Church Attendance and Religious change Pooled European dataset (CARPE): a survey harmonization project for the comparative analysis of long-term trends in individual religiosity

Biolcati, Ferruccio; Molteni, Francesco; Quandt, Markus; Vezzoni, Cristiano

Check for updates

Church Attendance and Religious change Pooled European dataset (CARPE): a survey harmonization project for the comparative analysis of long-term trends in individual religiosity

Ferruccio Biolcati[1] · Francesco Molteni[1] · Markus Quandt[2] · Cristiano Vezzoni[1]

Abstract

Despite the long-lasting interest in religious change, debates on the topic have been heated and are still far from being settled. In order to provide a reliable data source through which to study these dynamics, the CARPE project harmonizes well-known international surveys containing items concerning religiosity (the ESS, Eurobarometer, EVS, ISSP and WVS). This makes it possible to broaden the available observation window, both across countries and over time. Moreover, the opportunities this provides for comparing different survey programmes also enable researchers to analyse the consistency of the results, minimizing the impact of random fluctuations and providing useful information with respect to the degree of confidence which can be placed on the relevant estimates. The main focus of this cumulative approach is the variable regarding church attendance, which has been harmonized in various ways. All in all, the CARPE dataset contains figures of religious practice for 45 countries spanning the period of 1970–2016 and derived from 1665 national surveys. This results in a sample of approximately 1.8 million individual observations. The aim of this contribution is to present the dataset's composition, the harmonization procedure adopted, the strategy used to combine the single datasets and the reliability tests which have been performed. Finally, some possible applications of the CARPE dataset will be introduced.

✉ Ferruccio Biolcati
  ferruccio.biolcati@unimi.it

Extended author information available on the last page of the article

🙎 Springer

1 Introduction

The origin of CARPE (Church Attendance and Religious change Pooled European dataset)[1] lies in a research project started few years ago and devoted to religious voting in Europe. In that context, those of us involved in the project were asked to provide a reliable description of the change of individual religiosity in Italy over time. The basic idea of harmonizing different studies (the EVS, WVS, ESS, etc.) came from a paper on church attendance trend in Italy (Pisati 2000). Based on this contribution, we decided to go further in the harmonization process in order to include as many datasets as possible. The first results of this work on the Italian case have been published in Vezzoni and Biolcati-Rinaldi (2015). Analysing the harmonized data, the authors depict the peculiar trend of religious change in Italy, with church attendance decreasing at different paces (including a period of stability in the 1980s) during the period between 1968 and 2010.

Since the beginning of research on this subject, it has been clear how fruitful such an approach incorporating different survey programmes can be. It offers the opportunity to tackle multiple research questions: describing trends of religious change, examining subgroup variation, disentangling period and cohort effects, and providing aggregate information for research projects in which religiosity plays the role of contextual variable. In the context of a harmonized dataset, researchers may increase the reliability of the estimates. They may deal with the issue of total survey error from a variety of methodological perspectives: wording, administration modes, sampling, or weighting. Finally, it opens the opportunity for collaboration with researchers willing to replicate the same strategy in different contexts, or with an interest in the theoretical and methodological issues mentioned above.

Given these considerations, we proceeded to extend the Italian dataset into a comparative (European) dimension, with the intended scope to have the widest possible coverage of European countries for the past 40 years. The result was the CARPE project, to which readers will be introduced in this paper. Section 2 of this article explains the purpose and research strategy adopted for the project. Section 3 provides technical details concerning the selection of studies and harmonization procedures used. A reliability analysis of the CARPE estimates is performed in Sect. 4, while the following section illustrates possible research applications, with respect to both the individual and aggregate versions of the dataset. Finally, some limitations of the overall project are discussed. "Appendix 1" presents the main characteristics of the aggregate version of the CARPE dataset which is available through the Data Archive for the Social Sciences of the GESIS—Leibniz Institute for the Social Sciences based in Cologne (Biolcati et al. 2020).

2 Purpose and research strategy

The CARPE project mainly aims to inform the debate about the substance of religious changes that occurred in Europe over the past decades. This being done by using data from a multiplicity of sources with sometimes more, sometimes less differing approaches to

---

[1] Throughout the article, we will use the terms 'pooled dataset' and 'harmonized dataset' as interchangeably, since they refer to the same research practice even if focusing on different aspects (pooling the datasets rather than harmonizing the variables).

measuring religiosity, a major challenge lies in granting the equivalence of these measures across the various data sources. The variability in extant operationalizations partly stems from different theoretical approaches to the concept of religiosity, and partly from purely methodological choices. The former source of variability—that of different meanings attributed to the data—falls in line with what the sociology of religion knows as Glock's (1959) paradox. Glock highlighted how the different interpretations of religious change in the United States might derive from researchers considering different dimensions and indicators of religiosity. A possible answer to Glock's paradox consists in a thorough revision of concepts and measures: the concept of religiosity could be redefined, dimensions reshaped, and new indicators be developed. But even with some existing indicators already being good measures for different aspects of what a theoretical revision might yield, this approach could only be fully executed when designing new data collections. Covering all desirable dimensions in a backward-looking way is impossible, as in the existing surveys, there are always gaps in regional or temporal coverage for a number of those dimensions.

The second, methodological, source of variability stems from different survey practices being followed across countries, fieldwork organisations, and points in time, be those related to ad hoc decisions or to conscious attempts at refining (and thus changing) measurement practice. This introduces variation even into what might otherwise count as measures of a single dimension of religiosity, e.g. when church attendance is measured with a six-point scale in one survey and with a seven-point scale in another one, and when response category wordings vary across surveys etc. This requires often difficult decisions about ex post-harmonization, on top of the harmonization problems that naturally arise from the cross-country comparative setting (Wolf et al. 2016).

Knowing that some of this methodological variation, e.g. in relation to sampling methods or non-response treatment, is often not revealed by the available documentation (Kołczyńska and Schoene 2019) makes this issue alone enough of a practical challenge for any claims to a valid harmonisation effort. We therefore decided to focus only on a single indicator of religiosity, on church attendance. Among all candidate indicators, church attendance luckily by and in itself combines the greatest coverage across survey programmes and over time with the smallest opportunities for conceptual variation. Thus we arrive at a large number of surveys of a narrowly defined measure of religiosity. Each survey represents a measurement occasion that can be pooled in a time series (Vezzoni and Biolcati-Rinaldi 2015). This enables us to analyse the consistency of the observed results with much better statistical power, by minimizing the impact of random fluctuations and by possibly detecting and controlling systematic methodological variation, so that the degree of confidence in the final estimates does increase (Firebaugh 2008, pp. 90–119; Kwak and Slomczynski 2019).

Also in terms of concept and measurement practice, focusing on the indicator of church attendance is a natural choice. To begin with, church attendance taps into religious practice, which is admittedly only one aspect of religious change but one that is particularly effective for measuring the ritualistic dimension of individual religiosity. Moreover, it has often been used to detect strong forms of religiosity ('belonging' as the counterpart to 'believing') in cross-national research (Billiet 2002; Brenner 2016; Davie 1994; Ruiter and Van Tubergen 2009; Voas 2009; Reitsma et al. 2012). Secondly, church attendance has a social dimension that is important in two respects: it exposes the laity to the messages of the clergy and fosters regular social interaction and the strengthening of group identity. Thirdly, the question concerning church attendance is relatively simple to formulate and easy to understand for respondents, given that it requires simply reporting the frequency of a certain behaviour.

