed in all the national surveys conducted in EQLS and in most of those conducted in ESS, but to a lesser extent in EVS and ISSP. Monetary or material incentives were used primarily in ESS, far less frequently in other projects. Refusal conversions, that is procedures used to convince people refusing to be respondents to participate after all, were employed in 3/4 of all ESS surveys, and in over half of EQLS surveys, but in EVS and ISSP refusal conversion was used much less frequently, in only 1 in 5 national surveys. Finally, back-checking procedures were used in almost all ESS surveys, slightly less often in EQLS, and significantly less so (although still in more than half of the national surveys) in EVS and ISSP.

Cross-Project Differences in Survey Outcome Rates

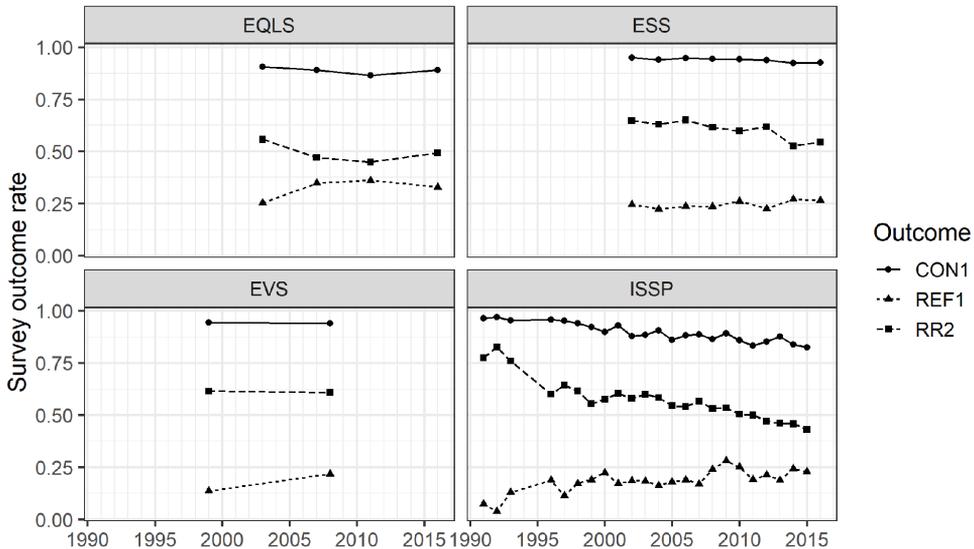
Fieldwork efforts have an impact on the final results, including on the survey outcome rates (cf. Sturgis, Williams, Brunton-Smith, & Moore, 2017). Figure 3 presents three survey outcome rates: the response rate (RR2), the contact rate (CON1), and the refusal rate (REF1), achieved in surveys in consecutive editions of the compared projects. In line with the AAPOR (2016) definitions, RR2 is the number of complete and partial interviews divided by the number of interviews (complete and partial) plus the number of non-interviews (refusal and break-off plus non-contacts plus others) plus all cases of unknown eligibility (unknown if housing unit, plus unknown, other). The CON1 treats all cases of indeterminate eligibility as eligible, and measures the proportion of all cases in which the survey reached any member of the housing unit. REF1 is the proportion of cases in which a housing unit or respondent refuses to give an interview, or breaks off an interview among all potentially eligible cases. Data that would allow calculating the values of RR2, CON1, and REF1 were not available for EB and CCEB, nor for some editions of the EVS (1981-1990) or the ISSP³ (1985-1990 and 1994-1995).

2) Additionally, in some countries two ISSP waves are sometimes fielded together, i.e. two questionnaires are administered to the same respondents (cf. ISSP, 2013, pp. 38-39).

3) As mentioned earlier, ISSP is sometimes fielded following another survey. ISSP documentation does not explain how reported outcome rates are related to the response rates of the preceding survey.

