Improving Knowledge Production in Comparative Survey Research: Cross-Using Data from Four International Survey Programmes*

BRINA MALNAR**
University of Ljubljana

LORNA RYAN
City, University of London

Abstract: In the context of flourishing cross-national and multi-level research, the variety of macro and micro data sources available to researchers have evolved into an interdependent ecosystem of social indicators. Focusing on four comparative social surveys, this paper examines the extent to which secondary data users take advantage of a range of complementary data sources to broaden the breadth or strengthen the robustness of their research. Using two Google Scholar-based datasets of 2789 and 796 publications, we find that, despite the complex equivalence issues in comparative survey research, users combine data to a considerable extent, aiming to increase conceptual, geographic, and temporal coverage and cross-validate findings. Selecting the example of the European Social Survey, 183 journal articles are qualitatively examined to identify specific epistemic gains attained by analysts when combining ESS survey data with data from other comparative programmes. The strategy involves risks, emanating from either analysts' own misjudgements or arising from the wider issues of comparability and transparency in cross-national survey research. However, a number of data harmonisation platforms have recently emerged that may facilitate the standardisation of measures across surveys, augmenting the possibilities for future theory development and research.

Keywords: comparative surveys, social indicators, academic users, data ecosystem, multi-level analysis

---

** Direct all correspondence to: Brina Malnar, Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia, e-mail: brina.malnar@fdv.uni-lj.si; Lorna Ryan, School of Arts and Social Sciences, City, University of London, Northampton Square, London, EC1V 0HB, UK, e-mail: Lorna.Ryan.1@city.ac.uk.

Cross-national surveys in the social science data ecosystem

In 1962 and 1970, Rabier pioneered the first European multi-nation surveys to investigate mass attitudes towards European integration. The surveys involved a limited number of western European countries and their success led to the launch of the Eurobarometer in 1974 [Norris 2007]. In the 1980s, large general-purpose cross-national surveys began to emerge around the globe with the specific aim to facilitate systematic cross-national analysis. These are now widely regarded as some of the main data sources on contemporary societies [Kuechler 1998; Mochmann 2008; Smith 2019]. They facilitate comparisons of large numbers of countries in a cross-sectional and cross-time perspective, a format that encourages analysts to determine causality or to test developmental societal theories, such as modernisation theory [Andress et al. 2019]. They offer ample possibilities to study public opinion in a variety of institutional and societal contexts in which analysts can quantify the extent to which differences in outcomes reflect differences in country-specific features, such as demographic structure, public policies, labour market characteristics, and many others [Norris 2009; Bryan and Jenkins 2015]. In this respect, comparative social research is sometimes considered the equivalent of an experimental research design in the natural and physical sciences [Mochmann 2008].

Making the most of these possibilities, an increasing share of studies based on cross-national surveys employ multi-level design, combining individual-level micro data with macro indicators, seeking explanations of social phenomena in interaction between actors and institutions [Andress et al. 2019; Smith 2019]. This strategy is facilitated by the growing availability of systematically collected country-level data, coupled with matching micro-level data that saturate a full multi-level analysis [Quandt and Luijkx 2015]. National statistical offices are among the largest providers of comparable macro indicators, along with numerous international organisations such as the UN, the World Bank, the OECD, Eurostat, the Global Health Observatory, and the IMF [Vezzoni 2015; UNECE 2017], while cross-national survey programmes usually provide comparable micro-level data.

This paper builds on the assumption that, from the perspective of secondary data users, a variety of macro and micro sources represent an ever more interdependent 'ecosystem' of indicators, rather than a collection of discrete segments. The concept of an ecosystem has been adopted by several non-biological academic disciplines in recent years, so that we now have, for example, software ecosystems, mobile application ecosystems, innovation ecosystems, and strategy research ecosystems [Oh et al. 2016; Seppanen et al. 2017]. Most recently, this has extended to an ecosystem of technologies for social science research [Duca and Metzler 2019]. While the term is often used metaphorically and the theoretical conceptualisations outside biology are still being developed, it generally refers to a group of interacting actors that depend on each other's activities [Jacobides et al. 2018]. Davies [2012] describes the elements in the open data ecosystem as autonomous and self-organising components, linked together in local and global

feedback loops. In a similar vein, EU science policy documents refer to research infrastructures as originally stand-alone undertakings that are 'becoming more and more part of a connected ecosystem forming a unique resource for advanced research' [ESFRI 2018: 18]. Given the complexity and distributed nature of the numerous data sources used by social science analysts, we adopt this general notion of an ecosystem to conceptualise their mutually enriching interconnectedness.

While we also investigate the extent to which micro and macro indicators are combined in academic publications, our main interest is in the bottom-up synergies between four comparative surveys which count among the leading sources of individual-level comparative findings. The World Values Survey (WVS) and the European Values Study (EVS) are global longitudinal survey research programmes that study changing values and their impact on social and political life. The International Social Survey Programme (ISSP) is a cross-national collaboration programme that conducts annual surveys on diverse topics relevant to the social sciences, and the European Social Survey (ESS) is an academically driven cross-national survey that measures the attitudes, beliefs, and behavioural patterns of diverse populations in Europe and aims to spread higher standards of rigour in cross-national research.[1] Having fielded their first waves in 1981 (WVS and EVS), 1985 (ISSP), and 2002 (ESS), the four programmes have generated rich time series of general-purpose social indicators, which have been widely used for longitudinal comparative research by generations of scholars in a variety of social scientific fields [Heath et al. 2005; Norris 2009].

The decision to focus on this group of data providers was motivated by two considerations: first, the availability of a large bibliographic dataset based on European Social Survey publications [Malnar 2019], which, among other bibliometric variables, documents cross-survey usage; second, the overlapping character of the four surveys in terms of themes, country membership, and temporal span creates ample possibilities for cross-usage. All four are general population cross-sectional programmes that produce comparative data on contemporary societies and deposit the data in public archives as a 'public good' for members of various academic communities to use in secondary analysis [Jowell et al. 2007; Vezzoni 2015]. Over recent decades, these four time series have become essential sources for academic research, bringing with their use the unavoidable limitations that come with indirect data collection. The trade-offs between the convenience of having ready data produced by someone else and the effort of dealing with data collected independently from the formulation of the research question are well-known [Damian et al. 2019: 4]. For example, the data may not cover all the concepts, the entire geographic region, the exact years, or the specific population that a researcher is interested in [Johnston 2014; Gonçalves 2016]. In this respect 'data completeness' is a useful concept, one that quantitative social scientists usually

---

[1] Survey websites: www.worldvaluessurvey.org (WVS); www.europeanvaluesstudy.eu (EVS); www.issp.org (ISSP); www.europeansocialsurvey.org (ESS).

associate with missing data [Faniel et al. 2016], but that can be extended to refer to issues of conceptual, geographic, and temporal coverage.

