**Appendix A:**

|  |
| --- |
| Table A1 Number of respondent by country per year. |
|  | **‘02** | **‘03** | **‘04** | **‘05** | **‘06** | **‘07** | **‘08** | **‘09** | **‘10** | **‘11** | **‘12** | **‘13** | **‘14** | **‘15** | **‘16** | **‘17** | **‘18** | **‘19** | **‘20** | **Total per country** |
| **Austria** | 0 | 19 | 0 | 15 | 0 | 0 | 21 | 0 | 16 | 33 | 8 | 0 | 4 | 39 | 35 | 0 | 80 | 3 | 0 | **273** |
| **Belgium** | 17 | 3 | 35 | 3 | 15 | 15 | 19 | 70 | 19 | 911 | 161 | 0 | 54 | 6 | 68 | 16 | 68 | 7 | 0 | **1487** |
| **Denmark** | 8 | 5 | 10 | 0 | 6 | 6 | 19 | 2 | 37 | 44 | 13 | 24 | 19 | 0 | 0 | 8 | 18 | 0 | 0 | **219** |
| **Finland** | 1 | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 4 | 0 | 2 | 0 | 9 | 0 | 8 | 3 | 4 | 0 | 0 | **35** |
| **France** | 0 | 0 | 0 | 0 | 20 | 15 | 67 | 3 | 37 | 102 | 30 | 48 | 53 | 14 | 41 | 12 | 94 | 40 | 0 | **576** |
| **Germany** | 1 | 15 | 22 | 4 | 20 | 2 | 28 | 2 | 55 | 639 | 92 | 53 | 17 | 9 | 36 | 38 | 71 | 10 | 0 | **1114** |
| **Iceland** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | **3** |
| **Ireland** | 0 | 2 | 0 | 0 | 2 | 4 | 1 | 7 | 3 | 14 | 4 | 11 | 6 | 1 | 1 | 8 | 4 | 10 | 0 | **78** |
| **Italy** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 13 | 0 | 0 | 0 | 13 | 13 | 32 | 0 | **73** |
| **Luxembourg** | 0 | 14 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **25** |
| **the Netherlands** | 30 | 1 | 23 | 2 | 28 | 9 | 26 | 17 | 40 | 767 | 71 | 8 | 45 | 2 | 18 | 56 | 43 | 3 | 0 | **1189** |
| **Norway** | 11 | 2 | 13 | 0 | 12 | 0 | 17 | 2 | 16 | 9 | 16 | 0 | 17 | 0 | 23 | 0 | 16 | 14 | 0 | **168** |
| **Portugal** | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 4 | 0 | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | **17** |
| **Spain** | 9 | 0 | 15 | 0 | 10 | 2 | 8 | 0 | 0 | 8 | 0 | 22 | 0 | 4 | 0 | 30 | 0 | 13 | 1 | **122** |
| **Sweden** | 18 | 0 | 8 | 2 | 18 | 4 | 12 | 7 | 13 | 102 | 40 | 38 | 29 | 1 | 27 | 8 | 19 | 23 | 0 | **369** |
| **Switzerland** | 4 | 0 | 4 | 3 | 6 | 1 | 25 | 10 | 18 | 635 | 26 | 5 | 22 | 2 | 34 | 62 | 24 | 1 | 0 | **882** |
| **United Kingdom** | 26 | 4 | 0 | 0 | 0 | 0 | 49 | 30 | 46 | 725 | 42 | 14 | 41 | 18 | 37 | 5 | 85 | 0 | 0 | **1122** |
| **Total per year** | **125** | **66** | **141** | **30** | **142** | **60** | **295** | **155** | **304** | **3991** | **506** | **239** | **316** | **96** | **329** | **260** | **539** | **157** | **1** | **7752** |
| *Note: included are the* *European Social Survey (ESS), the World Value Survey (WVS), the European Value Survey (EVS), 2000 Families, and EurIslam* |