Figure 3

Survey Outcome Rates RR2, CON1, REF1 in Different Editions of Survey Projects



Data shown in Figure 3 confirm reports about the growing difficulty faced by researchers conducting surveys (de Leeuw, Hox, & Luiten, 2018). In all analysed projects, later editions noted a significant rise in refusal rates, and a corresponding decline in contact and response rates. The latter is particularly evident in the ISSP: at the beginning of the 1990s the median response rate was about 80 percent, but in the last archived edition from 2015 (at the time data were prepared for this paper) it dropped to around 45 percent. This decline was caused by a significant reduction in contact rates, and to a lesser degree by the rise of refusals. Comparing editions of the projects that were conducted in the same year, the highest values of RR2 and CON1 were usually noted in ESS, and the lowest in ISSP. The highest refusal rates were noted in EQLS, and the lowest in ISSP.

Discussion and Conclusions

In this paper we identified properties of surveys that are important for sample comparability and described their variation in 1,537 national surveys from five survey projects carried out in Europe. Our results show substantial differences in sample designs and fieldwork procedures across these projects, as well as changes within projects over time. In some projects (e.g., ESS and EQLS), ensuring the comparability of survey procedures is given much more priority than in others, which is visible both in the quality of the

survey documentation in terms of its information content, in the length of the fieldwork, procedures of sample selection and fieldwork execution that allow an oversight of the interviewers' work, or the use of additional procedures improving survey outcome rates. The presented data also confirmed the observations made in prior studies about the growing difficulties with fieldwork execution, including a decline in response rates due to an increasing number of refusals to participate and decreasing contact rates.

Comparability requires that survey procedures result in similar amounts of errors, both random and non-random, in all samples included in the analysis. There is a rich literature on the different sources of error in the Total Survey Error (TSE) paradigm. However, much of this literature discusses different components of TSE separately or in turn (Smith, 2018). It is unclear how different sources of error interact, add up, or cancel out, thereby influencing overall survey comparability. In practice, the most realistic way of improving survey comparability is by minimizing all types of errors as much as possible, and thoroughly documenting the entire survey process including all potential sources of error.

With this in mind, the properties we identified and catalogued can be used to select surveys that meet comparability criteria for substantive analyses. Cut-offs for identifying surveys that do not meet these criteria depend on research goals and mode of analysis. In the Fisher-Neyman tradition of statistical modelling, only probability samples can be legitimately included, and non-probability samples should be discarded. Similarly, the lack of weights could be considered disqualifying, unless the survey relies on simple random sampling where all design weights are equal to 1 by definition. The situation becomes more complicated in the absence of adequate documentation that would enable the assessment of survey comparability. Thus, while specifying cut-off criteria may only be possible with a specific research goal in mind, we strongly recommend that survey data producers prepare the documentation in a way that enables secondary data users to understand the threats to the comparability of surveys, and to make informed decisions about which surveys to include in their analyses. At the very least, all survey documentation should include the types of information that we collected and analysed in the current study for each national survey separately, and not for entire rounds, as is the case for the Eurobarometer. While we observe that the quality of documentation has substantially increased over time in all four remaining projects, there is still room for improvement, including the standardization of the methodological information provided in the survey documentation. It is also worth noting that our analysis only includes surveys from Europe, which tend to be of high quality relative to cross-national survey projects carried out in other regions of the world (cf. Kołczyńska & Schoene, 2018).

Another way of using survey characteristics would be to include them directly into substantive models, e.g. to model the uncertainty associated with particular estimates of interest. Including properties of surveys in models requires decisions regarding the model design that depend crucially on the given research problem. Additionally, account-

ing for survey quality in the models would require quality indicators that quantify deviations from comparability, such as measures of nonresponse bias, which – as we have mentioned earlier – are not typically provided in survey documentation. While outside of the scope of the current paper, this constitutes a promising avenue for future research.

Moreover, sample comparability is only one of the challenges of cross-national comparative research with survey data. The other major challenge is related to measurement, and establishing configural, metric, and scalar invariance of latent constructs despite the different cultural contexts in which the studies are carried out (for a review see, e.g., Milfont & Fischer, 2010). Future research could address the combined effects of sample comparability and measurement equivalence on total survey comparability, both in terms of the nature of these effects – whether additive, interactive or otherwise – and the relative threat that sample comparability and measurement equivalence pose to the validity and reliability of results of comparative studies relying on cross-national survey data.

Funding: This work was supported by grants awarded by the National Science Centre, Poland (no. 2018/31/B/HS6/00403 and no. 2019/32/C/HS6/00421).

Competing Interests: The authors have declared that no competing interests exist.