From the perspective of secondary analysts, the key element of completeness is conceptual relevance, determining how a particular data source is able to fill existing knowledge gaps [Müller-Bloch and Kranz 2015]. Any survey's conceptual range is necessarily limited, and even more so in cross-national designs, where only certain subjects and only certain aspects of those subjects can successfully be measured cross-nationally [Jowell 1998; Lagos 2008]. Whereas the four comparative surveys share a number of concepts, they are also thematically distinct. The EVS and WVS were developed primarily to test the modernisation and secularisation theory and process, which is why they include fewer indicators of behaviours and more indicators of values, while the ISSP and ESS, with their (partial) modular format, were designed to enable the testing of a set of thematic theoretical approaches and include more behavioural items (for more on the historical development of large comparative surveys see Heath et al. [2005]; Norris [2009]; Groves [2011]; Hadler et al. [2015]; Quandt and Luijkx [2015]). Clearly, each individual survey has a narrower conceptual range than all four combined, which means that there may be improvements in data completeness that can be achieved by their cross-use.

Another limitation faced by secondary data analysts is geographic range. Although they partially overlap, the four surveys differ substantially in the size of their membership, which directly translates into comparative scope. Two of them are global and two European, but with very different coverage of their target areas. The WVS is sometimes referred to as covering 90% of the global population [Haerpfer and Kizilova 2016: 723] and the EVS in its recent waves included almost all European countries. The ISSP and the ESS cover considerably smaller shares of their target regions, with methodological rigour being the main rationale behind this approach. A discussion of the complex relation between the size of membership and equivalence issues is not the aim of this article. Attention here is instead focused on how secondary analysts handle the consequences of these two inclusion philosophies. When dealing with various aspects of data limitation, such as (a lack of) conceptual or geographic scope, the conventional solution is to adjust the research question to achieve a better fit with the chosen data set [Doolan and Froelicher 2009]. This study explores the prevalence of an alternative 'ecosystem' approach, in which analysts can respond to issues of incompleteness by taking advantage of the co-existence of multiple survey resources and their combined conceptual, temporal, and geographic breadth.

Besides coverage issues, the validation of findings is another vital aspect of comparative survey data use to which the ecosystem notion can be applied. Methodological literature documents an almost endless number of equivalence issues in comparative research, and highlights how designing successful cross-national studies is an elaborate and methodologically complex process in which there is much more potential for biases and errors than there is in single-nation

surveys [Lyberg et al. 2019; Smith 2019; Lynn et al. 2006; Damian et al. 2019]. Ensuring measurement equivalence when there are multiple distributed actors and many social and cultural contexts remains a challenge [Briceno-Rosas et al. 2020] and may not result in the desired level of quality, as the 'data from different countries were possibly collected using different methods, under different conditions, and in different cultural contexts' [Quandt and Luijkx 2015: 795]. The validation of results, always considered an element of good research practice [Mathison 1988], is therefore even more vital when we analyse comparative data. In this respect, the co-existence of multiple comparative time series is often viewed as an opportunity for replication and generalisation, as converging findings provide much stronger confidence in conclusions than comparisons between countries at a single point in time do [Tengö et al. 2014; Norris 2009; Heath et al. 2005]. This is something that our study will also examine.

While many publications in the field of comparative survey research discuss issues of governance, quality control, transparency of documentation and procedures, questionnaire development, harmonisation, translation, and fieldwork management [Jowell 1998; Harkness 1999; Lynn 2003; Lyberg et al. 2019; Hadler et al. 2015; Pennell et al. 2017; Lindstrøm and Kropp 2017], few studies empirically investigate the knowledge production side of comparative programmes [e.g. Damian et al. 2019], and we found none that study combined data use. This theme is more commonly discussed in the medical literature, particularly in relation to systematic reviews, based on the entire population of relevant studies [Oliver et al. 2005; Petticrew and Roberts 2006; Harden 2010]. Apart from the ubiquitous experimental method, it is also customary for medical surveys to be combined to improve target population coverage when samples are small, incomplete, or of poor quality [Schenker et al. 2002; Dong et al. 2014]. Judging from the findings of Schenker et al. [2002] and Dong et al. [2014], this practice often raises concerns about validity issues. Roberts and Binder [2009] find that the separate and pooled approaches to estimation lead to different results and caution that it is often not appropriate to combine similar data from more than one survey. Others highlight inconsistencies in question wording but conclude that 'in spite of complicating issues, combining information from multiple surveys appears to be potentially useful and an important area for further research' [Schenker and Raghunathan 2007], a point that seems equally relevant for social survey research.

The aim of our study was to determine the analytical benefits that support the cross-use of datasets in comparative survey research, a strategy that we term the 'ecosystem' approach. Specifically, we examine how analysts take advantage of the co-existence of the four comparative social survey programmes and their combined conceptual, temporal, and geographic breadth when answering research questions, dealing with issues of data incompleteness, and validating findings.

Data and measures

In order to address our research objectives, we relied on quantitative and qualitative information and two original databases.

Dataset 1 – publications based on the ESS

Our main data source is the European Social Survey bibliographic database [Malnar 2019], updated annually as part of an ongoing bibliographic monitoring exercise.[2] The data collection process uses Google Scholar to search for the key-phrase 'European social survey' in the past publication year. Google Scholar is the search engine of choice because it indexes scholarly literature across an array of publication formats and disciplines and thus attempts to index the totality of the realm of scientifically relevant documents, such as articles, books, chapters, reports, and theses [Mayr and Walter 2007; Ware and Mabe 2012; Harzing 2012]. An ESS-based publication is defined as any type of academic publication in the English language that used at least one ESS item in its primary analysis, a fact established through a case-by-case review of abstracts and texts. The version of the ESS bibliographic database we used contained 4914 records for the period 2004–2019, which were entered into an SPSS data file, and 2789 full texts (1821 of them journal articles), acquired through open-access and subscription-access schemes. This makes it a unique bibliographic resource for monitoring the academic usage of a major cross-national survey.

The ESS dataset is suitable for the purpose at hand as it contains a set of variables that measure the use of other data sources in the 2789 downloaded texts. With respect to macro indicators, it documents the presence of GDP and GINI, two widely employed measures of country-level affluence and inequalities, and the use of multi-level analysis, an indirect but robust indicator of the presence of macro measures in respective explanatory models. In addition, it records the use of other national or cross-national survey sources, as well as the specific presence of WVS, EVS, or ISSP data. In this way, we identified 304 publications in which ESS data are combined with data from the other three comparative programmes, and these were the key subset of publications we used in our study.

In order to examine the use of multi-level analysis and highlight thematic overlap between the four surveys, we used the topic variable, which was derived from 4914 ESS publications. The open coding approach, common in qualitative research [Blair 2016], involved a human coder who identified and catalogued the themes the authors addressed based on the publications' titles, abstracts, and

---

[2] The authors are directly involved in producing the bibliographic data file, which is to be made public and accessible from the ESS ERIC webpage by the end of 2021. Currently, all data that support the findings of this study are available from the authors upon request.

texts, an exercise that resulted in a list of 26 main topics. The same approach was used to compile a list of topics in Dataset 2. Finally, the full texts of 304 publications that combine ESS data with data from the three other comparative surveys were reviewed to identify the main analytical reasons behind this strategy. Six motivations for cross-using data were identified and a six-category variable was constructed to quantify their presence across the 304 publications by assigning them appropriate codes (see Table 3 in the results section).