3 Building the harmonized dataset

3.1 The selection of studies

We used three criteria for selecting the studies[2] to be included in the harmonized dataset. First, we decided to consider only studies produced for non-commercial purposes and with public access, since their data are easy to obtain and they usually provide decent documentation for data quality assessment. Secondly, we selected cross-national comparative studies with some thematic coverage of religion and full coverage of the adult population through nationally representative samples. Thirdly, we considered only repeated cross-sectional surveys, i.e. surveys pertaining to a study repeated over time on the same population but with different samples. This design was chosen in order to reconstruct trends in church attendance fully and comprehensively, since any new edition of the study provides an updated picture of the population and thereby improves our understanding of change at the aggregate level (Duncan and Kalton 1987). The issue of data quality (Blasius and Thiessen 2012; Groves 1989; Harkness 1999) has not been tackled directly; however, given the specific selection criteria we applied, the data quality of the selected studies, despite some heterogeneity across studies, rounds and countries, is well known in the scientific community and well assessed in the literature (Kołczyńska and Slomczynski 2019; Maineri et al. 2017; Oleksiyenko et al. 2019). Based on these criteria, five repeated cross-sectional studies were chosen:

- Eurobarometer (https://ec.europa.eu/public_opinion/),
- European Social Survey (ESS), (https://www.europeansocialsurvey.org/),
- European Values Study (EVS), (https://www.europeanvaluesstudy.eu/),
- International Social Survey Programme (ISSP) (https://www.issp.org/),
- World Values Survey (WVS) (https://www.worldvaluessurvey.org/).

The pooled dataset covers 45 European countries. Those countries are Albania, Austria, Armenia, Belgium, Bosnia Herzegovina, Bulgaria, Belarus, Croatia, Cyprus, Northern Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Kosovo, Latvia, Lithuania, Luxembourg, Macedonia, Malta, Moldova, Montenegro, the Netherlands, Norway, Poland, Portugal, Romania, the Russian Federation, Serbia, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, and the United Kingdom. Due to their historical specificities, a split version is also available for Germany (West and East) and the United Kingdom (Great Britain and Northern Ireland).

At this stage of the project, we decided to focus on European countries in order to limit the heterogeneity of the meaning of church attendance across religious traditions.

---

[2] The term 'study' is used as synonymous with the term 'survey programme'. The study develops in different countries and in different rounds (we prefer the term 'round' to the term 'wave' that is usually applied to panel studies). We refer also to the year when the fieldwork took place (in that country) to be distinguished from the planned round. So every survey belongs to a specific study, country, round and year.

3.2 The studies' specifications

All rounds of the different studies were included in CARPE, except when preliminary analyses detected quality problems or duplications. Specifications for the studies considered follow.

(a) *Eurobarometer*: the *Mannheim Eurobarometer Trend File 1970–2002* was combined with Eurobarometers 63.1, 63.4, 64.3, 65.1, 65.2, 65.3, 66.1 and 73.1, covering the period of 2005–2010, the *Central and Eastern Eurobarometer* (1990–1997) and the *Candidate Countries Eurobarometer* (2000–2004). Eurobarometer 46.1 was dropped from the harmonized dataset since it overestimates church attendance due to the provision of a specific filter question (self-described religiosity): church attendance was not asked to those not describing themselves as religious and considered as missing for these cases.
(b) *ESS*: Rounds 1 (2002) to 7 (2014) were included in the dataset.
(c) *EVS* and *WVS*: the integrated version, combining the EVS (four rounds: 1981, 1990, 1999 and 2008) and WVS (six rounds: 1981–1984, 1989–1993, 1994–1998, 1999–2004, 2005–2009 and 2010–2014) were included in the pooled dataset.
(d) ISSP: rounds from 1985 to 2016 were included in the dataset. For some study rounds, data referring to two consecutive ISSP 'modules' (equivalent to study rounds) have been collected within the same sample in several countries. In order not to duplicate the sample, the following 'country rounds' were not included in the harmonized dataset: Austria 1987, 2000, 2003, 2007 and 2009; Ireland 1990, 2004 and 2007; Italy 1986, 1988 and 1990; the Netherlands 2003; Portugal 2003, 2005 and 2008; and Switzerland 1996, 2004, 2006, 2008 and 2011.

There are two versions of the CARPE dataset: an individual level version containing 1,879,751 cases, and an aggregate version in which every case refers to a specific combination of study, year and country (Biolcati et al. 2020; see also "Appendix 1"). The composition of the harmonized dataset is summarized in Table 1. Each cell in the table reports the number of countries surveyed in the specific year-study combination. The harmonized dataset is made up of 1665 distinct surveys. The study contributing the most cases is Eurobarometer (48.0% of the surveys in the harmonized dataset comes from the European Commission study), followed by the ISSP (29.5%). The remaining surveys come from the ESS (10.8%), EVS (7.1%) and WVS (4.5%). The longest time series is Eurobarometer, which covers 37 years from 1973 to 2010. This is followed by the ISSP, which spans 31 years. The remaining studies cover different periods and have a different 'granularity'. The EVS has run every 9 years from 1981 to 2009, the observation window of the WVS stretches from 1989 to 2008 with approximately 5-year intervals between each round, and the ESS began in 2002 and rounds have been collected every 2 two years since.

Regarding the sample size, the ESS has the largest mean sample size (1453), followed by the EVS (1152), ISSP (1137), and WVS (1104). The surveys with the smallest mean sample sizes are those undertaken by Eurobarometer (805).

The distribution of surveys by study and country is reported in Table 2. The median number of surveys per country is 35 for the 44 years of the observation window. If countries with at least 20 surveys are selected, the median number rises to 47. The mean sample size is 1011, ranging from 454 for Luxembourg to 1640 for the Russian Federation.

Table 1 Surveys distribution by study and year, 1973–2016

Years	EB	ESS	ISSP	EVS	WVS	Total
1973	9					9
1975	10					10
1976	20					20
1977	20					20
1978	20					20
1980	10					10
1981	11			10		21
1982				2		2
1983				1		1
1984				1		1
1985	11		2			13
1986			5			5
1987			5			5
1988	13		4			17
1989	38		9		1	48
1990	42		7	19	4	72
1991	54		9	5	1	69
1992	48		8	1		57
1993	31		15	1		47
1994	44		18			62
1995	17		16		3	36
1996			13		13	26
1997			18		5	23
1998	17		21		6	44
1999			17	31		48
2000	13		16	2	1	32
2001	13		17		4	34
2002	26	15	23		2	66
2003	78	8	19			105
2004		18	20			38
2005	98	8	18		7	131
2006	123	21	16		9	169
2007		5	13		3	21
2008		18	24	39		81
2009		9	23	6	2	40
2010	34	16	20			70
2011		12	15		8	35
2012		18	18		3	39
2013		10	29		2	41
2014		15	20		1	36
2015		7	23			30
2016			11			11
Total	800	180	492	118	75	1,665
(%)	48.0	10.8	29.5	7.1	4.5	100.0
Mean sample size	805	1453	1137	1152	1104	1011

Table 2 Surveys distribution by study and country, 1973–2016

Countries	EB	ESS	ISSP	EVS	WVS	Total	Time coverage	Mean sample size
Albania	2	0	0	1	2	5	1991–2008	1075
Armenia	1	0	0	0	0	1	1992–1992	856
Austria	10	4	17	3	0	34	1986–2016	1112
Belarus	1	0	0	2	3	6	1990–2011	1269
Belgium	38	7	1	4	0	50	1973–2014	1068
Belgium-Brussels	0	0	3	0	0	3	2013–2015	503
Belgium-Flanders	0	0	10	0	0	10	2002–2015	1043
Belgium-Wallonia	0	0	3	0	0	3	2013–2015	557
Bosnia Herzegovina	0	0	0	1	1	2	2001–2008	1293
Bulgaria	19	4	9	3	2	37	1991–2013	987
Croatia	7	2	7	2	1	19	1996–2016	995
Cyprus	18	4	11	1	2	36	1996–2012	689
Czech Republic	20	6	23	3	2	54	1991–2015	1188
Denmark	38	7	15	4	0	64	1973–2014	1023
Estonia	20	6	3	2	2	33	1991–2015	994
Finland	10	7	15	3	2	37	1990–2015	1126
France	38	7	19	4	1	69	1973–2015	1194
Georgia	1	0	3	1	3	8	1992–2016	1300
Germany East	22	7	26	3	3	61	1990–2016	758
Germany West	38	7	31	4	3	83	1973–2016	1086
Great Britain	38	7	30	4	1	80	1973–2014	1010
Greece	29	4	0	2	0	35	1981–2011	1026
Hungary	21	7	25	3	2	58	1986–2015	986
Iceland	2	2	5	4	0	13	1984–2016	794
Ireland	38	7	13	4	0	62	1973–2014	1019
Italy	38	3	10	4	1	56	1973–2013	1000
Kosovo	0	0	0	1	0	1	2008–2008	1559
Latvia	20	0	14	3	1	38	1990–2016	949
Lithuania	20	3	7	2	1	33	1991–2015	960
Luxembourg	38	2	0	2	0	42	1973–2010	454
Macedonia	1	0	0	1	2	4	1992–2008	1076
Malta	18	0	0	4	0	22	1983–2010	467
Moldova	1	0	0	1	3	5	1992–2008	1027
Montenegro	0	0	0	1	2	3	1996–2008	895
Netherlands	38	7	18	4	2	69	1973–2014	1149
Northern Cyprus	5	0	0	1	0	6	2005–2008	464
Northern Ireland	36	7	11	4	0	58	1975–2014	344
Norway	10	7	25	3	2	47	1982–2015	1239
Poland	20	7	18	3	3	51	1987–2015	1156
Portugal	27	7	10	3	0	47	1988–2015	1007
Romania	19	0	0	3	3	25	1991–2012	965
Russian Federation	2	4	19	2	4	31	1990–2015	1640
Serbia	0	0	0	1	2	3	1996–2008	1221
Slovak Republic	20	5	13	3	2	43	1990–2016	1011