**Appendix B: Harmonization of Islamic religiosity variables**

**Attendance**

|  |  |  |
| --- | --- | --- |
| **Survey source** | **Formulation original questions** | **Original answering option and recoding thereof in colour codes.** |
| **Final variable**  |  | **0 Never to less than yearly** | **1 yearly to monthly** | **2 weekly** | **3 more than weekly** |
| European Social Survey | How often do you attend religious services apart from special occasions?   | never | less often | only on special holy days | at least once a month | once a week | more than once a week | every day |
| World Value Study / European Value Study | How often do you attend religious services?  | never, practi-cally never | less often | once a year | other specific holi-days | only on special holy days/ Christ-mas/ Easter days | once a month | once a week | more than once a week |
| European Value Study (as of round 2017) | Apart from weddings, funerals and christenings, about how often do you attend religious services these days?   | never, practi-cally never | less often | once a year | only on special holy days | once a month | once a week | more than once a week |
| 2000Families | 'How often do you attend religious services or go to a place of worship'.3'around once a month' 5'every day' | never | less often | only on special holy days' | around once a month' |  | once a week' | every day' |
| EurIslam | 'How often do you go to the Mosque or another place of worship for religious reasons?' | Never |  | Rarely/ only on special occasions |  |  | Weekly | Daily |

Religious identification

|  |  |  |
| --- | --- | --- |
| **Survey source** | **Question and answer formulation** | **Subdimension** |
| ESS | How religious are you?0 ‘Not at all religious’ – ‘10 very religious’ | Self-identified religiousness |
| WVS, EVS  | Are you a religious person?0'conviced atheist', 1'not a religious person', 2'a religious person' |
| WVS, EVS  | How important is god in your life?0 'not at all important' – 9'very important' | Importance |
| 2000familes | 'how important is religion to the way you live your life'.0'totally unimportant' 1'farily unimportant' 2'neither important nor unimportant' 3'important' 4'very important' | Importance |
| EurIslam | To what extent do you see yourself as a Muslim?'not at all'1 'hardly' 2 'somewhat' 3 'strongly' 4 'very strongly'. | Self-identified religiousness |

Frequency of prayer

|  |  |  |
| --- | --- | --- |
| **Survey source** | **Formulation original questions** | **Original answering option and recoding thereof in colour codes.** |
|  |  | 0 (practically) never | 1 infrequent | 2 at least weekly | 3 daily or more often |
| European Social Survey | How often do you pray apart from at religious services | never | less often | only on special holy days | at least once a month | once a week | more than once a week | every day |
| World Value Study  | How often do you pray? | only when attending religious service | never, practi-cally never | less often | once a year | only on special holy days/ Christ-mas/ Easter days | several times each weak | once a day | several times a day |
| European Value Study  | How often do you pray outside of religious services? | never | less often | several times a year | at least once a month | once a week | more than once a week | every day |
| 2000families | Apart from religious service, how often do you pray (namaz)? | never | less often | only on special holy days | at least once a month | once a week |  | every day | five times a day |
| EurIslam | How often do you pray? | never |  | only on special occasions |  | Once or a few times a week  |  | once a day | several times a day |

**Appendix C: Operationalization of hostile public opinion and veil ban laws**

*1. Public opinion*

We aggregated the answers to five items in the biannual European Social Survey data (2002-2018) (ESSa; ESSb; ESSc). This was done per country on their national representative full datasets. Missing score were imputed linearly if scores were available for the same country from a survey prior to and after the countryyear with a missing score. Where necessary, the answers were reversed to bring them in line with the direction of the final coding: a higher score indicate more hostility.

Allow many/few immigrants of different race/ethnic group from majority [imdfetn]

Allow many to come and live here; Allow some; Allow a few; Allow none

Allow many/few immigrants from poorer countries outside Europe [impcntr]

Allow many to come and live here; Allow some; Allow a few; Allow none

Immigration bad or good for country's economy [imbgeco]

0 Bad for the economy; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10 Good for the economy

Country's cultural life undermined or enriched by immigrants [imueclt]

0 Cultural life undermined; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10 Cultural life enriched

Immigrants make country worse or better place to live [imwbcnt]

0 Worse place to live; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10 Better place to live

2. Veil Ban in laws

Based on overview studies and reports, followed by per country checks, we coded whether veils, niqaab, burqas and other garment alike were banned from public places by law. The precise list of sources and data can be obtained from the authors, but the core sources include the following (all accessed 31/1/2022):

1. <https://www.dw.com/en/full-face-veil-ban-how-laws-differ-across-europe/a-44049185> (1/6/2018)
2. <https://www.ohchr.org/Documents/Issues/Women/WRGS/VeilinEuropereport.pdf> (2019 or later)
3. En.Wikipedia.org
	1. <https://en.wikipedia.org/wiki/Anti-mask_law>
	2. <https://en.wikipedia.org/wiki/Niq%C4%81b#Enforcement,_encouragement_and_bans>
	3. <https://en.wikipedia.org/wiki/Hijab_and_burka_controversies_in_Europe>
	4. <https://en.wikipedia.org/wiki/Hijab_by_country>

Codes are given from the year of the enactment of a law onwards, unless the law was modified or replaced in which case the a new score is given.