Acknowledgments: We would like to thank the reviewers for valuable suggestions for improving the article, and Robin Krauze for language editing.

Data Availability: Standardized metadata used in this article are freely available (see [Supplementary Materials](#) section).

Supplementary Materials

1. Table S1: European countries according to frequency of participation in cross-country projects (Jabkowski & Kołczyńska, 2020a).
2. Table S2: Survey documentation quality indicators (Jabkowski & Kołczyńska, 2020a).
3. Dataset on sampling and fieldwork practices in Europe (Jabkowski & Kołczyńska, 2020b).

Index of Supplementary Materials

Jabkowski, P., & Kołczyńska, M. (2020a). *Supplementary materials to: Sampling and fieldwork practices in Europe: Analysis of methodological documentation from 1,537 surveys in five cross-national projects, 1981-2017*. PsychOpen. <https://doi.org/10.23668/psycharchives.3460>

Jabkowski, P., & Kołczyńska, M. (2020b). *Supplementary materials to: Sampling and fieldwork practices in Europe: Analysis of methodological documentation from 1,537 surveys in five cross-national projects, 1981-2017 [Dataset]*. PsychOpen. <https://doi.org/10.23668/psycharchives.3461>

References

- American Association for Public Opinion Research. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (9th ed.). Retrieved from https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf
- Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., & Dever, J. A. ... Tourangeau, R. (2013). *Report of the AAPOR task force on non-probability sampling*. Retrieved from https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/NPS_TF_Report_Final_7_revised_FNL_6_22_13.pdf
- Bauer, J. J. (2014). Selection errors of random route samples. *Sociological Methods & Research*, 43(3), 519-544. <https://doi.org/10.1177/0049124114521150>
- Bauer, J. J. (2016). Biases in random route surveys. *Journal of Survey Statistics and Methodology*, 4(2), 263-287. <https://doi.org/10.1093/jssam/smw012>
- Bethlehem, J., Cobben, F., & Schouten, B. (2011). *Handbook of nonresponse in household surveys*. Hoboken, NJ, USA: John Wiley & Sons.
- Beullens, K., Loosveldt, G., Denies, K., & Vandenplas, C. (2016). *Quality matrix for the European Social Survey, Round 7*. Retrieved from https://www.europeansocialsurvey.org/docs/round7/methods/ESS7_quality_matrix.pdf
- Biemer, P. B. (2010). Total Survey Error: Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5), 817-848. <https://doi.org/10.1093/poq/nfq058>
- Chapman, D. W. (2003). To substitute or not to substitute – that is the question. *Survey Statistician*, 48, 32-34.
- Cornesse, C., & Bosnjak, M. (2018). Is there an association between survey characteristics and representativeness? A meta-analysis. *Survey Research Methods*, 12(1), 1-13. <https://doi.org/10.18148/srm/2018.v12i1.7205>
- Couper, M. P., & de Leeuw, E. D. (2003). Nonresponse in cross-cultural and cross-national surveys. In J. A. Harkness, F. J. R. van de Vijver, & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp. 157-177). New York, NY, USA: Wiley.
- de Leeuw, E., Hox, J., & Luiten, A. (2018). *International nonresponse trends across countries and years: An analysis of 36 years of labour force survey data*. Retrieved from <https://surveyinsights.org/?p=10452>
- Elliot, D. (1993). The use of substitution in sampling. *Survey Methodology Bulletin*, 33, 8-11.
- EQLS. (2013). *3rd European Quality of Life Survey* (Sampling report EU27 and non-EU countries). Retrieved from http://doc.ukdataservice.ac.uk/doc/7316/mrdoc/pdf/7316_3rd_eqls_sampling_report.pdf
- European Commission. (2017). *Eurobarometer 87.3, 2017* (Study description in DDI format DDI-lifecycle 3.2) [Data sets, questionnaires and other documents]. <https://doi.org/10.4232/1.12847>
- European Social Survey. (2008). *ESS4 - 2008 documentation report* (Ed. 5.5). Retrieved from https://www.europeansocialsurvey.org/docs/round4/survey/ESS4_data_documentation_report_e05_5.pdf