One limitation of the approach adopted, owing to the availability of Dataset 1 is that several parts of the analysis are focused on one of the four surveys, the European Social Survey. Nonetheless, considering the similarities in their mission and target academic audiences, findings on patterns of cross-survey use are likely to be relevant for all four surveys.

Dataset 2 – publications based on the WVS, the EVS, and the ISSP

In order to examine the thematic contrasts and complementarities between publications based on the four comparative surveys, as well as the rationale for using multiple survey waves, we constructed a supplementary dataset consisting of English-language journal articles that used data from the WVS, the EVS, and the ISSP. We used the same methodology for this dataset: a keyword search for the phrases 'World values survey', 'European values study', and 'International social survey programme' across the Google Scholarplatform, combined with human reviewing for primary data use. To limit the work effort, the search was restricted to journal articles and to selected publication years. The supplementary dataset thus consists of 796 articles, among them 193 WVS articles for the year 2016, 143 EVS articles for the years 2015 and 2016, and 261 ISSP articles for the years 2011, 2014, and 2016. EVS articles were sourced for two consecutive years to increase the sample size and ISSP articles from three scattered years to lessen the potential effect of fielding schedule on the structure of topics, knowing that modules recently deposited in data archives tend to be analysed more than older ones. The 199 ESS articles for the publication year 2016 were taken from the original ESS bibliographic database

In addition to the two datasets, our study draws on other publicly available information. To summarise and compare survey characteristics such as time span, fielding frequency, and geographic coverage, we gathered information from the official homepages of each survey.

Results

Combining ESS data with other micro- and macro-data sources

Using the ESS dataset (Dataset 1), we first examined the overall practice of combining micro- and macro-data sources in publications with ESS primary data use. This was found to be considerable (Table 1). GDP, the most used macro indicator, is referred to in about 37% of ESS publications, either as part of the analytical models or in the text, and a third of ESS-based journal articles use multi-level analysis, which suggests the use of macro indicators.[3] The presence of other micro data is also significant, with about a third of ESS publications containing data from additional national sources, but more often from other comparative sources, or both. The prevalence of combining data sources is best demonstrated by the fact that among the 1821 ESS journal articles analysed, the total share of those that used either other micro data, GDP or GINI macro indicators, or multi-level analysis is 69.9%. The share would undoubtedly increase if the bibliographic dataset documented the presence of other macro indicators besides GDP and GINI. This suggests that the majority of ESS-based publications rely on supplementary data sources. While this finding is not unexpected, the picture empirically corroborates and quantifies the notion that there is a rich landscape of interacting social indicators, or 'ecosystem', in international academic publications.

In order to refine our insight into the epistemological benefits of combining the ESS with macro indicators, we examined the use of multi-level analysis across individual topics (Table 2). We can see considerable variation in topic use, with some topics being much more frequently analysed in interactions between individual-level and contextual conditions (e.g. national-level policies or institutions) than others. On one hand, the structure of causality is likely to make some topics more reliant on the inclusion of high-quality comparative macro indicators, while other topics may rely primarily on individual-level explanations. On the other hand, the availability of standardised comparative macro indicators relevant for individual topics may also determine the scope of multi-level analysis use. In any case, we found this highly relevant comparative approach evident in at least 20% of ESS international publications across every topic and often more. The implications for knowledge production are considered below in the final section.

Combining ESS data with data from three comparative surveys

In the rest of our study, we narrowed our focus to examining the analytical motivations behind combining ESS data with other sources of comparative micro data. Specifically, we employed a combination of quantitative and qualitative approaches to obtain a systematic insight into 304 publications in which academic

---

[3] The use of multi-level analysis in the ESS bibliographic file is only available for journal articles.

Table 1. The share of publications that combine the ESS survey with other data sources

	%	N
Publications citing the GDP indicator	36.7	1023
Publications citing the GINI indicator	11.1	311
Journal articles using multi-level analysis*	34.8	634
Publications combining ESS data with other survey data	32.0	893
Cross-national survey(s)	18.2	508
National survey(s)	10.7	298
Both	3.1	87

Note: N = 2789 ESS publications; * N = 1821 ESS journal articles.

Table 2. Using multi-level analysis across topics in ESS journal articles

Topic	%	Topic (continued)	%
Nation, ethnicity	48.0	Social inequalities	28.7
Family, children, partners	41.5	Education	26.7
Welfare	40.5	Subjective well-being	26.1
Immigration	36.4	Religion, religiosity	25.4
Health	36.2	Culture, values	25.3
Gender issues	33.1	Crime	25.2
Work, employment	31.1	Social capital	25.1
Ageing, age groups	29.2	Economic issues	24.8
Politics	28.7	Civil society, volunteering	20.8

Notes: [a] Only topics with 100 or more publications are included; N = 1821.

authors pair ESS data with data from three other comparative programmes: 186 joint publications with WVS data, 149 with EVS data, and 99 with ISSP data. In 72.7% of the cases ESS is combined with one survey, while in the rest of the publications the data from two or all three surveys are present.

First, we qualitatively reviewed the 304 texts for the analytical reasons behind the combining of data and, as already noted, identified six categories that by and large saturate the broad rationale behind these strategies. Coding each publication into one or more categories (there can be more than one rationale for combining data sources), we obtained a summary quantitative picture (Table 3).

'Improving conceptual coverage' and the 'validation of findings' are the most frequent incentives for combining datasets. This echoes the debate about

Table 3. Analytical reasons for combining ESS data with data from the WVS, the EVS, and the ISSP (%)

	WVS (N = 166)	EVS (N = 81)	ISSP (N = 57)	Total (N = 304)
Reason for combining data [a]	(%)			
Combining concepts, indicators	51.8	64.2	43.9	53.6
Validation, robustness checks	34.3	27.1	43.9	34.2
Adding non-European countries	26.5	0.0	17.5	18.1
Adding time points	12.0	19.7	14.0	14.5
Pooling samples across surveys	1.2	9.9	5.2	4.3
Adding European countries	3.0	2.5	0.0	2.3

Note: [a] All relevant reasons were coded for each publication.

limitations in the scope of subjects that can be measured cross-nationally, as well as the need for validity checks in the face of significant potential for biases in cross-national research. In about a third of cases, analysts combine comparative data to tackle issues of geographic coverage, mostly to add non-European regions into the comparison, while in 15% of cases the rationale was to increase the number of measurement points. Pooling datasets and treating them as a single sample is the least frequent reason for combining data. The table also reveals specific complementarities between surveys – for example, combining indicators is most frequent in ESS – EVS combinations while robustness checks are most prevalent in ESS – ISSP combinations. This is not surprising considering that the two surveys have developed a number of similar thematic modules and seem to constitute the most suitable pair for verification of findings.

While this quantitative picture provides a valuable general insight, we sought to obtain a more detailed understanding of the practice of combining comparative data sources and the analytical motives for this. For this purpose, we qualitatively examined 183 journal articles from the 304 combined publications, assuming that scientific rigour and comparable format make this subset the most suitable choice for systematic examination. Following the order of reasons listed in Table 3, the next sections present an overview of our findings.