The item:

Are laws enacted that explicitly forbid to wear garments in public space including religious garments of women, whereby this enacted has taken place in the context of the post 2000 debates on Islam [Veilban]

Bans on face covering in public get togethers or garments that make identification difficult from the 1970s and 1980s (e.g. 1970s Italian ban; 1985 German ban) are not included (i.e. coded as 0)

0: no ban

1: a (ful) public ban of face coverings

0,5: either a ban of headscarfs in very specific places (education, court) or a full ban (of face covering) but only in parts of the country (e.g. Spain).

0,25: partial ban in only parts of the country

**Appendix D: EFA on macro variables**

|  |
| --- |
| **KMO and Bartlett's Test** |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | .652 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 1621.867 |
| df | 45 |
| Sig. | .000 |

|  |
| --- |
| **Total Variance Explained** |
| Factor | Initial Eigenvalues | Extraction Sums of Squared Loadings | Rotation Sums of Squared Loadingsa |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total |
| 1 | 4.583 | 45.830 | 45.830 | 4.415 | 44.148 | 44.148 | 4.186 |
| 2 | 2.321 | 23.205 | 69.035 | 2.080 | 20.800 | 64.948 | 2.215 |
| 3 | 1.087 | 10.870 | 79.905 | .560 | 5.601 | 70.549 | 2.174 |
| 4 | .666 | 6.663 | 86.568 |  |  |  |  |
| 5 | .604 | 6.041 | 92.608 |  |  |  |  |
| 6 | .350 | 3.498 | 96.106 |  |  |  |  |
| 7 | .205 | 2.048 | 98.154 |  |  |  |  |
| 8 | .106 | 1.062 | 99.216 |  |  |  |  |
| 9 | .060 | .599 | 99.815 |  |  |  |  |
| 10 | .018 | .185 | 100.000 |  |  |  |  |
| Extraction Method: Principal Axis Factoring. |
| a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance. |



|  |
| --- |
| **Structure Matrix** |
|  | Factor |
| 1 | 2 | 3 |
| VeilBan\_mean | -.128 | -.006 | .367 |
| proport\_mean | -.372 | .195 | .673 |
| AntiDisc\_mean | -.039 | .897 | .159 |
| Path2Cit\_mean | .176 | .923 | .106 |
| path2perm\_mean | .339 | .516 | -.194 |
| imdfetn\_mean\_mean | .984 | .208 | -.315 |
| impcntr\_mean\_mean | .940 | .219 | -.312 |
| imueclt\_mean\_mean | .835 | .328 | -.738 |
| imwbcnt\_mean\_mean | .859 | .195 | -.617 |
| imbgeco\_mean\_mean | .773 | -.135 | -.625 |
| Extraction Method: Principal Axis Factoring.  Rotation Method: Oblimin with Kaiser Normalization. |

