# Improving and Validating Survey Estimates of Religious Demography Using Bayesian Multilevel Models and Poststratification

Christopher Claassen* Richard Traunmüller†

November 29, 2016

*School of Social and Political Sciences, University of Glasgow
†Department of Social Sciences, Goethe University Frankfurt

# Improving and Validating Survey Estimates of Religious Demography Using Bayesian Multilevel Models and Poststratification

**Abstract**

Religious group size, demographic composition, and the dynamics thereof are of interest in many areas of social science, including migration, social cohesion, parties and voting, and violent conflict. Existing estimates however are of varying quality because many countries do not collect official data on religious identity. We propose a method for accurately measuring religious group demographics using abundant existing survey data: Bayesian multilevel regression models with post-stratification, or MRP. We illustrate this method by estimating the demography of Muslims, Hindus, and Jews in Great Britain over a 20-year period, and validate it by comparing our estimates to UK census data on religious demography. Our estimates are very accurate, differing from true population proportions by as little as 0.37 (Muslim) to 0.02 (Jewish) percentage points. These findings have implications for the measurement of religious demography as well as small group attributes more generally.

Keywords: Religion, Europe, demography, survey methods, multilevel regression with post-stratification (MRP), small-area estimation (SAE), Bayesian methods.

# 1. Introduction

The size of religious groups, their demographic composition, and their dynamics over time are of interest in many areas of the social sciences, including social cohesion and change (DiPrete et al 2011; Putnam and Campbell 2010; Voas et al 2002; Voas and Chaves 2016), migration (Geddes 2003; Levitt 2003), parties and voting (Gelman et al 2008; Lipset and Rokkan 1967), and violent conflict (Fearon and Laitin 2003; Toft et al 2011). Religious demographic change is, if anything, even more of interest outside the academy, especially in Europe, where religious diversity is increasing rapidly and religion now plays a central role in vexing political debates around migration, citizenship, and identity (Adida et al 2016).

Yet we know surprisingly little about the demography of small religious minorities, both in Europe and beyond, due to severe data limitations. For a variety of historical and political reasons, official census data on religious identity is only gathered in a few countries and is entirely lacking for major cases such as the United States, France, Germany, or the Netherlands.[1] In addition, while there have been a number of recent efforts to provide global measures of religious diversity (Brown and Patrick 2015; Johnson and Grim 2013; Maoz and Henderson 2013), these projects combine data of varying quality, often in an *ad hoc* fashion. They also only provide the marginal distributions of religious identity, rather than the full joint distributions. In other words, they provide no information regarding the socio-demographic composition of small religious groups, such as their age-structure, sex ratio, or socio-economic resources.

In this paper, we propose a method for measuring the size, demographic composition, and dynamics of religious minorities using secondary survey data. While official data on religious identity is rare, public opinion survey data is common. Indeed, questions on religious identity are typically included on several cross-national survey projects that have carried out thousands of surveys on national samples across dozens of countries over the last two or three

[1] Among the EU27 only 14 states included such a question in their most recent censuses (Johnson and Grim 2013).

decades.

The use of secondary survey data for measuring religious demography is not straightforward however. Religious minorities are too small to yield reliable estimates in typical survey samples of a few thousand respondents. Although surveys can be pooled across time to increase the size of the sample (e.g. Smith 2005; Tighe et al 2010), this prohibits us from analyzing change in group size and composition – particularly unsatisfactory a solution when considering newer religious minorities such as Muslims or Hindus in the United States or Europe. To make matters worse, survey samples of such newer religious minorities are also less likely to be representative because these groups are harder to reach than majority populations (Font and Méndez 2013).

In this paper, we apply Bayesian multilevel regression modeling with post-stratification, or MRP (Gelman and Little 1997; Park et al 2004), to overcome these problems. In an earlier paper, Tighe et al (2010) pioneer the use of MRP for estimating religious demography. However, they focus only on the American Jewish population and only at a single point in time. More importantly, due to a lack of US census data on religion, they were not able to validate their estimates against a true benchmark. In other words, it remains unknown whether – and to what degree – MRP can help us measure religious demography.