Improving conceptual coverage

We identified four typical strategies in publications where improving conceptual coverage was the main motivation for combining comparative survey sources:

Adding aggregate indicators into the explanatory models (28 articles). The most widespread approach identified was the addition of aggregate country-level in-

dicators or context conditions from other comparative surveys to complement analytical (regression) models based on the main ESS dataset. These indicators usually represented concepts or sub-concepts not present in the ESS, such as shares of materialists or post-materialists, indicators of gender and family roles or of national and supranational identity, and environmental values. In statistical terms, aggregate indicators come in many forms, such as country mean scores, population shares, factor scores, composite indicators derived from several items, estimates averaged across several surveys, and standard deviations.

Descriptive completeness (25 articles). Another common practice is a descriptive combination of indicators, with authors citing marginals from other surveys to support their general discussion or argument. In more elaborate instances, analysts combine blocks of indicators from various surveys in the form of charts or tables, often through various sub-sections in the text. These indicators usually cover a variety of complementary but not identical measures and are used to construct a more comprehensive descriptive picture of the phenomenon under investigation, such as family change, public opinion on gay rights, levels of social capital before and after an economic crisis, and similar phenomena.

Combined causal design (11 articles). Using this strategy, authors present two or more explanatory models or studies, based on two or more comparative datasets, addressing subjects that are complementary but not identical. The aim is to establish different dimensions of causality or to test different but related hypotheses, usually in the form of separate regression models, and use the combined result from all the datasets to answer a general research question. The main purpose is to combine diverse indicators, not to validate findings.

Country case studies (9 articles). Here analysts observe a phenomenon in a single country, using several (comparative) sources to address different aspects or dimensions, frequently using NUTS-level analysis in combination with regional administrative data. Sometimes no cross-country comparisons are made and cross-national datasets are simply combined for a national study. Other times the main dataset is a national survey or a primary study, which is the basis for an explanatory model, enhanced by comparative indicators from several comparative surveys for international comparisons.

In all cases, secondary users gain an epistemic advantage from the fact that the combined conceptual coverage across the four comparative surveys is greater than that arising from any single use.

Validating findings

The second most frequent reason for combining ESS data with data from the three other surveys was to validate findings, a strategy made possible by a number of overlapping themes, concepts, and, to a limited degree, individual indicators. To highlight their thematic complementarities, Table 4 presents the structure of topics analysed in journal articles based on each of the four comparative surveys

(Dataset 2). In the top two rows are topics frequently analysed using data from all four surveys, most notably politics. The middle rows consist of topics more often found in publications based on pairs of surveys, such as culture or religion in both values studies, while the bottom part presents topics that are more prevalent in one survey's publications, like immigration in the case of the ESS or the environment and social inequalities in the case of the ISSP. Topics with a greater than 10% share among publications based on each survey are marked with an asterisk.

Thematic correspondence measured in this way is clearly stronger for some survey pairs that others, but the findings also suggest that all four programmes offer opportunities for studying a large variety of similar topics and issues. Accordingly, the qualitative reviewing identified four validation strategies:

Cross-validating theoretical models or estimates (30 articles). The largest group are publications where hypotheses are corroborated across more than one dataset. Analysts usually present separate explanatory models based on two or more comparative surveys, examining the consistency of associations across analysed countries. This strategy is not always straightforward as replicating models across

Table 4. Distribution of topics in journal articles based on the four surveys (%)

Topic [a]	WVS (N = 193)	EVS (N = 143)	ESS (N = 199)	ISSP (N = 262)
		(%)		
Politics, democracy	14.5*	16.8*	25.1*	14.5*
Economy, recession	15.6*	13.3*	13.6*	9.2
Culture, values	24.9*	17.5*	6.0	5.3
Social capital	11.9*	10.5*	7.0	3.8
Religion	10.9*	14.7*	3.0	8.4
Welfare, policies	6.2	7.0	10.6*	21.0*
Work	4.7	6.3	15.1*	12.2*
SWB-QOL[b]	15.0*	8.4	12.6*	6.1
Family	5.7	10.5*	7.5	16.8*
Immigration	5.2	3.5	14.1*	6.1
Social inequalities	1.6	6.3	8.5	13.0*
Environment	5.8	0.7	3.0	10.7*
Health	4.7	4.9	10.6*	5.3
Gender issues	8.8	10.5*	7.5	5.3

Notes: N = 797; [a] up to two topics were coded per article; only topics investigated in 50 or more articles are presented; [b] subjective well-being and quality of life.

surveys may require the use of strategies such as scaling the models down to the smallest common denominator, using corresponding but not identical indicators, presenting regression tables with missing cells, or carrying out partial replication by cross-validating only some associations. In most cases the results are found to be consistent across surveys, but this may also be due to the known issue of publication bias [e.g. Franco et al. 2014], with authors who find conflicting results opting out of publishing their studies or parts of them. However, some authors report findings that are (partially) contradictory and suggest that fellow researchers should be cautious when drawing conclusions based on a single dataset.

Sensitivity testing (15 articles). Like the first group, these publications retest their main hypotheses or explanatory model using an alternative dataset, but with the explicit aim of controlling for measurement specification. They compare matching concepts or associations measured in different ways in terms of phrasing or scales, examine the effect of ambiguous wordings and similar. Examples include prospective and retrospective indicators of voting behaviours, alternative measures of multiculturalism, measures of globalisation exposure, etc. Again, the results are mostly found to be robust, and show congruency across different measures. But there are some notable exceptions where authors find no consistent patterns across surveys and conclude that the results are dependent on the choice of dataset, which is obviously a problem.

Validating trends (3 articles). In some cases, analysts use multiple comparative surveys to cross-validate trends across a set of countries for the same or a similar time period – for example, secularisation trends or the convergence of values.

Descriptive corroboration (3 articles). The least ambitious approach employed is that of seeking a descriptive convergence of results, typically by comparing the values of corresponding individual indicators from several surveys, such as trust in the legal system, party membership, or a belief that immigrants make crime problems worse.

In addition to these direct validation strategies, a number of the authors who used ESS data reported that they were replicating analyses from previous articles using WVS, EVS, or ISSP data. Such cases are not documented in the ESS bibliographic dataset and are therefore not part of our review. In this respect, the extent of cross-survey validation of findings in our study is underestimated.

Expanding the geographic scope

In about 20% of publications (Table 3) the main rationale for combining ESS data with the three other surveys is to add non-European countries into the comparison or, very marginally so, to add European countries. We identified three typical patterns:

Inserting or combining global country aggregates (11 articles). Because of its comprehensive thematic coverage and larger national samples, the ESS is probably the leading cross-national survey source for analysing individual-level out-

comes related to immigration. This results in a set of publications in which country aggregates from global comparative surveys, particularly the WVS, are used in ESS-based explanatory models to test the acculturation effects of immigration. In a typical design, estimates in ESS immigrant sub-samples are compared with identical or similar aggregate indicators in their countries of origin – for example, levels of trust or civic participation, attitudes towards homosexuality, cultural distance, and others.

Seeking transatlantic comparisons (7 articles). These publications compare European (ESS) countries with Western transatlantic regions (North America, Australia, and New Zealand) to look for convergence or contrasts. The United States, particularly with its specific value structure, is often an important comparison point for topics such as welfare chauvinism, preferences for redistribution, desired levels of immigration, and others. It should be noted though that a number of authors who seek to make this specific comparison chose to combine ESS data with US national datasets such as the General Social Survey.