|  |
| --- |
| **Appendix E:** |
| Table A2. Summarized Multilevel regression models estimating the context-dependent impact of Islamic religiosity on voting amongst Muslim Europeans (2002-2020). |
|  | Model 1a | Model 1b | Model 1c | Model 2a | Model 2b | Model 2c | Model 3a | Model 3b | Model 3c | Model 4a | Model 4b | Model 4c |
|  | **Exclusion policies** | **Veil ban** | **Hostile public attitudes** | **Disproportionality** |
| **FIXED EFFECTS** | B | B | B | B | B | B | B | B | B | B | B | B |
| *Attendance(more than once a week=ref)\*macro* |  |  |  |  |  |  |  |  |  |  |  |  |
|  Never  | .002 |  |  | .039 |  |  | -.002 |  |  | -.004 |  |  |
|  Few times a year  | .002 |  |  | -.004 |  |  | -.024 |  |  | -.002 |  |  |
|  Up to once a week | .002 |  |  | .056 |  |  | .023 |  |  | .003 |  |  |
| *Identification\*macro* |  | -.001 |  |  | -.007 |  |  | -.001 |  |  | -.000 |  |
| *Individual praying\*macro* |  |  | -.000 |  |  | -.013 |  |  | .010 |  |  | .001 |
| **INTERACTION VARIABLES** |  |  |  |  |  |  |  |  |  |  |  |  |
| *Exclusion Policies* | -.001 | .001 | .002 |  |  |  |  |  |  |  |  |  |
| *Veil ban* |  |  |  | .091 | .089. | .109 |  |  |  |  |  |  |
| *Hostile Public attitudes* |  |  |  |  |  |  | -.003 | -.014 | -.032 |  |  |  |
| *Disproportionality* |  |  |  |  |  |  |  |  |  | -.007 | -.007# | -.010\* |
| Intercept | -.749\* | -.871\* | -.917\*\* | -.682\* | -.805\* | -822\* | -.715 | -.733 | -.637 | -.685\* | -.789\* | -.797\* |
| **MODEL STATISTICS** |  |  |  |  |  |  |  |  |  |  |  |  |
| BIC | 9412.837 | 9318.128  | 9319.745 | 9380.527 | 9303.676 | 9304.281 | 9384.685 | 9307.927 | 9307.605 | 9401.614 | 9313960 | 9318.724 |
| Nindividual = 7752 ; Ncountryyear = 240 ; Nyear = 17\*\*\* P<0.001 \*\*p<0.01 \*p<0.05 p<0.1 (macro-level effects only)Notes: 1)Per dimension of religiosity the interaction models are estimated (i.e. all four macro-levels are interacted with one religiosity dimension per model. 2)These coefficients are included at the macro-level in the interaction models, leading to overestimation of the p value (Type-1 error); coefficient here are given to assess the shape of the effects (e.g. whether the effects are fanning in or out).3)Control variables were included with each model.  |
|  |

**Appendix F:**

In this appendix, we present more detailed results of the analyses conducted to assess how the synchronization of datasets might influence or explain results, i.e. whether some differences found (or lack thereof) is likely to be ascribed to differences between the dataset and procedures followed to synchronize them, or whether they are more likely due to substantive differences (some of which are tested in the interaction analyses in the manuscript. As will be described in more detail below, this is particularly important for the analyses of the content-dependency of the relationships.

In the below we take the following four steps to assess this:

1. We briefly describe the datasets in a bit more detail to provide a background to assessment in this appendix
2. Based on 1. and Appendix B, we discuss, how the differences in the measurement might influence the relationships and conclusion of focus in this study.
3. We present the overall descriptives of the core variables per dataset, the get a first grasp at the patterns.
4. Mann-Whitney U test on matched samples (i.e. the same ethnic groups in the same country of living and year) to assess where we find different distributions, even on seemingly (see 1. And 2.) similar samples.

Based on these assessment, we argue that (beside general caution in drawing too bold conclusions), two steps are crucial to study the overall relationships between the three religiosity dimensions and electoral participation across the diverse Muslim population in Western Europe:

1. As the dimension of religiosity might work differently between migration generation and because the differences between survey sources partly aligns with survey differences in terms of including first-generation migrants, the analyses need to be assessed separately for first-generation migrant Muslim citizens and other Muslim citizens.
2. As the differences between survey sources for religious identification partly aligns with survey differences in terms of countries and years covered, the cross-level interaction analyses need to be controlled for differential effects by survey source (i.e. control interaction terms between the religiosity dimensions and the survey source control variable need to be included).

**1. Sampling and item differences between datasets**

ESS, EVS and WVS are dataset based on random samples of the respective populations in a country, among which Muslim citizens are samples too. Their data collections rely on invitations and questionnaire in the formal language of the population survey. This is likely to lead to lower responses among Muslims among first-generation migrants who speak the destination-country language less; the demand on them filling out the surveys is higher. At the same time, all groups of Muslim citizens, also from smaller so-called origin countries and converts, are part of the sampling frame, Muslims in which Muslim lives are included, and the invitation does not highlight Islam as an major issue in the survey which prevents substantive selectivity in response.