The contribution of this paper is the provision of a dispositive test. To do so we estimate the size of three religious minorities – Muslims, Hindus, and Jews – in Great Britain[2] over a 20 year period. In addition to estimating the overall prevalence and dynamics of these religious identities, we also estimate their prevalence within age by gender by education subgroups. As such, we offer three contributions beyond the work of Tighe et al (2010). First, by focusing on three quite different groups, we provide a more general test of the method. In particular, While American (and European) Jewish populations are small, they are well-established groups that should pose no special sampling challenges. Second, we attempt to also estimate population dynamics, an especially challenging task for rapidly growing

[2] We focus only on England, Wales, and Scotland for reasons we outline in the methods section.

groups like British Muslims and Hindus. Finally, and most importantly, the UK government has included questions in religious identity in the 2001 and 2011 censuses, allowing us to determine the accuracy of our estimates, and the method of MRP.

Our results suggest that considerable optimism is warranted regarding the use of MRP applied to existing survey data for measuring religious demography. Our best performing models are able to very accurately measure the population proportions of British religious minorities: our estimates of overall group size differ from the "true" census values by only .37 (Muslim), .12 (Hindu), and .02 (Jewish) percentage points on average. Our estimates continue to be accurate, although somewhat less so, when we turn to the more arduous task of measuring the prevalence of the three religious identities within age by gender by education subgroups, with Muslim estimates differing, on average, from the true values by .62 percentage points; Hindu estimates, by .20 percentage points; and Jewish estimates, by .11 percentage points.

The findings of our paper also have important implications that go beyond the study of religious demography. Given the abundance of secondary survey data, it is in principle possible to study small socio-demographic groups and their dynamics in considerable detail, even in the absence of census data. We envision that this approach will be immensely valuable in the study of other social groups of interest, including but not limited to ethnic, political, or regional groups.

The remainder of this paper is structured as follows. We discuss existing approaches to measuring religious demography, the specific problems in using small-sample surveys to do so, and the potential that MRP holds in addressing these problems. We then describe our data and modeling approach before presenting and describing the results of our tests. We conclude with some advice to scholars interested in using this method to measure the size of religious or other demographic minorities.

## 2. Existing Approaches to Estimating Religious Demography

Researchers and policy-makers both require accurate measures of religious demography and diversity. As a consequence, a number of projects have arisen to provide such measures, including the World Religion Database (Johnson and Grim 2013), the World Religion Dataset (Maoz and Henderson 2013), and the Religious Characteristics of States data set (Brown and Patrick 2015).

Although the scope and comprehensiveness of these databases are admirable, and while they provide perhaps the only source of data for some regions and periods of time, there are nevertheless a number of limitations with their estimates. First, they incorporate data of varying quality. In some countries census or large-sample survey data are used, while smaller surveys are relied upon for others. Where neither of these exist, or they are out of date, subjective expert opinions inform the estimates (Johnson and Grim 2013). Although these databases rightly respect the adage that some data is preferable to none at all, we have no way of ascertaining the degree of uncertainty attached to any particular estimate because none are provided.[3] Without uncertainty estimates, analysts are led to treat census measures and expert opinions as equally valid.

Second, the methods used to adjust sample survey data, combine data, and obtain estimates when no data are available, are less than fully transparent. Adjusting, combining, interpolating, and extrapolating data require modeling. Yet neither the assumptions underlying the model nor the exact methods for doing so are fully specified. In addition, the uncertainty induced by modeling is again ignored.

Finally, these project only provide the marginal, not the joint demographic-religious distributions. Joint distributions are useful for at least three reasons. First, social scientific arguments on inter-group relations refer not only to uni-dimensional concepts based on mere

[3]Grim and Hsu (2011) develop a "data quality index" for religious demography data which rests on four criteria: geographic coverage, response rate, sampling, and questionnaire design but do not use this index to provide uncertainty estimates for any of these projects.

group size (e.g. fractionalization) but on multi-dimensional concepts, such as religion by ethnicity or socio-economic status (e.g. cross-cutting cleavages; Selway 2011). Second, for demographers to use estimates of religious group size in population projections, they require more detailed information on the sex-specific age-structure of these religious groups. Finally, reliable estimates of religious-demographic joint distributions can be used to adjust or post-stratify survey opinions from unrepresentative samples of religious minorities.