Table 2 (continued)

Countries	EB	ESS	ISSP	EVS	WVS	Total	Time coverage	Mean sample size
Slovenia	19	7	15	3	3	47	1992–2015	942
Spain	27	7	20	4	5	63	1981–2016	1231
Sweden	10	7	21	4	3	45	1982–2016	1078
Switzerland	2	7	14	1	3	27	1987–2015	1161
Turkey	17	0	6	0	0	23	2000–2015	1064
Ukraine	1	5	2	2	3	13	1992–2013	1564
Total	800	180	492	118	75	1665	1973–2016	1011

Individuals aged 18–70 were selected. For respondents older than 70, the indicator considered here becomes less valid, since the decrease in church attendance recorded from that age onwards does not necessarily reflect a change in religiosity but rather the increasing difficulties the elderly may have in physically getting to church.

3.3 Harmonization procedures

While the introductory sentence of the question regarding church attendance is quite constant over studies (e.g. "Apart from such special occasions as weddings, funerals, etc., how often do you attend religious services?"—ISSP 2016), the answer options are characterized by a certain degree of heterogeneity. In order to do harmonize the response formats, three different strategies were followed.

A first, easy way was to distinguish between respondents attending church 'at least weekly' versus respondents attending 'less than weekly'. The weekly threshold makes sense for Catholic countries, since participation in Sunday mass is intended to be a precept of Catholicism. This brings with it a loss of accuracy, but the threshold has a clear meaning for most respondents and can be identified in all the surveys considered. However, this choice seems inappropriate for non-Catholic countries, where weekly church attendance is not so widespread and the distinction between 'regular' churchgoers and others occurs at different thresholds.

Alternatively, a monthly threshold would work for most denominations, at the expense of losing precision for Catholic countries. However, this has a major drawback, since Eurobarometer response formats do not include a monthly answer category prior to 2002 (see Table 3). This strategy thus results in the loss of a large number of cases, since our pooled dataset relies extensively on the Eurobarometer.

A third strategy to follow does not require setting a fixed threshold for church attendance. We proceeded by imputing an *implied probability* of weekly church attendance, as proposed by Hout and Greeley (1998, p. 116), derived from the reported frequency of church attendance. Our indicator of implied probability is obtained by converting reported church attendance frequency into a ratio between the weeks per year in which a person attended church and the total number of weeks per year. For example, when respondents declared that they attended church once a month, it was presumed that they attended church in approximately 12 of the 52 weeks of a year. Their implied probability of attending church in any given week therefore was equal to 0.23 (i.e. = 12/52). Table 3 reports

Table 3 Coding scheme for the implied probability of weekly church attendance, by study and years

Wording of response options	Eurobarometer (mannheim trend file + candidate countries)	Eurobarometer (central and eastern countries)	Eurobarometer (2005–2010)	ESS	EVS/WVS	ISSP 1985	ISSP 1986–1987	ISSP 1988–2000	ISSP 2001–2016
Every day				0.99					
Several times a week + everyday									0.99
Several times a week	0.99	0.99							
More than once a week			0.99	0.99	0.99				
Once a week or more								0.99	0.99
Once a week	0.99	0.99	0.99	0.99	0.99	0.99	0.99		
Two–three times a month								0.58	0.58
Once to three times a month						0.50	0.50		
At least once a month				0.23					
Once a month					0.23			0.23	0.23
About once a month			0.23						
Several times a year		0.12				0.12	0.12	0.12	0.12
Few times a year	0.12								
About every 2 or 3 months			0.10						
Twice a year, once a year							0.03		
Only on special holy days			0.03	0.03	0.03				
Other specific holy days					0.03				
Once a year	0.02				0.02				0.02
About once a year		0.02	0.02						
Less frequently		0.01			0.01	0.03	0.01	0.03	0.01
Less often			0.01	0.01					
Rarely		0.01							
Never	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

the coding plan for all the surveys included in the harmonized dataset. This transformation allows to treat church attendance as a continuous variable and to compare attendance across different studies while retaining the accuracy of the original response formats.

Following Hout and Greeley (1998), a probability of 0.99 has been assigned to those who declared to attend religious services every week. This choice seems reasonable to account for possible random events that could prevent people to attend mass, even if they would like to. In practical terms, the results are not affected by this choice. Some further clarifications might be useful. For "Two–three times a month" we chose a midpoint (30 weeks a year) between two times a month (24 weeks a year) and three times a month (36 weeks a year). Similarly, for "Once to three times a month" we chose the midpoint of 2 times a month and we assumed that this equals a 0.50 probability of weekly church attendance. For "Several/Few times a year" we assumed 6 weeks a year (1 week every 2 months). For "About every 2 or 3 months" we chose a midpoint (5 weeks a year) between every 2 months (6 weeks a year) and every 3 months (4 weeks a year). For "Only on special holy days" we assumed 1–2 (1.5) weeks a year. Considering the overall scale, we assigned a probability of 0.01 to "Less often" and "Rarely". The implied probability assigned to the answer option "Less frequently" for ISSP changes over time to accommodate for some options that were not present in certain rounds.

Beyond church attendance, a few other target variables were harmonized. We selected four background variables that at the same time were relevant for the analysis of church attendance and feasible with a sustainable workload. The following target variables are included in the dataset: gender, age, level of education, religious denomination. To facilitate the harmonization process, we defined very broad response options for the denomination (Catholic, Protestant, Orthodox, other) and the level of education (lower secondary or less, upper secondary, tertiary).

4 Reliability checks

For CARPE to be used extensively, the quality of the measurements it provides needs first to be assessed.[3] While assessing their validity is ultimately a question for the substantive research this is meant to invite, assessing certain aspects of their reliability is something that should come before any substantive application. Reliability, when speaking of measures, implies both "consistency" and "repeatability". A very common definition of reliability is provided by Carmines and Zeller, who affirm that "Reliability concerns the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials" (1979, p. 11). Similarly, Nunnally (1978) stresses the fact that such measurement consistency should be obtained over a variety of conditions, and Corbetta (2003) emphasizes the reproducibility of measurement results.