Observing global patterns (6 articles). The most ambitious approach identified includes publications that seek to test the universal validity of theories or findings, with analysts combining several comparative datasets to observe relevant relationships across a range of countries from around the globe. Examples include examining value trends, system responsiveness, changes in unionisation, popular support for health care, motivational selectivity of migrants, and others.

Again, in all cases, secondary users seek to benefit from the greater geographic coverage across the four surveys.

Expanding the time scope and scale

The four datasets differ considerably in terms of their time-series length: the EVS, the WVS, and the ISSP date back to 1981 and 1985, while the ESS was first fielded in 2002. Their measurement frequency is also very different, with the ESS being fielded biannually, the WVS roughly in 5-year intervals, and the EVS and ISSP in close to 10-year intervals.[4] We identified three strategies related to the time dimension.

Combining time series (13 articles). In the majority of cases, analysts merge two or more survey series, usually partially overlapping, to create a single and more dense string of measurements, sometimes observing short-term, medium-term, and long-term dynamics on different datasets, depending on their time span and fielding frequency. Such strategies enable the observation of shared indicators or relationships across a longer time period, often across several decades

---

[4] In the case of the ISSP we refer to the typical fielding frequency of the same ISSP substantive module, which usually, though not always, defines the frame of interest for academic users, while the demographic section is fielded annually.

and in a number of countries. Trends typically explored this way include religiosity and religious practice, subjective well-being, inequality and redistribution, and more. Sometimes multiple comparative surveys are used to observe trends in analytically related yet distinct concepts measured by separate datasets, such as various aspects of the secularisation trend or the effects of the economic crisis on various aspects of social capital.

Combining cross-section and trend (5 articles). Some publications use one comparative dataset for cross-section analysis and explanatory model building, usually the more recent one or the one with richer conceptual coverage and obtain a simpler cross-time picture from another comparative survey that goes further back in time. An example is an ESS-based model of the political participation of young Europeans, with EVS/WVS data used to illustrate trends in three non-electoral forms of participation.

Compensating for time-series gaps (3 articles). This strategy is mostly pragmatic, used by some analysts who examine larger sets of countries in a cross-time perspective and face a problem of inconsistent participation in their main dataset. In such cases they may resort to other comparative surveys to fill in the missing time points for one or more countries, using identical or similar indicator(s) at the closest available time points.

Pooling data

Merging surveys to use estimation techniques appropriate to a single pooled sample was the least frequent reason for combining data. Harmonisation issues, rather than user preferences, are the likely reason for these low numbers, as analysts are generally keen to increase sample sizes and measurement frequency. This is demonstrated in Table 5, which shows that within-survey pooling of samples, where no harmonisation barrier exists, is a markedly widespread analytical

Table 5. Multiple waves use in four comparative surveys (Dataset 2)

	WVS (N = 193)	ESS (N = 199)	EVS (N = 143)	ISSP (N = 262)
	(%)			
Articles using multiple rounds	39.9	39.1	28.0	25.6
To conduct cross-time analysis[a]	36.5	41.7	87.5	68.7
To pool samples	59.5	73.6	37.5	55.2
N =	193	199	143	262

Note: [a] Publications combining cross-time and pooled cross-sectional analysis were counted in both categories.

strategy, particularly in the cases of the ESS and the WVS with a denser fielding frequency, where combined sample sizes across all waves and countries come close to or surpass 400 000 cases. The other key reason to use multiple rounds is, of course, cross-time analysis, a prevailing motivation in EVS and ISSP publications.

Only 7 articles containing pooled cross-survey datasets were considered; these typically included a dependent concept, such as attending religious services, social mobility, or the effects of schooling reforms in a cross-time perspective in a narrow social group, such as single-country churchgoers, single-country immigrant group, birth cohorts, or occupational category. In these research designs, relevant sample sizes or, sometimes, adequate measurement density cannot be achieved by cumulating rounds across individual comparative surveys alone, so cross-survey pooling is the solution of choice.

Discussion and conclusion

Our study, which is descriptive in nature, examined what we term the 'ecosystem' approach to answering research questions in comparative research, an approach characterised by analytically motivated reliance on multiple data sources. Specifically, we investigated the combined usage of four comparative surveys, one of the key segments of individual-level data providers. We found that analysts benefit from inserting country aggregates from each other's repository of indicators, building complementary explanatory models, cross-validating theoretical models, comparing European and global estimates, and observing the dynamics of phenomena across a longer combined timespan. In the majority of cases, the use of multiple data sources was not a marginal exercise but a necessity in order to answer the research questions or to answer them in a more complete or valid way. Such practices are consistent with the broad notion of an ecosystem as adapted by the non-biological sciences. The four comparative programmes are characterised as interacting 'actors', whose rich series of longitudinal indicators are combined by academic authors to enhance or validate their analytical approaches. We could further extend the notion of interacting actors to the exchange of best practices in comparative methodology, such as fieldwork monitoring, management, or transparency [Lyberg et al. 2019; Jowell et al. 2007; Briceno-Rosas et al. 2020], but the present study was mainly focused on the analytical dimension, which, to our knowledge, has not yet been empirically explored. Not least because in the light of limited research funds the surveys can easily be perceived as competitors, while the aspect of epistemic complementarity and analytical cross-usage is seldom highlighted.

A potential weakness of our study is the lack of attention to the ways in which the 'ecosystem' of relevant social indicators increasingly includes big social data and other types of IT-based high-volume data, a development widely expected to open new epistemological possibilities for studying contemporary

societies [Olshannikova et al. 2017]. The reason this aspect is missing is that we did not detect any significant combined usage of (social) big data and data from comparative survey programmes, which is in itself a relevant but perhaps somewhat disappointing finding. Creating methodological and theoretical synergies between big data and traditional data can be considered one of the key future challenges for social and data scientists [White and Breckenridge 2014: 336–337].

Our study also brings into view some further issues and limitations. The first one arises, perhaps paradoxically, from the proliferation of multi-level comparative studies [Smith 2019; Andress et al. 2019], a widely used comparative approach that relies on a web of standardised cross-national indicators measuring individual and macro-level variables. Besides achieving greater explanatory power, the multi-level trend may also negatively impact existing disparities in knowledge production between countries and regions by relying on a more costly multi-data model of data collection and training. According to the ESS bibliographic report, approximately a third of authors offering reasons for excluding individual countries from analysis, despite their micro data being available in the cumulative ESS data file, cite the absence of standardised macro indicators for doing so [Malnar 2019: 25]. As noted in the literature, the OECD world is more completely documented with respect to social science statistics than other geographic areas, which biases insights towards the prosperous parts of the world [Goerres et al. 2019; Wysmułek 2018; Kroneberg 2019]. In this vein, missing macro indicators may lead to what could be termed 'secondary' exclusion, in addition to which less affluent countries are also less likely to field the surveys regularly in the first place, or less likely to field them in full accordance with the centrally prescribed protocols [Norris 2009; Lynn 2003], which are additional reasons for exclusion. With other analysts cautioning that future levels of funding for surveys that produce official statistics and social indicators are unlikely to grow and may well decline [Keeter 2012; Massey and Tourangeau 2013], the problem of imbalanced knowledge production could become even more persistent.