## 3. Measuring Religious Demography Using Sample Surveys

A potentially rich source of data for measuring religious demography is provided by the vast and rapidly growing cache of public opinion survey data that are collected for a variety of purposes, but happen to include information on respondents' religious identity. Indeed, questions on religious identity are now routinely included on cross-national survey projects, and such projects have fielded thousands of nationally representative surveys across dozens of countries over the last two or three decades.

There are, however, two difficulties in using these public opinion data to measure religious group prevalence. First, religious minorities are small groups – less than one percent of national populations in the case of European Jews. Standard public opinion survey samples are thus likely to include only tiny samples of religious minorities. Such small samples produce highly variable estimates. They also preclude further disaggregation by demographic subgroup.

One response to this challenge has been to pool survey samples gathered at different points in time, effectively creating a "mega-poll" with tens or hundreds of thousands of respondents (e.g. Smith 2005). This is a reasonable solution for estimating demographic size and composition of static groups such as American or European Jews (e.g. Tighe et al 2010). However, other minority groups of interest, such as European Muslims, are experiencing dramatic population growth. For such groups, pooling across time is not a desirable solution because it does not allow analysts to investigate the dynamics of the group's demography.

The second difficulty in using public opinion surveys to measure religious diversity is that these samples are not, in fact, truly random. Response rates are falling (Keeter et al 2006). Even well-funded, high quality national survey projects face this problem. For example, the UK sample of the European Social Survey saw response rates drop from 55.5 percent in 2002 to 43.6 percent in 2014, while the German sample saw a drop from 55.7 percent to 31.4 percent.[4]

To compound the problem, samples of religious minorities are particularly likely to be unrepresentative, because such individuals are more likely to have ethnic minority status and be of immigrant background, and less likely to speak the national language. As such, religious minorities constitute groups that are fundamentally more difficult to sample than majority populations (Font and Méndez 2013).

In sum, although decades of survey data exist on religious identity across virtually all European countries (and beyond), the samples of religious minorities are likely to be very small and additionally unrepresentative.

## 4. Improving Survey-Based Estimates Using Multilevel Regression with Poststratification

In response to the similar challenge of estimating population attributes within small areal units using survey data, statisticians developed methods of small area estimation (SAE). These methods uses regression models to "borrow strength" across time and areal units, often also incorporating supplementary census or administrative data (e.g. Rao 2013).

Of particular interest is the method of Gelman and Little (1997) and Park et al (2004), Multilevel Regression with Post-stratification, or MRP: a SAE model designed for estimating public opinion within areal units. MRP involves modeling some survey response using a set of geographic and demographic categories, such as areal unit, gender, and age. The analyst uses the model to predict the prevalence of that opinion within each cell of

[4] See `http://www.europeansocialsurvey.org/data/deviations_7.html`.

the joint demographic distribution (i.e. areal unit by gender by age), then weights each of these cells by its population size. Finally, the analyst then aggregates the weighted estimates to the areal unit of interest to obtain a modeled and post-stratified estimate. Although the multilevel component of this procedure can be estimated using classical methods such as maximum likelihood (e.g. Lax and Phillips 2009), Park et al (2004) propose the use of Bayesian methods.

The advantages of this Bayesian implementation of MRP are fivefold.[5] First, the use of a multilevel (or hierarchical) regression model allows estimates to be "partially pooled", or smoothed, across areal units and demographic groups to the extent that sample sizes in these subgroups are small (Gelman and Hill 2007). Partial pooling reduces the degree to which model predictions are affected by outlying observations and also provides for more efficient uncertainty estimates. It further allows for "deep" interactions among demographic-areal categories, which can lead to unreliable estimates in a classical (i.e., non multilevel) regression framework (Ghitza and Gelman 2013). Second, because the multilevel framework includes higher-level models for the areal unit of interest, area-level covariates can be added. These allow for more accurate estimates by incorporating additional non-survey information (Warshaw and Rodden 2012). Third, when combining data from several survey projects, MRP can adjust for the project-specific differences in methodology, such as sampling frames and interview modes. Although classical regression methods can include project fixed effects, it is not straightforward to then produce opinion estimates within small areas because the effect of the omitted project category becomes embedded within the model intercept. Fourth, the post-stratification step uses census or other official data to adjust for possible non-representativeness among small area subsamples. Finally, the use of Bayesian simulation methods of estimation provides more accurate, as well as more readily accessible, measures of inferential uncertainty (Gill 2008).