The 2000Families data was sample with a focus on high-sending regions in Turkey from a perspective of labour migrants in the 60s and 70s. The samples were random, but the sampling frame was not based on destination country. The sending regions were selected based on connections with specific countries and the final dataset covers multiple West European countries, most notably Austria, Belgium, Denmark, France, Germany, the Netherlands, Sweden, Switzerland, and the United Kingdom. Evidently, the data only include Turkish Muslims, which is however the largest Muslim population in Western Europe. The surveys were conducted in Turkish. It might be expected that the 2000Families respondent are relatively more religious and well-integrated in the Turkish community and to include the subgroups that are less acculturated in Western European politics better than the prior discussed data collections.

The EURISLAM data sample population in six European countries: Belgium, France, Germany, The Netherlands, Switzerland, and the United Kingdom. In each country the goal was to collect data on 150 to 250 Muslims per each of the following ethnic groups: Moroccan, Pakistani, Turkish and (ex-)Yugoslavian. In some of these countries certain groups are relatively small and little organized, while in others the are the largest group and well organized. EURISLAM used bilingual questionnaires, which likely include the first-generation migrant-background Muslim population well. The focus of the research might have led to more religious participation to reply more. The codebook also notes that particularly among UK Muslims in EURISLAM all identify religious. Moreover, the sampling took pace based on surname from online telephone directories, which is likely to oversample older and undersample interethnically married Muslims.

**2. The potential impact of measurement differences**

Considering this all, there is clear overlap between 2000Families and EURISLAM and between EES, EVS and WVS; both sets having there unique strengths in terms of reaching populations the other sets of survey excludes or is likely to reach. There is however no direct reasons to expect that the relationships between religiosity and electoral participation differ greatly based on these sampling frames beyond the differences in terms of migration generation and language comprehension, which we take explicitly into account in our results section. The difference in degree of organization might be considered relevant as well but this is part of the mechanisms tested (i.e. it might influence the degree of mosque attendance) more than that it directly leads to expectation about attendance functioning differently.

In Appendix B, the different ways in which the dimensions or religiosity are measured is discussed in more detail. For **attendance** and **prayer**, the concepts are surveyed with fairly fundamentally similar questions and very comparable ordinal answering options (i.e. not ‘often’ or ‘hardly’, but ‘weekly’ and ‘less than once a month’) that could be regrouped to create similar scale across surveys. Consequently, certainly after controlling for survey source in the models, it is unlikely that differences in the relationships between attendance and prayer on the one hand and electoral participation on the other between the survey sources are due to measurement. They might be due to the above discussion differences in population as well as chance. However, our focus is not primarily on explaining all such differences, but on establishing whether overall influences on electoral participation of the different religiosity dimensions. Consequently, we will come back to such differences below and in the main text, but our main interest is in whether we find a relationship and in what direction.

If the distributions are very, though, this might distort the models in which we model differences in the effects of the religiosity dimensions, as the countries and years available are clearly dependent on the survey source. In other words, the differences in item and answer formulation are by definition related to the macro-level indicators’ score. If differences in distribution exist this might distort the **cross-level interaction** models. Most likely this would lead to increased noise and a higher chance of false negative conclusions.

For the strength of **religious identification** the items used to create a variable across data sets are more varied that for attendance and prayer, and this more easily leads to differences in distributions and consequently to more serious distortions, on two fronts. As discussed directly above, this might bias the cross-level interactions effect. In addition, if the different items used for different data sources to measure identification, really tap different (sub)dimensions – something that was impossible for prayer and attendance given that questions were fundamentally similar – that might actually also harm the direct effect estimated and this then cannot be solved by including the survey type control and multilevel estimations. Here we should bring to memory that for religious identification we already took part of this issue into account when constructing the variable, by the z-score of each respondent compared to all respondent having answer the same survey item, which takes into account that some items might lead to more spread to the question formulation. At the same time, this cannot account for differences in meaning between items, which might be part of the issue too.

Interestingly though, the degree to which the distribution for attendance and prayer difference between data sources provides some of a bench mark which help to assess how likely differences in distribution found for the degree of religious identification as likely the result of the different items being used across survey types. If similar differences (in amount, location and direction) are found for identification compared to prayer and attendance, than this seems less likely. As the differences more noteworthy than additional caution and analyses are needed.