- European Social Survey. (2010). *ESS5 - 2010 documentation report* (Ed. 4.2). Retrieved from https://www.europeansocialsurvey.org/docs/round5/survey/ESS5_data_documentation_report_e04_2.pdf
- European Social Survey. (2017). *ESS 8 – 2016 documentation report* (Ed. 2.1). Retrieved from https://www.europeansocialsurvey.org/docs/round8/survey/ESS8_data_documentation_report_e02_1.pdf
- European Social Survey. (2019, June 6). *Round 10 survey specification for ESS ERIC member, observer and guest countries* (Version 1). Retrieved from http://www.europeansocialsurvey.org/docs/round10/methods/ESS10_project_specification.pdf
- European Values Study. (2016). *European values study 2008: Integrated dataset (EVS 2008)* [Data sets, questionnaires, code books and other documents]. GESIS Data Archive. <https://doi.org/10.4232/1.12458>
- Gaziano, C. (2005). Comparative analysis of within-household respondent selection techniques. *Public Opinion Quarterly*, 69(1), 124-157. <https://doi.org/10.1093/poq/nfi006>
- Grauenhorst, T., Blohm, M., & Koch, A. (2016). Respondent incentives in a national face-to-face survey: Do they affect response quality? *Field Methods*, 28(3), 266-283. <https://doi.org/10.1177/1525822X15612710>
- Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L. E., & Mohler, P. P. ... Smith, T. W. (2010). *Survey methods in multinational, multiregional, and multicultural contexts*. Hoboken, NJ, USA: John Wiley & Sons.
- Heeringa, S. G., & O’Muircheartaigh, C. (2010). Sample design for cross-cultural and cross-national survey programs. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, & T. W. Smith (Eds.), *Survey methods in multicultural, multinational, and multiregional contexts* (pp. 251-267). Hoboken, NJ, USA: John Wiley & Sons.
- Hubbard, F., Lin, Y., Zahs, D., & Hu, M. (2016). Sample design. In *Guidelines for Best Practice in Cross-Cultural Surveys* (pp. 99-149). Ann Arbor, MI, USA: Survey Research Centre, Institute for Social Research, University of Michigan. Retrieved from <http://www.ccsr.isr.umich.edu/>
- International Social Survey Programme. (1999). *International Social Survey Programme: Social inequality III - ISSP 1999* [Data sets, questionnaires and other documents]. GESIS Data Archive. Retrieved from <https://dbk.gesis.org/dbksearch/sdesc2.asp?no=3430&db=E&tab=3>
- International Social Survey Programme. (2013). *International Social Survey Programme Study Monitoring 2011 Health* (Report to the ISSP General Assembly on monitoring work undertaken for the ISSP by the Methodology Committee). Retrieved from https://dbk.gesis.org/dbksearch/file.asp?file=ZA5800_mr.pdf
- International Social Survey Programme. (2017). *International Social Survey Programme: Work orientations IV - ISSP 2015* [Data sets, questionnaires, code book and other documents]. GESIS Data Archive. <https://doi.org/10.4232/1.12848>
- Jowell, R. (1998). How comparative is comparative research? *The American Behavioral Scientist*, 42(2), 168-177. <https://doi.org/10.1177/0002764298042002004>