[5] There is now an extensive literature on this topic. See, e.g., Enns and Koch (2013); Hanretty, Lauderdale, and Vivyan (forthcoming); Lax and Phillips (2009); Park et al (2004); Selb and Munzert (2011); Warshaw and Rodden (2012); Leemann and Wasserfallen (forthcoming)

## 5. The Present Study

MRP has chiefly been used to measure public opinion within areal units. Tighe et al (2010), however, demonstrate the utility of the method for measuring religious demography. They apply MRP models to a pooled dataset of 50 surveys, finding that American Jews form 1.86 percent of the US population. Tighe et al. also go beyond measuring the single marginal distribution of religious identity – the overall size of the Jewish population – to estimating joint religious-demographic distributions. Although their estimates appear quite reasonable, Tighe et al. do not test whether they are accurate. With such a wealth of survey data, it could well be the case that any method – even a simple mean of the raw data – would be accurate.

We follow the example of Tighe et al. in using MRP for measuring the size and demographic composition of religious minorities. In contrast to Tighe et al., however, we seek to test the accuracy of the method by comparing our estimates with official data on religious group size and demography extracted in the 2001 and 2011 UK censuses. This will not only allow us to determine the degree to which our estimates are accurate, but will also provide some guidance to other scholars interested in measuring religious demography using using existing survey data.

Our paper differs from that of Tighe et al. in two additional ways: first, we aim to measure the size of several religious minorities, not just one: Muslims and Hindus, as well as Jews. Second, we also attempt to measure the dynamics of these three groups's demography over 20 years. While British (and American) Jewish populations are small, they are long-established and demographically stable groups.[6] Muslim and Hindu Britons, although slightly larger, are rapidly growing groups. Muslims, for example, increased from 2.8 to 4.8 percent of the UK population between 2001 and 2011, with Hindus showing an increase from 1.0 to 1.5 percent. In addition, survey data on Muslim Britons, in particular, is

[6]According to the 2011 census, Jews formed 0.5 percent of the UK population.

likely to suffer from poor sampling given that 92 percent of this group report ethnic minority status in 2011, 53 percent report being born outside the UK, and 48 percent are younger than 25 years of age. Thus, estimating the religious demography of these three groups over 20 years epitomizes the difficulties we have highlighted in measuring religious demography using survey data, and thus provide a stern, and hopefully generalizable test, regarding the utility of our method.

Because the multilevel regression models at the heart of MRP can be specified in a number of different ways, we estimate and test six Bayesian hierarchical models of Muslim, Jewish, and Hindu prevalence, thus providing six sets of estimates for each group. Models all include a time component, but vary in the number and complexity of the demographic predictors. In addition to the simple additive model, we will also include a model with all two-way demographic interactions, as well as a model where the time component varies by demographic factor. Each of these three models will be tested both with and without ethnicity, which might be expected to increase the accuracy of estimates but comes at the expense of some loss of survey data.

Because the method of MRP is both computationally challenging and requires a fair amount of data manipulation, we additionally model the data using a set of six simple logit regressions. Each of these models corresponds as closely as possible to one of the six hierarchical models, with the exception that demographic factors are estimated using classical, non-multilevel methods, and the post-stratification weighting is done using a vector of survey weights – as one might find in an off-the-shelf survey dataset. These classical regressions with survey weights, which we refer to as CRSW estimates, offer a quick and convenient alternative for for measuring demographics using survey data.

Finally, we also estimate religious group demographics by finding the raw proportion of the population holding each of the religious identities. We refer to these as “disaggregated” estimates.

# 6. Data and Methods

## 6.1. Data

Our ambition is to test a method that might be used in other national settings. We thus include data only from major cross-national survey projects: the International Social Survey Program (ISSP), and its British component, the British Social Attitudes Survey (BSAS); the Eurobarometer (EB); and the European Social Survey (ESS).[7] We include all BSAS, ESS, and EB surveys that ask respondents their religious identity, and that were conducted between 1995 and 2014, both inclusive. Within these constraints, we obtain 20 survey samples from the BSAS, 15 from the EB and three from the ESS, with 91,862 respondents in total. These are summarized in Table 1 below. To validate our estimates, we use the publicly available five percent census samples for the 2001 and 2011 censuses.[8]

It is important that these sources of data have comparable religious identity questions and response sets. Yet as Table 2 indicates, although the three questions used in the different survey projects are virtually identical, the census question differs slightly. There is also some variation in response sets: while the census only includes one response for Christian religions, the survey projects all allow respondents to choose Christian denominations.