**3. Descriptive statistics for religiosity per data source**

Above we discussed the differences between the datasets in substantive and methodological terms. Completely ignored that the dataset sample different part of West European Muslim population, we report the means and standard deviation for the three religiosity dimensions per data set below. While we should be cautious interpreting this figures, they do give a first indication of whether the figures are markedly different and how. In other words, it might help guide our focus in the remainder of this assessment.

Given all differences as discussed above, it is rather noticeable that the score are rather similar actually. In all datasets average attendance is between 1 and 1.5, meaning between irregularly up to a few times a month and far removed from weekly, with all standard deviations between .78 and 1.09. Prayer’s mean hoovers between 1.6 and 2.2, centering around 2 ‘weekly but not daily’, in all survey sources. Moreover, in all sources the average score on prayer is higher than the average score on attendance.

Lastly, regarding identification, the means are around 0 and the standard deviation around 1, which is a direct artefact of the harmonization procedure used. As standardized is done per item over all countryyears, it is noteworthy though that even after splitting the survey sources using the same item (see Appendix B) still have rather similar means (EVS, WVS) suggesting that also populations across surveys are rather similar.

|  |
| --- |
| Mean and s.d. per religiosity dimension per data source. |
|  | Survey Source (N) |
|  | Max sample (7940-7978) | ESS (2906-2915) | WVS(153-161) | EVS(585-612) | 2000Families(1140-1146) | EurIslam(3130-3159) |
| Attendance (0-3) |
| Mean | 1.14 | 1.21 | 1.25 | 1.04 | 1.38 | 1.00 |
| Sd | .89 | .93 | 1.09 | .99 | .78 | .83 |
| Praying (0-3) |
| Mean | 1.88 | 2.03 | 1.87 | 2.17 | 1.60 | 1.78 |
| sd | 1.15 | 1.12 | 1.07 | 1.02 | 1.04 | 1.21 |
| Identification (-5.12 - 1.18) |
| Mean | -.03 | .01 | .09 | .01 | -.05 | -.07 |
| Sd | .99 | 1.00 | .82 | .86 | 1.00 | 1.01 |

Moreover, in each of the five dataset all three indices correlate positively with each other. All correlations are between .244 and .539 (all p<.003). On the one hand, this indicates that multicollinearity is unlikely to be an issue, while also indicating that all dimensions relate positively to each other. In other words, the direction of the three dimensions correspond similarly in each of the survey sources. Together with the other information provided here, in Appendix B and the manuscript indicate that the operationalized variables are very likely to measure the same dimension in very similar ways.

**4. Mann-Whitney U Tests on matched subsamples**

Against the background of the above, we will now seeks similar subsamples from different survey sources and test whether the religiosity dimensions have unequal distributions. To test this we will use the **Mann-Whitney U Test**, which is fitting for non-parametric data.

When we talk about matched subsamples, we refer to samples from the same country of living, sampled from the same origin group and the same year (but see below), with at least 30 observation in each subsample, which allows for differences to actually be statistically significant (i.e. not biasing results in favor of not finding differences in distribution).

As the WVS represents a very small and diver sample, this is not included below. Here we should note that the WVS and EVS sampling frames and measurements are almost identical (see 1, and Appendix B), making this comparison being absent less critical than for instance not being able to make comparisons across the two identified clusters (ESS, EVS, WVS vs 2000Families, EURISLAM).

For each of the matched samples that show a significantly different distribution we also provide the distributions of the variables graphically (see below), which allows assessing the direction and extent of a difference (please note that the graph produced by SPSS are providing absolute numbers; the comparison between the two should not be made based on the absolute length of the bars but the relative length of the bars compared to the length of the other bars for the same survey source subsample). The differences found, will be assessed in relations to the differences between the surveys as discussion under 1. and 2.