With regard to CARPE, two main issues concerning reliability arise: the first is that of the data coming from different sources, the second is that of the losses or biases that the harmonization process might introduce into the original data. For reliability assessments, the different samples from which the aggregate CARPE data set is built can be considered as different trials (Carmines and Zeller 1979), or at least as being a similar procedure performed under different conditions (Nunnally 1978). Measures of the same property

---

[3] Stata syntax is provided in "Appendix 2".
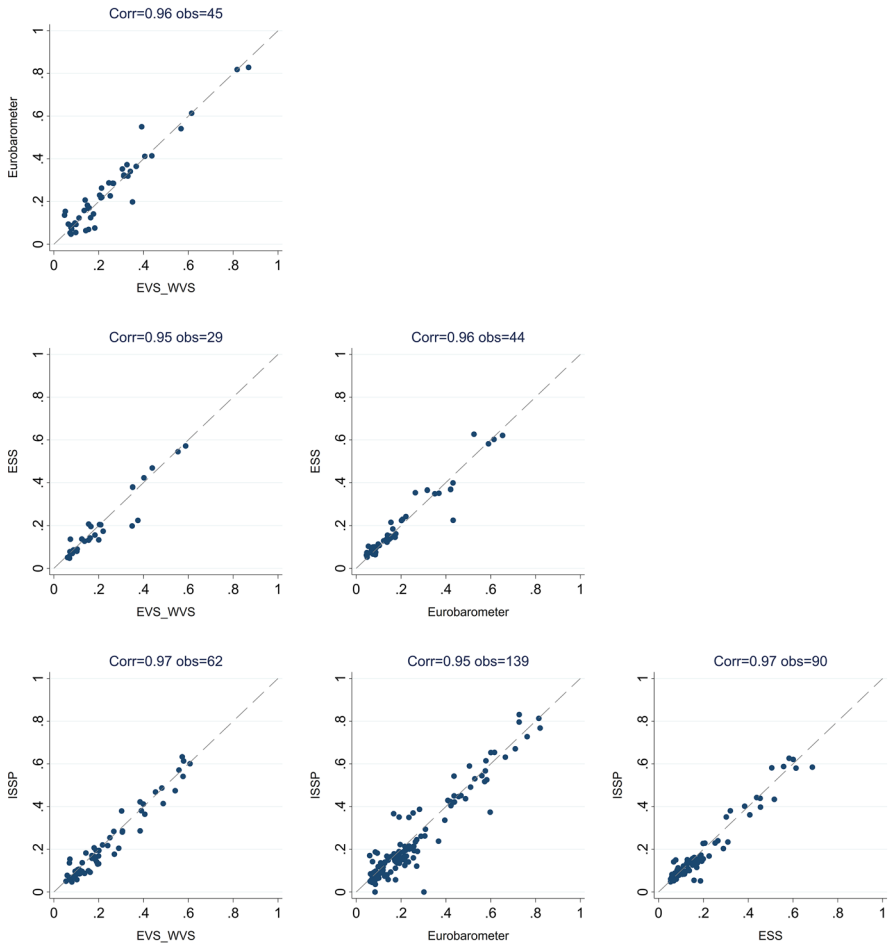
Fig. 1 Implied probabilities of weekly church attendance, country-year averages coming from different studies and the correlations between them

(frequency of church attendance in a certain year in a certain country) coming from different sources should ideally be consistent. To put this another way, the different sources should not produce systematically biased measurement outcomes.

First, our focus turns to the consistency between aggregate estimates coming from different survey programmes. The scatter plots in Fig. 1 show the country-year observations shared by different study pairs. For example, the first graph reports the mean implied probabilities of church attendance for the years and countries that overlap in Eurobarometer and EVS/WVS surveys. With the correlations at or above 0.95 for all pairs, this clearly demonstrates that observations coming from the different sources involved are fairly equivalent.

Moving on to a multivariate and multi-level perspective that also allows the precision of the estimates within each sample to be accounted for, it is possible to reach similar

Table 4 Testing the effects of studies on implied probability of weekly church attendance (4 level variance partition multilevel model and 3 levels fixed effects and random slopes multilevel models)

	Variance partition, 4 levels	Fixed effects, 3 levels	Random slopes, 3 levels
Fixed part			
Constant	0.240***	0.234***	0.246***
ESS *(ref: EB)*	–	0.010***	−0.015***
EVS/WVS *(ref:EB)*	–	−0.001	−0.014*
ISSP *(ref: EB)*	–	0.017***	−0.003
Random part			
Country			
variance(cons)	0.030***	0.030***	0.029***
Country-year			
variance(cons)	0.002***	0.004***	0.005***
variance(ESS)	–	–	0.002***
variance(EVS/WVS)	–	–	0.004***
variance(ISSP)	–	–	0.004***
Study-country-year			
variance(cons)	0.001***	–	–
Individual			
variance(cons)	0.108***	0.108***	0.108***
variance (total)	0.141	0.142	0.161
VPC (individual)	76.596	76.056	67.081
VPC (study-country-year)	0.709	–	–
VPC (country-year)	1.418	2.817	14.907
VPC (country)	21.277	21.127	18.012
Statistics			
N	1,885,921	1,885,921	1,885,921
Ll	−578,394	−579,870	−578,320

$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

conclusions. Table 4[4] presents three alternative multi-level estimation approaches. The first column (VARIANCE PARTITION, 4 LEVELS) basically assesses how individual variance in service attendance is accounted for by four levels of nesting implied within the harmonized dataset: the individual level, the study-country-year level, the country-year level and the country level. What is of interest here is the share of variance accounted for by the second level, that of study-country-year. This is because that level refers to the differences between samples (and thus, between survey programmes) within the different country-years (which are modelled as higher levels). This corresponds to the variance accounted for

---

[4] The analyses presented in Table 4 are performed after dropping countries with fewer than 5000 overall observations (Armenia, Bosnia and Herzegovina, Northern Cyprus, Montenegro, Serbia and Kosovo) as well as Turkey as the only Muslim-majority country in the dataset. With the exception of Northern Cyprus and Turkey, all these countries also report observations for just one year and one survey programme.

All information about the survey programmes is introduced at the highest level of aggregation possible. Thus, the Eurobarometer category includes the Mannheim trend file, Candidate countries, Central and Eastern Eurobarometers and individual Eurobarometer samples, and the EVS and WVS were considered together due to the similarity of their question formats.

the different survey programmes after having controlled for the variance due to the various countries and years in which the interviews took place. As the VPC[5] shows, the share of variance due to survey differences within country-years is only 0.7%. This confirms the negligible impact that the survey sources have on the overall measurement.

The other two models follow a different strategy to obtain almost the same results. In the FIXED EFFECTS and RANDOM SLOPES 3 LEVELS models, the differences between surveys are not modelled as additional levels like before, but rather as covariates in the fixed part of the model. The only difference between the two is that in the latter model the effects of such covariates are allowed to vary between country-years, as this controls for the different coverage (in terms of years and countries) of the different studies. This, again, makes it possible to model whether different studies lead to different measurement results given that it excludes the effect of the different studies designs in term of period covered and granularity. Generally speaking, the random slope model shows a better fit if compared to the fixed effect one (lr test = 0.000). This, together with the fact that the fixed effects are very small (less than 0.02 on a 0–1 scale) suggests that on average the different survey programmes in the same country-year capture the same level of church attendance.[6]

Overall, the tests presented here lead to the same conclusion: the different survey programmes on which CARPE is based (the EVS/WVS, EB, ISSP and ESS) are convincingly consistent as a foundation on which to build a harmonized form of measurement of church attendance: neither the differences in the sources nor the effects of the implied probability-transformation seem to present any biases or distortions.

5 Possible research applications

Generally speaking, the CARPE dataset can be useful in many ways. Among them, we would stress the following ones:

- Providing long-term country trends in religious change for individual countries.
- Comparing and grouping church attendance and religious change trends.
- Building country and country-year level figures to be used as contextual measures in multilevel models.
- Disentangling the effects of cohort replacement and period succession in explaining religious change.

While the first three tasks can be accomplished by relying on the aggregate version of the CARPE dataset, the last one needs to be based on individual-level observations. Here we present one application for each observation level of the CARPE dataset. In the first, we have applied group-based trajectory models (GBTM) to the aggregate dataset in order to identify the best way for clustering European countries according to their trends in church attendance. In the second, we have applied a cross-classified multi-level model

---

[5] The Variation Partition Coefficient (VPC) measures the percentage of variance accounted for by a specific level.

[6] Coefficients for ESS and EVS/WVS are statistically significant also in the random slope model, but given the sample size in these analyses (N = 1,885,921), some formally significant differences are to be expected even at unsubstantial effect sizes.

to the individual-level dataset in order to disentangle the relative contributions of cohort replacement and period succession in describing religious change.