The UK Office of National Statistics (ONS) has tested whether these differences produce different results (ONS 2009). They found significantly different percentages for the "Christian", "Sikh", and "No religion" responses when comparing the census and BSAS questions and response sets. Fortunately, these are not the groups we are fundamentally interested in. Nevertheless, to avoid bias, we do not attempt to estimate the size and demographics of all religious groups, but, rather, we treat Muslim, Jewish, and Hindu as separate categories and collapse all other responses, including refusals and "don't knows," into an "Other" category.[9] We treat non-response in both census and survey data as a separate

[7]In order to not limit the generalizability of our results, we only use data from cross-national projects, and do not use UK specific data sources, such as the UK Labour Force Survey.

[8]Census 2001: Small Area Microdata; Census 2011: Microdata Individual Safeguarded Sample.

[9]The census question is optional, so non-response does occur. This actually makes the census data more

response category rather than removing these observations.

In addition, the public release of the Northern Ireland census data does not include detailed breakdowns of religious minorities. We therefore exclude Northern Ireland from all the census and survey data we collect. We also remove people younger than 16 from the census results because such respondents are excluded, by design, from the surveys we examine.[10] Our data and estimates are thus representative of the population aged 16 and older and living in mainland Britain (England, Wales, and Scotland).

Finally, we also extract data on gender, age, household size, level of education, and ethnicity from the surveys and census data. These will be used as categorical predictors in our models and to post-stratify the resulting estimates. Non-response and "don't know" responses were coded as missing values, and all respondents with at least one missing value for these variables (5,162; 5.6 percent) were removed from the dataset, leaving a final sample size of 86,664 respondents.[11]

### 6.2. Model Specification

MRP entails predicting and post-stratifying the survey response of interest within joint demographic categories. The first step in specifying our model is to select demographic variables that are available in both our survey and census datasets. We select these variables according to four criteria. First, we examine the degree to which the survey data are unbalanced, compared to the census data, on each of the demographic factors. Second, we examine the bivariate associations between demographic variables and each religious identity. Third, we consider the extent of missing data. Finally, since our overarching goal is to produce a model that we can use in settings where official data on religion do not exist, we favor as simple, and thus generalizable, a model as possible. Using these criteria, we arrive at a basic set of four

comparable to the survey data, where, of course, all questions are optional.

[10] While most of these surveys sample those 18 years of age or older, there are some 16 and 17 year olds in our survey dataset. Our choice of age categorization is also constrained by the fact that the publicly available census data is released with age already categorized (for example, 16-19).

[11] See the supplementary materials for further details on the type of coding and degree of missingness for each variable.

demographic variables that will be used in the regressions and the post-stratification step: gender (male and female); age (16-29, 30-49, 50-64, and 65 and older); education (degree or no degree); and household size (1, 2-4, and 5 or more adults).[12]

To address the changes in religious demography, we extend the MRP method by additionally modeling population dynamics (e.g. Pacheco 2011). Given the deterministic nature of population dynamics, we find that a simple linear trend model effectively captures the dynamics of our religious group populations.

We are now in a position to describe our basic model more formally. We model the percent of the survey sample, $y_i$, where $i \in N$, holding each religious identity, in each year, $t$, as a independent Bernoulli variable. Our most basic version of the Muslim model is as follows (where we use the notation of Gelman and Hill (2007) in referring to individual survey respondents, $i$, being nested within $g$ gender groups or $r$ age groups, for example):

$$\Pr(y_{it} = \text{Muslim}) = \text{logit}^{-1}\left(\alpha + \beta \text{t} + \mu^{gen}_{g[i]} + \mu^{age}_{r[i]} + \mu^{hhs}_{h[i]} + \mu^{edu}_{d[i]} + \mu^{proj}_{p[i]}\right),$$