Returning to our main focus, combining micro indicators from comparative social surveys, the most obvious concern relates to validity risks. To what extent do considerable differences in questionnaire and sampling designs, translation procedures, fielding modes, and monitoring protocols, etc., make these strategies a viable analytical choice? As our study shows, analysts combine indicators across surveys in a variety of ways, with the scientific quality of these strategies largely depending on their expertise and analytical rigour. There is always a risk of errors originating from their own misjudgements, in terms of the selection of indicators and hard-to-replicate harmonisation decisions [Winters and Natcher 2016], or from larger problems with measurement equivalence in cross-national survey research. The latter highlights the importance of the transparency of survey methodology and protocols, which is essential for enabling analysts to make informed decisions [Lyberg 2019; Norris 2009; Harkness 1999]. In worst case scenarios, combining data can result in a decrease in the validity of findings and estimates.

On the other hand, our study shows that many analysts find the use of cross-survey data to be an important strategy for reducing uncertainties with respect to the validity of findings, allowing them to carry out robustness checks and sensitivity testing using independent, yet conceptually overlapping data sources. In a way, surveys represent each other's criterion-validity point [Drost 2011; Boukes and Morey 2018], at least relatively so, by providing opportunities to assess the agreement of their measures and results with external benchmark data. When discussing the issues surrounding cross-country data comparability in the context of the globalisation of public opinion research, Heath et al. [2005: 329] suggest that in particular 'data of lower than ideal quality should be checked against the results of other surveys that cover the same topics'. Such strategy may indeed help reduce uncertainties in many instances, but no comparative programme will ever be able to eliminate all the sources of survey and comparison error [Smith 2018], and if the cross-checking process ends up in conflicting outcomes, analysts must rely on informed methodological and theoretical judgements to decide how to proceed.

Finally, if the modest thesis that combining comparative data is 'potentially useful' [Schenker and Raghunathan 2007] is accepted, the question of how to facilitate such approaches arises. Among the many issues of cross-survey comparability, the factor that most directly hampers the efforts to combine data is the low standardisation of measures. Sharing a similar general mission, the four comparative surveys we examined have in common a number of widely used concepts and themes, but specific items, scales, and question wordings are rarely identical, with the exception of the two values surveys that originate from the same project and have shared between 100% and 30% of their content [Klingemann 2017: 139]. As noted by some scholars, this situation tends to confine the user to one particular programme, making it difficult or impossible to compare findings between projects [Heath et al. 2005; Tomescu-Dubrow and Slomczynski 2014]. In the light of our study, this situation forces analysts to take calculated risks by relying on similar but not harmonised measures, or to opt out of cross-survey data usage, and particularly from the pooling of samples, despite the obvious affinity for a larger number of cases that is indicated by the popularity of the intra-survey merging of waves in our study. Nonetheless, as the number of major general-purpose comparative surveys is relatively small, it is not inconceivable that their governing bodies would, at some point, consider pursuing the partial convergence of indicators, either of their own accord or at the prompting of other stakeholders, such as users or funders.

Alternatively, while ex-post harmonisation approaches are more complex and may not achieve the same level of comparability, there are several examples of national surveys on household income, health, mobility, and others being successfully transformed into integrated datasets with comparable measures, increased cross-national variation, and larger combined sample sizes [Dubrow and Tomescu-Dubrow 2016; Burkhauser and Lillard 2005]. There is also a grow-

ing number of digital platforms available for secondary data users, developed specifically for output harmonisation and replication purposes [Winters and Natcher 2016; Jeffers et al. 2017]. The availability of such resources suggests an increasing demand for cross-survey data use, driven by the need for improved timeliness, geographic or subpopulation detail, and statistical efficiency in social science data [NASEM 2017]. It may also signal the emergence of a new era in cross-survey standardisation and collaboration, with augmented possibilities for theory development and policy research.

BRINA MALNAR *is a research associate at the Faculty of Social Sciences, University of Ljubljana, and the head of the Public Opinion and Mass Communication Research Centre. As a member of the European Social Survey Core scientific team she leads the bibliographic monitoring task where her research activities are focuses on academic use of comparative survey data.*

LORNA RYAN *is a member of academic staff at the School of Arts and Social Sciences, City University of London. Her research interests include modes of science communication and EU research policy. She is a Research Manager with the European Social Survey (ERIC) at ESS ERIC HQ, City, University of London.*

References

Andress, H.-J., D. Fetchenhauer and H. Meulemann. 2019. 'Cross-National Comparative Research – Analytical Strategies, Results, and Explanations.' *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie* 71 (S1): 1–28, https://doi.org/10.1007/s11577-019-00594-x.

Blair, E. 2016. 'A Reflexive Exploration of Two Qualitative Data Coding Techniques.' *Journal of Methods and Measurement in the Social Sciences* 6 (1): 14–29, https://doi.org/10.2458/azu_jmmss.v6i1.18772.

Boukes, M. and A. C. Morey. 2018. 'Survey Context Effects and Implications for Validity: Measuring Political Discussion Frequency in Survey Research.' *Journal of Survey Statistics and Methodology* 7 (2): 201–226, https://doi.org/10.1093/jssam/smy008.

Briceno-Rosas, R., S. Butt and J. Kappelhof. 2020. 'Improving Central Monitoring of Fieldwork in Cross-National Surveys: The Case of the Fieldwork Management System in the European Social Survey.' *Survey Methods: Insights from the Field*. Retrieved 20 September 2021 (https://surveyinsights.org/?p=11866).

Bryan, M. L. and S. P. Jenkins. 2015. 'Multilevel Modelling of Country Effects: A Cautionary Tale.' *European Sociological Review* 32 (1): 3–22, https://doi.org/10.1093/esr/jcv059.

Burkhauser, R. V. and D. R. Lillard. 2005. 'The Contribution and Potential of Data Harmonization for Cross-National Comparative Research.' *DIW Diskussionspapiere* 486. Retrieved 20 July 2020 (https://www.econstor.eu/bitstream/10419/18337/1/dp486.pdf).

Curty, R. G. 2016. 'Factors Influencing Research Data Reuse in the Social Sciences: An Exploratory Study.' *International Journal of Digital Curation* 11 (1): 96–117, https://doi.org/10.2218/ijdc.v11i1.401.

Damian, E., B. Meuleman and W. van Oorschot. 2019. 'Transparency and Replication in Cross-National Survey Research: Identification of Problems and Possible Solutions.' *Sociological Methods & Research.* Retrieved 20 July 2020 (https://journals.sagepub.com/doi/abs/10.1177/0049124119882452), https://doi.org/10.1177/0049124119882452.

Davies, T. 2012. 'Open Data: Infrastructures and Ecosystems.' Paper presented at the ACM Web Science Conference, Evanston, Illinois. Retrieved 20 July 2020 (https://www.emergentworks.net/sites/default/files/ikmemergent_archive/Social_Life_of_Data_-_Infrastructure_and_Ecosystem_Paper.pdf).