These two examples are primarily focused on the variable concerning church attendance, as it represents the main target of the harmonization procedure. However, this does not exclude the possibility to differentiate such analyses by gender, different denominations, age groups or levels of education. As described, CARPE also contains a harmonized version of those variables.

5.1 A group-based trajectory model for clustering European countries

The main aim of this application is to cluster European countries' religious trends in the most meaningful and economical way possible. Other attempts have been made in the relevant literature to perform the same exercise, but all were based on data coming from single studies, or even data only from specific rounds of such studies. An ex-ante clustering mainly based on territorial distributions has been used by Voas and Doebler (2011) to present some general trends in religious change across cohorts in Europe. Burkimsher (2014) has also drawn some country trends by inspecting inter-cohort differentials for the 1950s–1980s. Finally, Brenner (2016) has provided a concise summary of religious service attendance trends.

Three main methods have been proposed in the literature for modelling trends and general developmental trajectories (Jen et al. 2010). The first is hierarchical modelling (Bryk and Raudenbush 1987), the second is growth curve modelling (Meredith and Tisak 1990), and finally there is group-based trajectory modelling (Nagin 2005). The first two approaches assume that the higher-level parameters are distributed according to a normal distribution, whereas group-based trajectory modelling does not (Jen et al. 2010). Given this relevant feature, this third approach is more flexible when it comes to identifying the clustering solution that best fits our data.

When applying this strategy to the aggregate version of CARPE, the best-performing solution is a 7-group one in which almost all cluster trends are modelled as a 2nd order polynomial except for cluster 3 (1st order), and cluster 5 (3rd order). Table 5 shows the results of this model estimation.

Table 5 Group-based trajectory model estimation results (description of the 7 groups and corresponding parameters form a third level polynomial interpolation)

	Intercept	Year	Year^2	Year^3
Group 1: Low and increasing	0.12	− 0.0031	0.00014	
Group 2: Low and stable	0.08	− 0.0003	0.00001	
Group 3: Low and slowly declining	0.19	− 0.0017		
Group 4: Average and rapidly declining	0.33	− 0.0072	0.00006	
Group 5: High and rapidly declining	0.46	− 0.0501	0.00042	− .00001
Group 6: High and slowly declining	0.41	0.0001	− 0.00006	
Group 7: Very high and rapidly declining	0.92	− 0.0107	0.00001	

A brief look at the coefficients makes it possible to outline the main features of the seven groups.[7] The first group consists of just one country (Ukraine) showing an increasing U-shaped trend. This is followed by a very large group of countries (comprising almost 35% of the observations) whose trend is essentially flat (very small linear and quadratic components) and with very low starting points (around a probability of 0.08). Then come three different groups in which decreasing trends can be observed, starting from different levels; the higher the starting point the steeper the trend. In addition, Group 5 also displays a relevant quadratic component. Group 6 is special: it starts from a quite high level of 0.41 but both its linear and quadratic components are close to 0. An inspection of the trends in the countries belonging to this group shows a great heterogeneity in their functional forms: overall, the shape can be described as (very) slowly declining. The remaining group (7) is the second-smallest group: its trend starts very high (around the probability of 0.92) and declines almost linearly. The features of the seven groups' trends can be visualized better by looking at Fig. 2.

Figure 2 clearly depicts the wide observation window spanning from 1973 to 2016. This has been possible only because of the cumulative approach used in building the CARPE dataset.

Starting with this clustering solution, it then becomes possible to provide a unique example of how a harmonized dataset can accentuate the strengths of different studies while at the same time minimizing their weaknesses. Figure 3 isolates two of the studies behind CARPE in order to highlight the potentials of such data harmonization. The EVS has been chosen because of its wide thematic coverage of religion, as well as its large and regular observation window (every 9 years starting with 1981). To contrast with this, the ESS has been chosen because of its much shorter time coverage, but also for its high granularity (every 2 years starting with 2002). As will be shown, CARPE combines the strengths of both studies. Panel 1 shows the trends of the 7 clusters that are drawn by relying only on EVS data, while Panel 2 does the same for ESS data. Panel 3 plots these two figures onto the same graph. From these three panels, the characteristics of the two studies emerge clearly. Panel 4 shows the same 7 trends based on combined EVS and ESS data, while Panel 5 compares these latter trends with those based on the entire CARPE dataset. Panel 6 shows the trends based only on CARPE. Especially when looking at Panel 5, it clearly emerges how a harmonized dataset such as CARPE is able both to increase the granularity and extend the observation window of a cumulative approach investigation in relation to the windows of the individual studies involved, while simultaneously providing very reliable estimations.

---

[7] *Low and increasing* Ukraine.

Low and stable Bulgaria, Belarus, Czech Rep., Denmark, Estonia, Finland, E. Germany, Iceland, Latvia, Norway, Russian Fed., Sweden.

Low and slowly declining France, Great Britain, Hungary.

Average and rapidly declining Belgium, W. Germany, Lithuania, Netherlands, Switzerland.
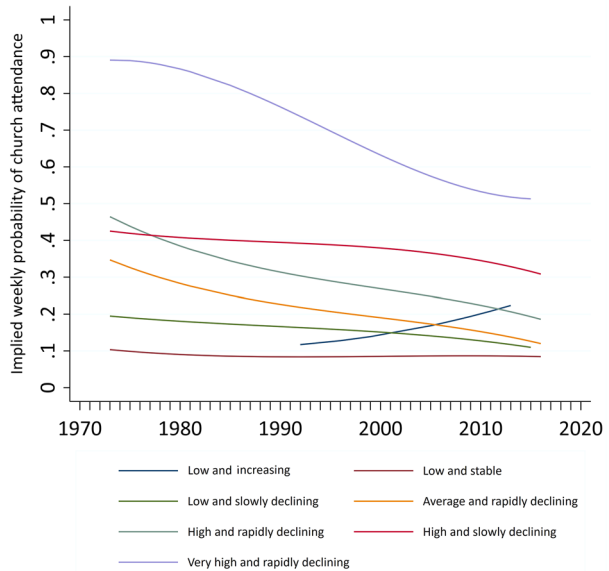
High and rapidly declining Austria, Greece, Luxembourg, Romania, Slovenia, Spain.

High and slowly declining Croatia, Cyprus, Italy, Portugal, Slovakia.

Very high and rapidly declining Ireland, Poland.

Fig. 2 7 Groups clustering solution for the implied probability of church attendance in Europe between 1973 and 2016

5.2 A cross-classified model to disentangle period and cohort effects on religious change

A second empirical application aims to disentangle the effects of cohort replacement and period succession on religious change. Far from being a simple empirical exercise, this distinction turns out to have a very strong theoretical value. If religious decline is mainly a matter of period succession (the effect of the passage of time), then the focus of both theoretical argumentation and empirical analysis should be on the phenomena or events that make people move away from religion. If instead cohort replacement is the leading mechanism, then the focus should be on the ways in which religious values, attitudes and behaviours are (no longer) transmitted from older to younger generations.

Separating these two mechanisms of change is complicated by the Age–Period–Cohort problem, which arises from the fact that possible effects of periods and cohorts are perfectly confounded by the effect of people aging. The methodological debate around this is often dominated by very technical solutions trying to solve the main problem directly: the problem of identification arising from the perfect collinearity of age, periods and cohorts (A = P–C).

Generally, the techniques attempting to deal with the Age–Period–Cohort problems can be divided into two main groups. On the one side are approaches trying to find a general procedure that can be applied to any APC problem (Yang and Land 2008). In this group we can place the constrained generalized linear models (Yang and Land 2013), the intrinsic estimator (Yang et al. 2004), and the Hierarchical Age-Period-Cohort models (HAPC) (Yang and Land 2008). On the other side are alternative approaches arguing for a more subject-specific perspective, which would exclude one of the three factors of the equation on theoretical grounds (Bell and Jones 2017).