Population dynamics are specified using a linear trend, captured with coefficient $\beta$.[13] The grouping variables of gender, age, household size, education, and survey project are then modeled as random effects drawn from normal distributions with variances, $\sigma^2$ to be estimated from the data:

$$\begin{aligned}
\mu^{gen}_{g[i]} &\sim N(0, \sigma^2_{gen}),\\
\mu^{age}_{r[i]} &\sim N(0, \sigma^2_{age}),\\
\mu^{hhs}_{h[i]} &\sim N(0, \sigma^2_{hhs}),\\
\mu^{edu}_{d[i]} &\sim N(0, \sigma^2_{edu}),\\
\mu^{proj}_{p[i]} &\sim N(0, \sigma^2_{proj}).
\end{aligned}$$

[12] Full details are provided in the Appendix.
[13] The 20 year time period is standardized to range from 0 to 1.

Survey project is included as a grouping variable but we make no use of its parameter estimates when estimating predicted effects. Rather, we allow our model to partial out the particular effects of each survey project, leaving us with survey project-adjusted estimates.[14]

We can easily add further complexities to this basic multilevel model. We are particularly interested in three additional features: the inclusion of ethnicity, two-way interactions among all demographic factors, and varying dynamic effects by demographic factors.

*Ethnicity.* We include an additional demographic category, ethnicity (white and non-white), in half of our models. Ethnicity is not available for 17 percent of the observations in our survey dataset – mainly because this variable is not included on the Eurobarometer questionnaire. However, the strong relationship we observe between ethnicity and Hindu and Muslim identity[15] suggests that is worth considering whether including ethnicity increases predictive power enough to offset the negative effects of data loss.[16] Model 2 extends Model 1 by including varying intercepts for ethnicity: $\mu^{ethnic}_{n[i]}$.

*Demographic interactions.* We specify two models where we incorporate all two-way interactions between demographic categories, one with ethnicity included, and another, without. These models allow the relationships between demographic predictors and religious identity to depend on other demographic predictors. While such interactions are possible within a classical regression framework, they may result in unreliable parameter estimates because of increasingly sparse survey data within the joint demographic distributions (Gelman 2007). Multilevel models, in contrast, partially pool information across all categories of each predictor, allowing “deep interactions” (Ghitza and Gelman 2013). Thus model 3

[14] We utilize the fact that the set of project intercepts is modeled with mean of 0. By leaving out the project effects when using our model to predict, we in effect partial out the effects of survey project on religious minority identity.

[15] See the supplementary materials for evidence.

[16] Although the model with ethnicity has to make do with less data, we consider these model comparisons to be “fair” in the sense that they represents the actual trade-off researchers face when using MRP to estimate religious demography.

extends model 1:

$$\begin{aligned}\Pr(y_{it} = \text{Muslim}) = \text{logit}^{-1}\Big(&\alpha + \beta \text{t} + \mu^{gen}_{g[i]} + \mu^{age}_{r[i]} + \mu^{hhs}_{h[i]} + \mu^{edu}_{d[i]} + \\ &\mu^{gen.age}_{g\cdot r[i]} + \mu^{gen.hhs}_{g\cdot h[i]} + \mu^{gen.edu}_{g\cdot d[i]} + \mu^{age.hhs}_{r\cdot h[i]} + \\ &\mu^{age.edu}_{r\cdot d[i]} + \mu^{hhs.edu}_{h\cdot d[i]} + \mu^{proj}_{p[i]}\Big)\end{aligned}$$

*Demographically varying trends.* Finally, we also allow our linear dynamic effects to vary across demographic groups. Again, we specify one such model with the basic four demographic categories and another with ethnicity added. These models allow differential rates of religious minority population growth within demographic subgroups. This helps to model situations such as higher rates of minority immigration among men or young people. Such a feature is again possible within a classical regression framework – in the form of time by demographic category interactions – but the risk is again that parameter estimates are too extreme. Model 5 extends model 1 through the addition of additional slope parameters ($\gamma$) for the time trend that vary by demographic group:

$$\begin{aligned}\Pr(y_{it} = \text{Muslim}) = \text{logit}^{-1}\Big(&\alpha + (\beta + \gamma^{gen}_{g[i]} + \gamma^{age}_{r[i]} + \gamma^{hhs}_{h[i]} + \gamma^{edu}_{d[i]})\text{t} + \\ &\mu^{gen}_{g[i]} + \mu^{age}_{r[i]} + \mu^{hhs}_{h[i]} + \mu^{edu}_{d[i]} + \mu^{proj}_{p[i]}\Big)\end{aligned}$$