Dong, Q., M. R. Elliott and T. E. Raghunathan. 2014. 'Combining Information from Multiple Complex Surveys.' *Survey Methodology* 40 (2): 347–354.

Doolan, D. M. and E. S. Froelicher. 2009. 'Using an Existing Data Set to Answer New Research Questions: A Methodological Review.' *Research and Theory for Nursing Practice* 23 (3): 203–215, https://doi.org/10.1891/1541-6577.23.3.203.

Drost, E. 2011. 'Validity and Reliability in Social Science Research.' *Education Research and Perspectives* 38 (1): 105–123.

Dubrow, J. K. and I. Tomescu-Dubrow. 2015. 'The Rise of Cross-National Survey Data Harmonization in the Social Sciences: Emergence of an Interdisciplinary Methodological Field.' *Quality & Quantity* 50 (4): 1449–1467, https://doi.org/10.1007/s11135-015-0215-z.

Duca, D. and K. Metzler. 2019. *The Ecosystem of Technologies for Social Science Research (White Paper).* London, UK: Sage, https://doi.org/10.4135/wp191101.

European Strategy Forum on Research Infrastructures (ESFRI). 2018. *ESFRI Roadmap 2018 – Strategy Report on Research Infrastructures*. Retrieved 20 July 2020 (http://roadmap2018.esfri.eu/).

Faniel, I. M., A. Kriesberg and E. Yakel. 2015. 'Social Scientists' Satisfaction with Data Reuse.' *Journal of the Association for Information Science and Technology* 67 (6): 1404–1416.

Franco, A., N. Malhotra and G. Simonovits. 2014. 'Publication Bias in the Social Sciences: Unlocking the File Drawer.' *Science* 345 (6203): 1502–1505, https://doi.org/10.1126/science.1255484.

Goerres, A., M. B. Siewert and C. Wagemann. 2019. 'Internationally Comparative Research Designs in the Social Sciences: Fundamental Issues, Case Selection Logics, and Research Limitations.' *KZfSS Kölner Zeitschrift für Soziologie Und Sozialpsychologie* 71: 75–97, https://doi.org/10.1007/s11577-019-00600-2.

Groves, R. M. 2011. 'Three Eras of Survey Research.' *Public Opinion Quarterly* 75 (5): 861–871, https://doi.org/10.1093/poq/nfr057.

Hadler, M., A. Eder, M. Haller and F. Höllinger. 2015. 'Methodological Problems of Quantitative Comparative Social Research.' *International Journal of Sociology* 45 (4): 247–255, https://doi.org/10.1080/00207659.2015.1098056.

Haerpfer, C. and K. A. Kizilova. 2016. 'The World's Largest Social Science Infrastructure and Academic Survey Research Program: The World Values Survey in the New Independent States.' *RUDN Journal of Sociology* 16 (4): 719–741.

Harden, A. 2010. 'Mixed Methods Systematic Reviews: Integrating. Quantitative and Qualitative Findings.' *FOCUS – Technical Brief No. 25*. Retrieved 8 July 2020 (https://ktdrr.org/ktlibrary/articles_pubs/ncddrwork/focus/focus25/Focus25.pdf).

Harkness, J. 1999. 'In Pursuit of Quality: Issues for Cross-National Survey Research.' *International Journal of Social Research Methodology* 2 (2): 125–140, https://doi.org/10.1080/136455799295096.

Harzing, A.-W. 2012. 'A Preliminary Test of Google Scholar as a Source for Citation

Data: A Longitudinal Study of Nobel Prize Winners.' *Scientometrics* 94 (3): 1057–1075, https://doi.org/10.1007/s11192-012-0777-7.

Heath, A., S. Fisher and S. Smith. 2005. 'The Globalization of Public Opinion Research.' *Annual Review of Political Science* 8 (1): 297–333, https://doi.org/10.1146/annurev.polisci.8.090203.103000.

Jacobides, M. G., C. Cennamo and A. Gawer. 2018. 'Towards a Theory of Ecosystems.' *Strategic Management Journal* 39 (8): 2255–2276, https://doi.org/10.1002/smj.2904.

Jeffers, K., M. King, L. Cleveland and P. Kelly Hall. 2017. 'Data Resource Profile: IPUMS-International.' *International Journal of Epidemiology* 46 (2): 390–391, https://doi.org/10.1093/ije/dyw321.

Johnston, M.P. 2017. 'Secondary Data Analysis: A Method of Which the Time Has Come.' *Qualitative and Quantitative Methods in Libraries* (QQML) 3: 619–626.

Jowell, R. 1998. 'How Comparative Is Comparative Research?' *American Behavioral Scientist* 42 (2): 168–177, https://doi.org/10.1177/0002764298042002004.

Jowell, R., C. Roberts, R. Fitzgerald and G. Eva. 2007. *Measuring Attitudes Cross-Nationally. Lessons from the European Social Survey*. London: Sage, https://doi.org/10.4135/9781849209458.

Keeter, S. 2012. 'Presidential Address: Survey Research, Its New Frontiers, and Democracy.' *Public Opinion Quarterly* 76 (3): 600–608, https://doi.org/10.1093/poq/nfs044.

Klingemann, H.-D. 2017. 'Prolegomena to a Comparative Second-order Analysis of the World Values Survey.' Pp. 133–144 in *Societal Enlightenment in Turbulent Times*, edited by B. Malnar and K. Müller. Vienna: Echoraum.

Kroneberg, C. 2019. 'Theory Development in Comparative Social Research.' *KZfSS Kölner Zeitschrift für Soziologie Und Sozialpsychologie* 71: 29–51, https://doi.org/10.1007/s11577-019-00604-y.

Kuechler, M. 1998. 'The Survey Method: An Indispensable Tool for Social Science Research Everywhere?' *American Behavioral Scientist* 42 (2): 178–200, https://doi.org/10.1177/0002764298042002005.

Lagos, M. 2008. 'International Comparative Surveys: Their Purpose, Content and Methodological Challenges.' Pp. 580–593 in *The SAGE Handbook of Public Opinion Research*, edited by W. Donsbach and M. Traugott. London: SAGE, https://doi.org/10.4135/9781848607910.n55.

Lindstrøm, M. D. and K. Kropp. 2017. 'Understanding the Infrastructure of European Research Infrastructures – The Case of the European Social Survey (ESS-ERIC).' *Science and Public Policy* 44 (6): 855–864, https://doi.org/10.1093/scipol/scx018.

Lyberg, L., L. Japec and C. Tongur. 2019. 'Prevailing Issues and the Future of Comparative Surveys.' Pp. 1053–1082 in *Advances in Comparative Survey Methods*, edited by T. Johnson, B. Pennell, I. Stoop and B. Dorer. Hoboken: Wiley, https://doi.org/10.1002/9781118884997.ch48.

Lynn, P. 2003. 'Developing Quality Standards for Cross-National Survey Research: Five Approaches.' *International Journal of Social Research Methodology* 6 (4): 323–336, https://doi.org/10.1080/13645570210132848.

Lynn, P., L. Japec and L. Lyberg. 2006. 'What's So Special about Cross-National Surveys?' *ZUMA Nachrichten Spezial* 12: 7–20.