Our contribution is characterized by three main features. (1) It draws on the Hierarchical Age–Period–Cohort models (HAPC) in which age is treated as a fixed-effect
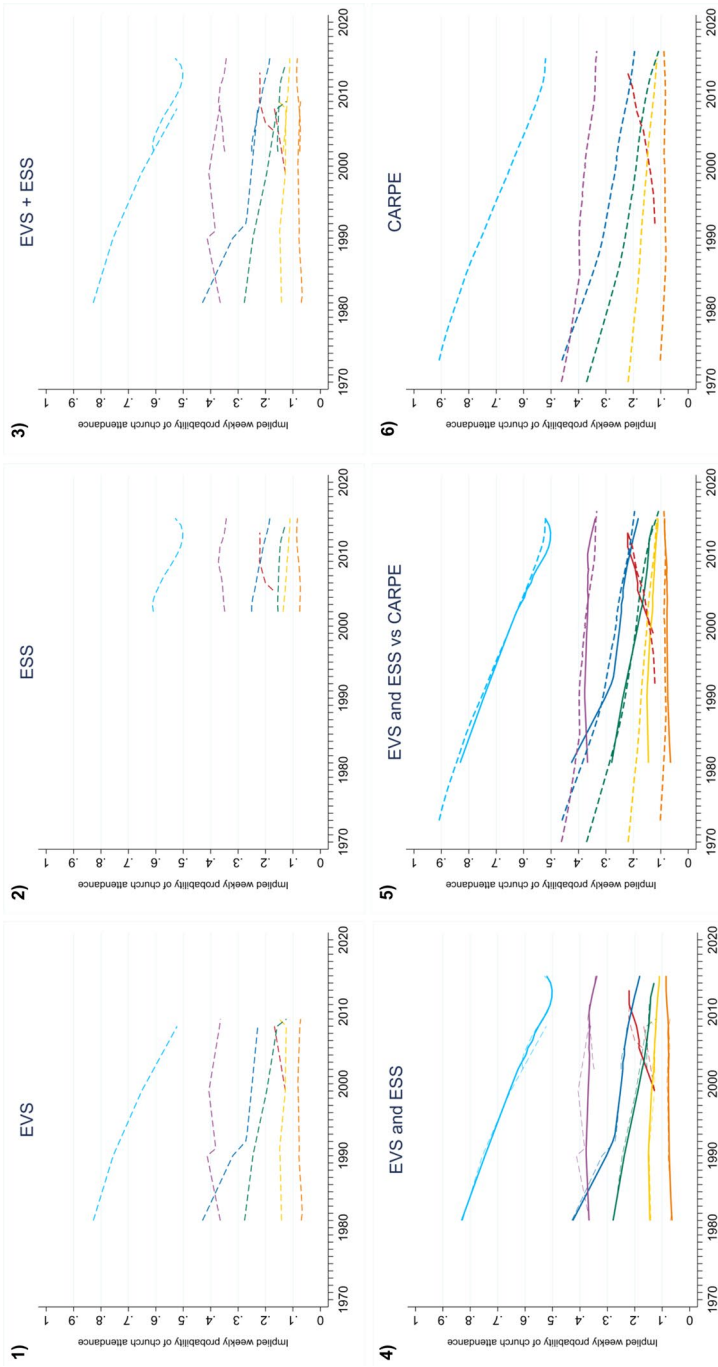
Fig. 3 Combining different studies to draw aggregate trends: an example from EVS and ESS

at the individual level and periods and cohorts as two cross-classified higher levels. However, (2) it considers age as irrelevant for explaining long-term aggregate religious change, thus introducing it only as a categorical control variable calibrated to measure three key life-stages for religious behaviour (< 30, 30–59 and > 59). Finally, (3) it takes full advantage of the harmonized CARPE dataset in order to minimize the main estimation bias potentially linked with such a way of modelling. Concerning this last point, the main critique levelled against HAPC models is that they are very sensitive to the ranges of periods and cohorts. As Bell and Jones note, "when a sample based on a small range of cohorts is taken, such that the period range is much greater than the cohort range, the results produced are very different to those produced when cohort groups span a much wider range than periods, as is structurally the case with repeated cross-sectional data" (2017, p. 1). This happens because the model tends to assign higher parameter variance where there is higher variability in the data. Clearly, a cumulative design such as that of CARPE makes it possible to rely on a much more balanced structure (71 birth cohorts and 42 periods) than single cross-sectional designs such as the EVS (just 4 rounds) or ESS (just 9 rounds).

Overall, the use of the individual-level version of the CARPE dataset makes it possible not only to control for this bias, but also to extend the time frame (back to the 1920 cohort and the year of 1973 respectively), while at the same time estimating parameters for every single year, thus providing very reliable and precise figures.

The entire analytical procedure for this analysis is based on three different modelling steps. The first one (M1) is based on periods, cohorts and countries as cross-classified higher levels. This means that every individual in the sample belongs to a unique combination of period, cohort and country. The second step (M2) adds the interaction between periods, cohorts, and countries as additional cross-classified levels, which allows country-specific trends to diverge from the general one. The third step (M3) simply adds age as a three-category fixed effect term.

As before for the reliability assessment, we turn to the VPC statistics to evaluate the relative impact of periods and cohorts in explaining religious change. These estimates are reported in Table 6 and appear to be coherent between the three models: the effect of cohort replacement in explaining religious change is overwhelmingly larger than that of period succession. The main empirical evidence of this is the VPC for cohorts always being higher (7.52 in the last model) than that of periods (0.54). Generally speaking, the proportion of variance that can be explained by cohort replacement is 14 times the size of that which can be explained by period succession.

6 Discussion and conclusions

The limited scale of the CARPE project highlights the burden as well as the potential benefits of its harmonization strategy. When human and financial resources are scarce, this kind of project becomes difficult to start, extend or update. A lot of work is involved in data preparation before analysis can even start: this implies high costs for researchers while returns are not certain. In cases where harmonization is successful, the theoretical and methodological benefits are remarkable. Researchers can pose new questions or assess old questions on a more reliable empirical basis. In addition, the number of methodological issues for which a harmonized dataset can provide potential solutions is remarkable.

Table 6 Cross-classified models to disentangle period and cohort effect. VPC results for the three different specifications

	M1 (Year-cohort-country)	M2 (M1 + Interactions)	M3 (M2 + Age)
Constant	0.24	0.25	0.23
Age 30–59 (ref: < 30)	–	–	− 0.00
Age 60–100 (ref: < 30)	–	–	0.02
VPC individual	*72.93*	*65.11*	*65.03*
VCP cohort	4.51	9.15	7.52
VPC year	0.34	0.67	0.54
VPC country	22.21	22.42	24.24
VPC cohort * country	–	1.80	1.82
VCP year * country	–	0.86	0.85
VPC TOTAL	100.00	100.00	100.00
N	1,718,770	1,718,770	1,718,770

With reference to the CARPE project, some limitations need to be highlighted. First, church attendance is a single indicator referring to a single dimension of a concept that is itself just as complex as individual religiosity. This obviously also restricts the conceptual validity of the harmonized measure. Adding further dimensions of religiosity to the dataset surely represents the most interesting possible expansion of the CARPE project. For example, adding subjective religiosity and daily praying could allow to construct a harmonized latent variable that could also be equivalent over time in a theoretically richer way than the single indicator of service attendance can (Billiet 2002). That the meaning attached to not attending church weekly by the Catholic church changed from a "dead sin" before the Vatican II Council to an undesired option thereafter points to changes in the nature of religiosity that a simple quantifying behavioural measure as ours per construction cannot possibly tap into.

Secondly, when investigating church attendance there is always the well-known problem of overreporting bias (for the Italian case, see Castegnaro and Dalla Zuanna 2006). Beyond the traditional explanation in terms of social desirability (Hadaway et al. 1993), recent contributions focus more on an identity-claiming mechanism (Brenner 2011; Vezzoni and Biolcati 2019). In the context of a longitudinal dataset, the potentially changing magnitude of such overreporting should be considered (Vezzoni and Biolcati-Rinaldi 2015, p. 105).