- Koch, A. (2018). Within-household selection of respondents. In T. P. Johnson, B. Pennell, I. A. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods* (pp. 93-111). Hoboken, NJ, USA: John Wiley & Sons. <https://doi.org/10.1002/9781118884997.ch5>
- Koch, A., & Blohm, M. (2006). Fieldwork details in the European Social Survey 2002/2003. In J. Harkness (Ed.), *Conducting cross-national and cross-cultural surveys: Papers from the 2005 meeting of the International Workshop on Comparative Survey Design and Implementation (CSDI)* (pp. 21-52). Mannheim, Germany: GESIS-ZUMA. Retrieved from <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49740-1>
- Kohler, U. (2007). Survey from inside: An assessment of unit nonresponse bias with internal criteria. *Survey Research Methods*, 2(1), 55-67. <https://doi.org/10.18148/srm/2007.v1i2.75>
- Kołczyńska, M., & Schoene, M. (2018). Survey data harmonization and the quality of data documentation in cross-national surveys. In T. P. Johnson, B. Pennell, I. A. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods* (pp. 963-984). Hoboken, NJ, USA: John Wiley & Sons.
- Lohr, S. L. (2008). Coverage and sampling. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 97-112). New York, NY, USA: Taylor & Francis.
- Lynn, P., Gabler, S., Häder, S., & Laaksonen, S. (2007). Methods for achieving equivalence of samples in cross-national surveys. *Journal of Official Statistics*, 23(1), 107-124.
- Lynn, P., Japac, L., & Lyberg, L. (2006). What's so special about cross-national surveys? In J. A. Harkness (Ed.), *Conducting cross-national and cross-cultural surveys: Papers from the 2005 meeting of the International Workshop on Comparative Survey Design and Implementation (CSDI)* (pp. 7-20). Mannheim, Germany: GESIS ZUMA. Retrieved from <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49740-1>
- Marker, D. A., & Stevens, D. L., Jr. (2009). Sampling and inference in environmental surveys. In D. Pfeffermann & C. Rao (Eds.), *Sample surveys: Design, methods and applications* (Vol. 29A, pp. 487-512). Amsterdam, Netherlands: Elsevier.
- Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111-130. <https://doi.org/10.21500/20112084.857>
- Mohler, P. Ph. (2019, March 18-20). *Why do we prefer second best quality indicators?* Paper presented at the Comparative Survey Design and Implementation Workshop, 2019, Warsaw, Poland. Retrieved from <https://csdiworkshop.org/wp-content/uploads/2019/05/Why-Do-We-Prefer-Second-Best-Quality-Indicators.pdf>
- Mohler, P. P., Pennell, B.-E., & Hubbard, F. (2008). Survey documentation: Towards professional knowledge management in sample surveys. In E. de Leeuw, J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 403-420). New York, NY, USA: Taylor & Francis.

- Park, A., & Jowell, R. (1997). *Consistencies and differences in a cross-national survey: The International Social Survey Programme (1995)*. London, United Kingdom: SCPR.
- Smith, T. W. (2002). Developing nonresponse standards. In R. Groves, R. Dillman, D. A. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 27-40). Hoboken, NJ, USA: John Wiley & Sons.
- Smith, T. W. (2018). Improving multinational, multiregional, and multicultural (3MC) comparability using the Total Survey Error (TSE) paradigm. In T. P. Johnson, B. Pennell, I. A. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods* (pp. 13-43). Hoboken, NJ, USA: John Wiley & Sons. <https://doi.org/10.1002/9781118884997.ch2>
- Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). *Improving survey response: Lessons learned from the European social survey*. New York, NY, USA: John Wiley.
- Stoop, I., Koch, A., Halbherr, V., Loosveldt, G., & Fitzgerald, R. (2016). *Field procedures in the European Social Survey Round 8: Guidelines for enhancing response rates and minimising nonresponse bias*. Retrieved from http://www.europeansocialsurvey.org/docs/round8/methods/ESS8_guidelines_enhancing_response_rates_minimising_nonresponse_bias.pdf
- Sturgis, P., Williams, J., Brunton-Smith, I., & Moore, J. (2017). Fieldwork effort, response rate, and the distribution of survey outcomes: A multilevel meta-analysis. *Public Opinion Quarterly*, 81(2), 523-542. <https://doi.org/10.1093/poq/nfw055>
- Vandenplas, C., Loosveldt, G., & Beullens, K. (2015, September). *Better or longer? The evolution of weekly number of completed interviews over the fieldwork period in the European Social Survey*. Paper presented at the Nonresponse Workshop, 2015, Leuven, Belgium.
- Vehovar, V. (1999). Field substitution and unit nonresponse. *Journal of Official Statistics*, 15(2), 335-350.
- von der Lippe, E., Schmich, P., Lange, C., & Koch, R. (2011). Advance letters as a way of reducing non-response in a National Health Telephone Survey: Differences between listed and unlisted numbers. *Survey Research Methods*, 5(3), 103-116. <https://doi.org/10.18148/srm/2011.v5i3.4657>
- Wuyts, C., & Loosveldt, G. (2019, May 29). *Quality matrix for the European Social Survey, Round 8* (Overall fieldwork and data quality report). Retrieved from https://www.europeansocialsurvey.org/docs/round8/methods/ESS8_quality_matrix.pdf



Methodology is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology Information (ZPID), Germany.