In total we could identify 13 pairs of matched subsamples to run the tests on (see below). The EVS and ESS (2018) both sample from the general population so we consider this to be the same ‘ethnic background’ or ‘country of origin’. In both sufficient cases are present to make comparison for Austria, France and the United Kingdom. As each of these samples is rather small, and given the sampling procedures, we also combined the three samples (weighing the subsamples so that for each source survey the relative size of each country sample is equally large, while keeping the overall samples sizes the same) – see 1.4 in the table below. All other comparisons include larger subsamples, so these were kept separate. In comparison 3 and 4 (see table below), we pooled the ESS data from 2010 to 2012 and did the same for the 2000Families data. By doing this, we could make these additional comparisons while comparing rather substantive sample. Grouping our adjacent years would not lead to more useful matched samples.

|  |
| --- |
| Independent-Samples Mann-Whitney U Test *Retain or reject null hypotheses that the distribution of the variable is the same across samples (p value is given)* |
| Data sets and year | Group in country | Atten-dance | Praying | Identi-fication |
| 1. EVS and ESS 2018  | 1. Muslims in random sample general population Austria (32-35 and 50 cases) | .482 | .733 | .383 |
|  | 2. Muslims in random sample general population France (64-65 and 30 cases) | .678 | .947 | .004 |
|  | 3. Muslims in random sample general population United Kingdom (49 and 36-37 cases) | .864 | .941 | .018 |
|  | 4. Muslims in random sample general population Austria, France, United Kingdom (146-149 and 108-109 cases) | .781 | .464 | .830 |
| 2. EURISLAM and 2000Families 2011  | 1. Turkish background Muslims in Belgium (225 and 263 cases) | .929 | .001 | .001 |
|  | 2. Turkish background Muslims in Germany (119 and 164 cases) | .001 | .877 | .672 |
|  | 3. Turkish background Muslims in The Netherlands (204 and 130 cases) | .002 | .973 | .115 |
| 3. EURISLAM 2011 and ESS (2010-2012) | 1. Pakistani background Muslims in Belgium (37 and 106 cases) | .714 | .056 | .001 |
|  | 2. Turkish background Muslims in Belgium (33 and 224 cases) | .758 | .122 | .008 |
|  | 3. Turkish background Muslims in 5. Germany (32 and 119 cases) | .874 | .329 | .626 |
|  | 4. Pakistani background Muslims in United Kingdom (41 and 256 cases) | .001 | .001 | .012 |
| 4. ESS and 2000-Families (2010-2012  | 1. Turkish background Muslims in Belgium (33 and 156 cases) | .399 | .925 | .001 |
|  | 2. Turkish background Muslims in Germany (32 and 125 cases) | .016 | .670 | .710 |

|  |
| --- |
| Distributions for the matched sample that were significantly differently distributed |
| Row in table above | Attendance | Praying | Identification |
| 1.2 |  |  | A graph with blue and green squares  Description automatically generated |
| 1.3 |  |  | A graph with blue squares and black text  Description automatically generated |
| 2.1 |  | A picture containing text, screenshot, diagram, font  Description automatically generated | A picture containing text, screenshot, diagram, font  Description automatically generated |
| 2.2 | A picture containing text, diagram, screenshot, font  Description automatically generated |  |  |
| 2.3 | A picture containing text, diagram, screenshot, font  Description automatically generated |  |  |
| 3.1 |  |  | A graph with a bar chart  Description automatically generated |
| 3.2 |  |  | A graph with blue squares and black text  Description automatically generated |
| 3.4 | A graph with blue squares and black text  Description automatically generated | A graph with blue and black squares  Description automatically generated | A graph with blue and green bars  Description automatically generated |
| 4.1 |  |  | A graph with blue and green squares  Description automatically generated |
| 4.2 | A graph with blue squares and black text  Description automatically generated |  |  |

For each of the thirteen pairs three comparisons are tested. Of these comparisons, two-thirds (26 of 39) show comparative distributions between datasets (and often with high p-values). This suggest that, even despite the different samples being included (see 1), the measurement are rather comparable. But before drawing conclusions, let us zoom in on the comparison that do show different distributions.

First, it is noteworthy, that the only comparison showing three different distributions is found for the only comparison including the Muslim population in the UK in the EURISLAM data, as it also this groups that was singled out in the codebook of EURISLAM (see 1.) as showing a clear sampling bias, Without these, 72% of the comparisons show no difference in distribution.