Thirdly, while there is a shortage of other indicators of religiosity, there is also a shortage of indicators of other phenomena related to religiosity. From this perspective, the scientific community should exploit the recurrence of the source datasets over the different harmonization projects: a sort of harmonized ID would give the opportunity to harmonize different datasets according to different research questions.

Fourthly, the issue of weighting is not considered. While all the studies considered here provide weights (but not all of them for all the years and all the countries), the CARPE harmonization project does not include them. Cross-national surveys are characterized by a significant degree of heterogeneity in weighting methods across studies, countries and over time. "Yet the variability in how these weights were calculated, their quality, and their adjustment effects are considerable" (Zieliński et al. 2019, 1035). Since weights are calculated in the different studies according to different strategies, we decided to discard them altogether. A different strategy could be to calculate new harmonized weights using

international official statistics. More research is needed on this point. Similarly, it would be useful to assess possible, if any, distortions due to the different sampling, answer options, and placement of questions within the questionnaire.

Lastly, more research is needed to further validate the use of the 'implied probabilities' transformation to the underlying discrete church attendance variable. This is not an issue of the harmonization process per-se, but becomes a relevant point to assess while reasoning about modelling requirements. This kind of reasoning, we believe, must be part of every harmonization design.

To conclude, the application of a harmonization strategy to a target variable reporting a behavior (i.e. church attendance) was illustrated. We think that such a strategy would be applicable to other survey items that measure the frequency of a particular behavior. Nevertheless, the item we dealt with presents two peculiarities that we believe important for the success of the harmonization strategy: firstly, church attendance is a behaviour very well-defined in normative terms; secondly, the wording of the question is relatively stable. As this is not the case for all the items about reported behavior, the success of the harmonization strategy might be more uncertain for some of them. This is something that should be discussed when considering a new harmonization project.

Appendix 1

Church attendance and religious change pooled European dataset (CARPE): aggregate version

The aggregate version of the CARPE dataset is available through the Data Archive for the Social Sciences of the GESIS—Leibniz Institute for the Social Sciences based in Cologne (Biolcati et al. 2020).

The main characteristics of this version are introduced in Sect. 3.

Both SPSS and STATA versions of the dataset are available.

The codebook is reproduced below.

study—study name. (1: EB Mannheim, 2: ESS, 4: ISSP, 5: EVS, 6: WVS, 8: Central and Eastern Eurobarometer, 9: Candidate Countries Eurobarometer).

study2—study name (synthetic version). (1: Eurobarometer, 2: ESS, 3: EVS/WVS, 4: ISSP).

round—survey round (i.e. 731 for Eurobarometer 73.1).

break—Unique identifier for survey-round (break = (STUDY*10,000) + round).

yearsurvey—year in which the fieldwork took place.

For surveys in which fieldwork spanned more than one calendar year, the year with more observations was coded.

ncases—number of cases on which the country-survey-year observations are based.

country—country of survey with splits (Germany (west and east), United Kingdom (Great Britain and Northern Ireland) and Belgium (Flanders, Wallonia and Brussels)). Codes are the ISO 3166–1 numeric codes.

country2—country of survey without splits. Codes are ISO 3166–1-Codes.

cath_p—percentage of people who declare to belong to a Catholic denomination.

prot_p—percentage of people who declare to belong to a Protestant denomination.

orth_p—percentage of people who declare to belong to an Orthodox denomination.

othden_p—percentage of people who declare "Other religion" as their denomination.

noden_p—percentage of people who declare not to belong to any denomination.

attendw_p—percentage of people who declare weekly church attendance.

attendm_p—percentage of people who declare monthly church attendance.

attendp—average measure of implied probability to attend church weekly.

The implied probability to attend church weekly measures the probability to attend church in any given week. For details, see Hout, M., Greeley, A.: What Church Officials' Reports Don't Show: Another Look at Church Attendance Data. American Sociological Review. 63(1), 113–119 (1998).

female—proportion of females over total.

ager—average age of respondents.

low_p—percentage of people with lower secondary education or less.

upp_p—percentage of people with upper secondary education.

tert_p—percentage of people with tertiary education.

Appendix 2

Stata syntax for Section 4: reliability checks

```
/********************************************************************************
*       Do-File for the replication of                                         *
*    Church Attendance and Religious change Pooled European dataset (CARPE).   *
*    A survey harmonization project for the comparative analysis               *
*    of long-term trends in individual religiosity                             *
*    Section 4: reliability checks                                             *
*                                                                              *
*       Authors: Ferruccio Biolcati, Francesco Molteni,                        *
*                Markus Quandt, Cristiano Vezzoni                              *
********************************************************************************/


******************************************************************
**Graphs*********************************************************
******************************************************************


use "...\CARPE_V4_relevant_variables.dta", clear      //Individual   version   CARPE
dataset
global working_folder ...\Materials

replace country2=899 if country2==276      //Minor    fix   for   Germany   already
implemented in the last version of the Dataset


//AGGREGATE VALUES

collapse (mean) attendp, by(yearsurvey country study2)
decode study2, gen(study2str)
replace study2str="EVS_WVS" if study2str=="EVS/WVS"

//CREATING DATASET

#delimit ;
levelsof study2str, local(study2);
foreach x of local study2 {;
preserve;
keep if study2str=="`x'";
rename attendp `x';
save "$working_folder/`x'.dta", replace;
restore;
                        };

#delimit cr

use "$working_folder/ESS.dta", clear
merge 1:1 yearsurvey country using "$working_folder/EB.dta", nogen
merge 1:1 yearsurvey country using "$working_folder/EVS_WVS.dta", nogen
merge 1:1 yearsurvey country using "$working_folder/ISSP.dta", nogen

save "$working_folder/Combined_observed.dta", replace

erase "$working_folder/EB.dta"
erase "$working_folder/EVS_WVS.dta"
erase "$working_folder/ISSP.dta"
erase "$working_folder/ESS.dta"
```

```
local allfiles : dir "$working_folder/" files "*.gph"
display `allfiles'
cd "$working_folder/"
graph combine  EVS_WVS_EB_.gph  EVS_WVS_ESS_.gph ESS_EB.gph  ///
                    EVS_WVS_ISSP_.gph EB_ISSP_.gph ESS_ISSP_.gph ,  ///
                                    holes (2 3 6)iscale(0.4) imargin(small)
rows(3) cols(3) ///
                                    graphregion(fcolor(white)     lstyle(none)
ilwidth(none) lcolor(gs16) margin(zero)) plotregion(fcolor(white) ilwidth(none) ///
                                    lstyle(none)  lwidth(none) margin(zero))
graph export TOTAL_observed.png, width(7000) height(7500) replace



********************************************************************
**Multilevel models**********************************************
********************************************************************

global data ...\ARTICLE - CARPE (Special Issue Data Harmonization)\Analisi\Data
global output ...\ARTICLE - CARPE (Special Issue Data Harmonization)\Analisi\Output
global syntax ...\ARTICLE - CARPE (Special Issue Data Harmonization)\Analisi\Syntax


use "$data\CARPE_v4_relevant_cleaned", clear    //Individual version CARPE dataset

decode country, gen(C)
tostring yearsurvey, gen(Y)
drop if Y=="9999"          //Small fix implemented in the final version of the
dataset
decode study2, gen(S)


gen CYS= C + "_" + Y + "_" + S
gen CY= C + "_" + Y


global MLwiN_path "C:\Program Files (x86)\MLwiN v2.35\x64\mlnscript.exe"
gen cons=1


runmlwin attendp cons, ///
     level4(C: cons) ///
     level3(CY: cons) ///
     level2(CYS: cons) ///
     level1(id: cons) ///
     nopause forces mlwinsettings(optimat) noversioncheck

est store CYS


runmlwin attendp cons i.study2, ///
     level3(C: cons) ///
     level2(CY: cons) ///
     level1(id: cons) ///
     nopause forces mlwinsettings(optimat) noversioncheck

est store FIXED

tabulate study2, generate(s)

runmlwin attendp cons i.study2, ///
     level3(C: cons) ///
     level2(CY: cons s2 s3 s4) ///
     level1(id: cons) ///
     nopause forces mlwinsettings(optimat) noversioncheck

est store RANDOM

lrtest FIXED RANDOM

est table CYS FIXED RANDOM ,stats(N ll) b(%4.3f) stfmt(%4.0f) varwidth(30) star
```