Malnar, B. 2019. *ESS Annual Bibliographic Report 2019 (ESS ERIC Deliverable 11.8)*. Retrieved 21 May 2020 (http://www.europeansocialsurvey.org/docs/findings/ESS-ERIC-Annual-Bibliographic-Report-2019.pdf).

Massey, D. S. and R. Tourangeau. 2013. 'Introduction. New Challenges to Social Measurement.' *The ANNALS of the American Academy of Political and Social Science* 645 (1): 6–22, https://doi.org/10.1177/0002716212463314.

Mathison, S. 1988. 'Why Triangulate?' *Educational Researcher* 17 (2): 13–17, https://doi.org/10.3102/0013189X017002013.

Mayr, P. and A.-K. Walter. 2007. 'An Exploratory Study of Google Scholar.' *Online Information Review* 31 (6): 814–830, https://doi.org/10.1108/14684520710841784.

Mochmann, E. 2008. 'Improving the Evidence Base for International Comparative Research.' *International Social Science Journal* 59 (193): 489–506, https://doi.org/10.1111/j.1468-2451.2009.01681.x.

Müller-Bloch, C. and J. Kranz. 2015. 'A Framework for Rigorously Identifying Research Gaps in Qualitative Literature Reviews.' *Proceedings of the 36th International Conference on Information Systems (ICIS)*, Fort Worth, USA, 13–16 December 2015.

NASEM – National Academies of Sciences, Engineering, and Medicine. 2017. *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps*. Washington, DC: The National Academies Press.

Norris, P. 2009. 'The Globalization of Comparative Public Opinion Research.' Pp. 520–540 in *The Sage Handbook of Comparative Politics*, edited by T. Landmann and N. Robinson. Thousand Oak, CA: Sage.

Oh, D.-S., F. Phillips, S. Park and E. Lee. 2016. 'Innovation Ecosystems: A Critical Examination.' *Technovation* 54: 1–6, https://doi.org/10.1016/j.technovation.2016.02.004.

Oliver, S., A. Harden, R. Rees, J. Shepherd, G. Brunton, J. Garcia and A. Oakley. 2005. 'An Emerging Framework for Including Different Types of Evidence in Systematic Reviews for Public Policy.' *Evaluation* 11 (4): 428–446, https://doi.org/10.1177/1356389005059383.

Olshannikova, E., T. Olsson, J. Huhtamäki and H. Kärkkäinen. 2017. 'Conceptualizing Big Social Data.' *Journal of Big Data* 4 (1). Retrieved 20 July 2020 (https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0063-x#citeas), https://doi.org/10.1186/s40537-017-0063-x.

Pennell, B., K. C. Hibben, L. E. Lyberg, P. P. Mohler and G. Worku. 2017. 'A Total Survey Error Perspective on Surveys in Multinational, Multiregional, and Multicultural Contexts.' Pp. 179–201 in *Total Survey Error in Practice*, edited by P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, C. Tucker and B. West. Hoboken, NJ: Wiley, https://doi.org/10.1002/9781119041702.ch9.

Petticrew, M. and H. Roberts. 2006. *Systematic Reviews in the Social Sciences. A Practical Guide*. Malden, MA: Blackwell, https://doi.org/10.1002/9780470754887.

Quandt, M. and R. Luijkx. 2015. 'Data Bases and Statistical Systems: International Comparative Research.' Pp. 789–797 in *International Encyclopedia of the Social & Behavioral Sciences, Volume 5*, edited by J. D. Wright. 2nd ed. Oxford: Elsevier, https://doi.org/10.1016/B978-0-08-097086-8.41017-2.

Roberts, G. and D. Binder. 2009. 'Analyses Based on Combining Similar Information from Multiple Surveys.' Pp. 2138–2147 in *Proceedings of the Joint Statistical Meetings Section on Survey Methods Research*. Retrieved 10 November 2021 (https://www.semanticscholar.org/paper/Analyses-Based-on-Combining-Similar-Information-Roberts-Binder/70ef9ec615640c12f029fe0d026b0166ca88e2a7).

Schenker, N., J. F Gentleman, D. Rose, E. Hing and I. M. Shimizu. 2002. 'Combining Estimates from Complementary Surveys: A Case Study Using Prevalence Estimates from National Health Surveys of Households and Nursing Homes.' *Public Health Reports* 117 (4): 393–407, https://doi.org/10.1016/S0033-3549(04)50177-2.

Schenker, N. and T. E. Raghunathan. 2007. 'Combining Information from Multiple Surveys to Enhance Estimation of Measures of Health.' *Statistics in Medicine* 26 (8): 1802–1811, https://doi.org/10.1002/sim.2801.

Seppanen, M., S. Hyrynsalmi, K. Manikas and A. Suominen. 2017. 'Yet Another Ecosystem Literature Review: 10+1 Research Communities.' Pp. 1–8 in *2017 IEEE*

European Technology and Engineering Management Summit (E-TEMS). Munich: IEEE Institute of Electrical and Electronic Careers, https://doi.org/10.1109/E-TEMS.2017.8244229.

Smith, T. W. 2019. 'Improving Multinational, Multiregional, and Multicultural (3MC) Comparability Using the Total Survey Error (TSE) Paradigm.' Pp. 13–44 in *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts*, edited by T. P. Johnson, B. Pennell, I. Stoop and B. Dorer. Hoboken, NJ: Wiley, https://doi.org/10.1002/9781118884997.ch2.

Tengö, M., E. S. Brondizio, T. Elmqvist, P. Malmer and M. Spierenburg. 2014. 'Connecting Diverse Knowledge Systems for Enhanced Ecosystem Governance: The Multiple Evidence Base Approach.' *AMBIO* 43 (5): 579–591, https://doi.org/10.1007/s13280-014-0501-3.

Tomescu-Dubrow, I. and K. M. Slomczynski. 2014. 'Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective.' *Ask: Research and Methods* 23 (1): 103–114.

United Nations Economic Commission for Europe (UNECE). 2017. *Guide on Poverty Measurement.* New York and Geneva: United Nations.

Vezzoni, C. 2015. 'Secondary Analysis in the Social Sciences and Its Relation to Futures Studies.' *On the Horizon* 23 (2): 128–139, https://doi.org/10.1108/OTH-02-2015-0006.

Ware, M. and M. Mabe. 2012. *The STM Report. An Overview of Scientific and Scholarly Journal Publishing*. The Hague: International Association of Scientific, Technical and Medical Publishers.

White, P. and R. Saylor Breckenridge. 2014. 'Trade-Offs, Limitations, and Promises of Big Data in Social Science Research.' *Review of Policy Research* 31 (4): 331–338, https://doi.org/10.1111/ropr.12078.

Winters, K. and S. Netscher. 2016. 'Proposed Standards for Variable Harmonization Documentation and Referencing: A Case Study Using QuickCharmStats 1.1.' *PLOS ONE* 11(2): e0147795, https://doi.org/10.1371/journal.pone.0147795.

Wysmułek, I. 2018. 'Europe of Uneven Data: Country Representation in International Surveys on Corruption, 1989–2017.' *Ask: Research and Methods* 27 (1): 87–104, https://doi.org/10.18061/ask.v27i1.0005.