Second, focus on which surveys differ from which it must be observed that in line with the discussion above (see 1.), the **ESS<>EVS** comparison shows the fewest differences (17%). Comparisons 2.1-2.3 indicate that while **EURISLAM and 2000Families** both surveyed in Turkish, there are marked differences (44%) in distribution; these differences are difficult to ascribe to the formulation of the questions formulation as for instance the attendance questions in these two sources are highly similar (see Appendix B). Still having fewer answering options in EURISLAM might have forced people to pick a higher or lower frequency; however, the figures below (row 2.2 and 2.3) show that the daily attendance category is chosen less frequent in EURISLAM, while the next option in both surveys was weekly, which implies that less frequent choice for daily in EURISLAM is not an artefact of having to choose between daily and a lower next option than in 2000Families. Moreover, in EURISLAM relatively more people selected never, whereas 2000Families offer an option between never and only on special occasions (“less often”) which we grouped with never. If anything that should bias towards more people in 2000Families being classified in the lowest category, not fewer. In other words, the figures suggest that the difference found is at least partly, if not largely, substantively linked differences in samples, not linked to item formulation. In line with the sampling being different, we also find differences in distribution for the **ESS<>EURISLAM** and **ESS<>2000Families** pairs, and more so than for the more similar ESS<>EVS. Respectively 22% or 40% (depending on whether the UK sample is included – see above) and 33% of the pairs show a significantly different distribution. We see that the ESS samples have higher identification than the EURISLAM samples in three of four cases and that in one case the 2000Families show a higher identification than ESS. In both comparisons, this might be linked to differences in the items used to measure identification (Appendix B). Altogether, considering the results from the perspective of the differences between the data sources across the comparison, suggests that most differences are (partly) explained by sample differences not differences in item formulation, with exception of the differences found for identification.

Third, and directly related to the points above, we can consider where distributions differ per dimension of religiosity. This shows that respectively 69% and 85% of the comparisons for attendance and prayer show similar distributions, while this is only 46% for identification. As discussed above, the use of similar items for attendance and prayer across surveys make that the results for these two dimensions can also serve as a benchmark for judging the results for identification. In line with the conclusion at the end of the paragraph above, it seems prudent to be more careful when drawing conclusions on the impact of identification across surveys. More specifically, it is more likely that for the dimension of identification differences in results across surveys are (partly) related to the differences in the items included and not just due to differences in the sample. Given the positive correlations (see 3. above) and the inclusion of controls for the source surveys in the models this is not expected to have the largest impact on the direct effects of identification (i.e. differences in the relationship between identification and electoral participation can be found, but this is still likely due to sample differences). However, given that the extent to which specific groups of first- and second-generation migrants are sampled and which countries and years are included in each survey, these differences in distributions across survey might related to the macro-level variables and thus the differences in effect across country-year as well relate to different effect between generations.

Based on this assessment, we thus argue that beside general caution in drawing too bold conclusions, two steps are crucial to study the overall relationships between the three religiosity dimensions and electoral participation across the diverse Muslim population in Western Europe:

1. As the dimension of religiosity might work differently between migration generation and because the differences between survey sources partly aligns with survey differences in terms of including first-generation migrants, the analyses need to be assessed separately for first-generation migrant Muslim citizens and other Muslim citizens.
2. As the differences between survey sources for religious identification partly aligns with survey differences in terms of countries and years covered, the cross-level interaction analyses need to be controlled for differential effects by survey source (i.e. control interaction terms between the religiosity dimensions and the survey source control variable need to be included).

**Appendix G:**

|  |
| --- |
| B-coefficients per country based on Model 1b from Tabel 3 being estimated per country separately |
| Country | Attendance | Identification | Praying |
| Austria | -0,04 | -0,04 | 0,06 |
| Belgium | 0,01 | 0 | -0,01 |
| Denmark | -0,05 | -0,02 | **0,08** |
| Finland | -0,08 | -0,04 | 0,05 |
| France | 0,02 | -0,02 | 0,03 |
| Germany | -0,00 | 0,03 | -0,00 |
| Ireland | 0,02 | -0,04 | 0,07 |
| Italy | -0,06 | -0,05 | 0,17 |
| Netherlands | -0,01 | -0,01 | **0,03** |
| Norway | -0,02 | 0,02 | 0,02 |
| Spain | -0,05 | -0,05 | ***0,08*** |
| Sweden | -0,00 | 0,01 | 0,02 |
| Switzerland | ***-0,04*** | -0,01 | 0,01 |
| United Kingdom | 0,02 | 0,01 | 0,02 |
| Notes: Countries with fewer than 50 cases are given in grey; coefficients in bold are significant at **p<.05**, coefficients in italics and bold at ***p<.1***. |