References

Bell, A., Jones, K.: The Hierarchical Age-Period-Cohort Model: Why Does It Find the Results That It Finds? Qual. Quant. **52**, 783–799 (2018)

Billiet, J.: Proposal for questions on religious identity. In: European Social Survey Core Questionnaire Development (2002). https://www.europeansocialsurvey.org/docs/methodology/core_ess_questionnaire/ESS_core_questionnaire_religious_identity.pdf. Retrieved 26 June 2020

Biolcati, F., Lomazzi, V., Molteni, F., Quandt, M., Vezzoni, C. Church Attendance and Religious change Pooled European dataset (CARPE) (Version: 1.0.0) (2020). https://doi.org/10.7802/2040

Blasius, J., Thiessen, V.: Assessing the Quality of Survey Data. Sage, London (2012)

Brenner, P.S.: Exceptional behavior or exceptional identity? Overreporting of church attendance in the US. Public Opin. Q. **75**(1), 19–41 (2011)

Brenner, P.S.: Cross-national trends in religious service attendance. Public Opin. Q. **80**(2), 563–583 (2016)

Bryk, A.S., Raudenbush, S.W.: Application of hierarchical linear models to assessing change. Psychol. Bull. **101**(1), 147–158 (1987)

Burkimsher, M.: Is religious attendance bottoming out? An examination of current trends across Europe. J. Sci. Study Relig. **53**(2), 432–445 (2014)

Carmines, E.G., Zeller, R.A.: Review of Reliability and Validity Assessment. SAGE, Thousand Oaks (1979)

Castegnaro, A., Dalla Zuanna, G.: Studiare la pratica religiosa: differenze tra rilevazione diretta e dichiarazione degli intervistati sullafrequenza alla messa. Polis. XX 1, 85–112 (2006)

Corbetta, P.: Social Research, Methods and Techniques. Sage, London (2003)

Davie, G.: Religion in Britain Since 1945: Believing Without Belonging. Blackwell, Oxford (1994)

Duncan, G.J., Kalton, G.: Issues of design and analysis of surveys across time. Int. Stat. Rev. **55**(1), 97–117 (1987)

Firebaugh, G.: Seven Rules for Social Research. Princeton University Press, Princeton (2008)

Glock, C.Y.: The religious revival in America. In: Zahl, J. (ed.) Religion in the Face of America. University of California Press, Berkeley (1959)

Groves, R.M.: Survey Errors and Survey Costs. Wiley, New York (1989)

Hadaway, C.K., Marler, P.L., Chaves, M.: What the polls don't show: a closer look at US church attendance. Am. Sociol. Rev. **58**(6), 741–752 (1993)

Harkness, J.A.: In pursuit of quality: issues for cross-national survey research. Int. J. Soc. Res. Methodol. **2**, 125–140 (1999)

Hout, M., Greeley, A.: What church officials' reports don't show: another look at church attendance data. Am. Sociol. Rev. **63**(1), 113–119 (1998)

Jen, M.H., Johnston, R., Jones, K., Harris, R., Gandy, A.: International variations in life expectancy: a spatio temporal analysis. Tijdschrift Voor Economische En Sociale Geografie. **101**(1), 73–90 (2010)

Kołczyńska, M., Schoene, M.: Survey data harmonization and the quality of data documentation in cross-national surveys. In: Johnson, T.P., Pennell, B.-E., Stoop, I., Dorer, B. (eds.) Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC), pp. 963–984. Wiley, Hoboken (2019)

Kołczyńska, M., Slomczynski, K.M.: Item metadata as controls for ex post harmonization of international survey projects. In: Johnson, T.P., Pennell, B.-E., Stoop, I., Dorer, B. (eds.) Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC), pp. 1011–1034. Wiley, Hoboken (2019)

Kwak, J., Slomczynski, K.M.: Aggregating survey data on the national level for indexing trust in public institutions: on the effects of lagged variables, data-harmonization controls, and data quality. In: *Harmonization: Newsletter on Survey Data*, vol. 5 (2019)

Maineri, A.M., Scherpenzeel, A., Bristle, J., Pflüger, S-M., Butt, S., Zins, S., Emery, T., Luijkx, R.: Evaluating the quality of sampling frames used in European cross-national surveys. SERISS (2017). https://seriss.eu/wp-content/uploads/2017/01/SERISS-Deliverable-2.1-Report-on-the-use-of-sampling-frames-in-European-studies.pdf. Accessed 9 Sept 2020

Meredith, W., Tisak, J.: Latent curve analysis. Psychometrika **55**(1), 107–122 (1990)

Nagin, D.S.: Group-Based Modeling of Development. Harvard University Press, Boston (2005)

Nunnally, J.C.: Psychometric Theory. McGraw-Hill, New York (1978)

Oleksiyenko, O., Wysmulek, I., Vangeli, A.: Identification of processing errors in cross-national surveys. In: Johnson, T.P., Pennell, B.-E., Stoop, I., Dorer, B. (eds.) Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC), pp. 985–1010. Wiley, Hoboken (2019)

Pisati, M.: La domenica andando alla messa. Un'analisi metodologica e sostantiva di alcuni dati sulla partecipazione degli italiani alle funzioni religiose. Polis XIV 1, 115–136 (2000)

Reitsma, J., Pelzer, B., Scheepers, P., Schilderman, H.: Believing and belonging in Europe. Eur. Soc. **14**(4), 611–632 (2012)

Ruiter, S., van Tubergen, F.: Religious attendance in cross-national perspective: a multilevel analysis of 60 countries. Am. J. Sociol. **115**(3), 863–895 (2009)

Voas, D.: The rise and fall of fuzzy fidelity in Europe. Eur. Sociol Rev. **25**, 155–168 (2009)

Vezzoni, C., Biolcati, F.: Calibrating self-reported church attendance questions in online surveys. Experimental evidence from the Italian context. Soc. Compass. **66**(4), 596–616 (2019)

Vezzoni, C., Biolcati-Rinaldi, F.: Church attendance and religious change in Italy (1968–2010). A multilevel analysis of pooled datasets. J. Sci. Study Relig. **54**(1), 100–118 (2015)

Voas, D., Doebler, S.: Secularization in Europe: religious change between and within birth cohorts. Relig. Soc. Centr. East. Eur. **4**(1), 39–62 (2011)

Wolf, C., Schneider, S.L., Behr, D., Joye, D.: Harmonizing survey questions between cultures and over time. In: *The SAGE Handbook of Survey Methodology*, pp. 502–524. London: SAGE Publications Ltd (2016)

Yang, Y., Land, K.C.: Age–period–cohort analysis of repeated cross-section surveys: Fixed or random effects? Sociol. Methods Res. **36**(3), 297–326 (2008)

Yang, Y., Land, K.C.: Age–Period–Cohort Analysis: New Models, Methods, and Empirical Applications. CRC Press, Boca Raton (2013)

Yang, Y., Fu, W.J., Land, K.C.: A methodological comparison of age-period-cohort models: the intrinsic estimator and conventional generalized linear models. Sociol. Methodol. **34**, 75–110 (2004)

Zieliński, M.W., Powałko, P., Kołczyńska, M.: The past, present, and future of statistical weights in international survey projects: implications for survey data harmonization. In: Johnson, T.P., Pennell, B.-E., Stoop, I., Dorer, B. (eds.) Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC), pp. 1035–1052. Wiley, Hoboken (2019)

Affiliations

Ferruccio Biolcati[1] · **Francesco Molteni[1]** · **Markus Quandt[2]** · **Cristiano Vezzoni[1]**

Francesco Molteni
francesco.molteni@unimi.it

Markus Quandt
markus.quandt@gesis.org

Cristiano Vezzoni
cristiano.vezzoni@unimi.it

[1] Department of Social and Political Sciences, University of Milan, Via Conservatorio, 7, 20122 Milan, MI, Italy

[2] Research Data Center "International Survey Programs", GESIS – Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, 50667 Koeln, Germany