The intercepts ($\mu$) and trend slopes ($\gamma$) for each demographic category are modeled using a bivariate normal distribution, with intercepts and slopes correlated ($\rho$). For example, the gender level intercepts and slopes are modeled as follows:

$$\begin{pmatrix} \mu^{gen}_{g[i]} \\ \gamma^{gen}_{g[i]} \end{pmatrix} \sim N(0, \Sigma_{gen}), \quad \Sigma_{gen} = \begin{pmatrix} \sigma^2_{\mu^{gen}} & \\ \rho^{gen}\sigma^2_{\mu^{gen}}\sigma^2_{\gamma^{gen}} & \sigma^2_{\gamma^{gen}} \end{pmatrix}$$

In addition to these six multilevel regressions, we also include a corresponding set of six classical logit regressions (see Table 3). Rather than post-stratifying the estimates from the classical logit models, as we do with our multilevel models (see next section), we instead use survey weights to, in effect, find the weighted likelihoods. We calculate the vectors of weights in question ourselves, using our population data on age by gender by education by household size by (as appropriate) ethnic groups. Our goal in using classical logit models with survey weights is to test a computationally simple alternative to Bayesian MRP for analysts who might prefer the convenience of doing so.

### 6.3. Post-stratification

We post-stratify the estimates from our six multilevel models. To do so, we first obtain predictions from each model of the proportion of adults in Great Britain holding each of the three religious identities within demographic subgroups, $j$. For example, we denote model MRP1's estimates as $\phi_j^{MRP1}$. For the models without ethnicity, there are 48 demographic subgroups (2 age by 2 gender by 4 education by 3 household size) for each of the 20 years, and thus 960 predictions in total. When ethnicity is included in the model there are 96 demographic subgroups for each of the 20 years, and thus 1,920 total predictions. We next weigh each subgroup prediction by the proportion of the adult population in Great Britain each year that falls in that subgroup, $N_j$.[17] Finally, we obtain weighted estimates for a smaller set of 32 (2 age by 2 gender by 4 education by 2 years) demographic subgroups, which we denote $k$, by aggregating the set of $j = 48$ subgroups across the three household size groups.[18]

Using our first model, MRP1, the post-stratified estimate of the proportion of the population, $\pi_k^{MRP1}$, holding a particular religious identity within each of the 32 target demographic subgroup, $k$:

$$\pi_k^{MRP1} = \frac{\sum_{j \in k} N_j \phi_j^{MRP1}}{\sum_{j \in k} N_j}$$

For comparison, we also produce "disaggregated" estimates: the raw proportion of the

[17]Our data for these population proportions come from the 2001 and 2011 censuses. We use linear interpolation to smooth the population estimates for each of the 48 subgroups between the census years of 2001 and 2011. Before and after these years, we use the unadjusted estimates from either the 2001 or 2011 census.

[18]When ethnicity is included we aggregate across household sizes and ethnicity to produce our 32 estimates.

population, $\pi_j^{Disagg}$, holding a particular religious identity within each of our 32 demographic subgroup, $k$:

$$\pi_k^{Disagg} = \frac{\sum_{i \in k} y_i}{N_k}$$

### 6.4. Estimation

We use Bayesian Markov Chain Monte Carlo (MCMC) methods to estimate the multilevel models. Bayesian MCMC allows for more accurate estimation of hierarchical variance parameters (e.g. $\sigma^2_{gen}$) than maximum likelihood estimation (Gill 2008). It also allows for estimation uncertainty to be measured in a straightforward fashion.

Weakly informative $N(0, 2.5)$ priors are used for the dynamic effects ($\beta$) and $N(0, 10)$ priors for the grand intercept ($\alpha$). For the models with varying dynamic effects, the hierarchical covariance matrix is decomposed into variances and a correlation matrix, and the latter given an LKJ prior (Lewandowski et al 2009; Stan Development Team 2016). The hierarchical variance parameters are then given weakly-informative $Gamma(1, 1)$ priors.

Estimation is conducted using the `Stan` program, which implements Hamiltonian Monte Carlo sampling (Stan Development Team 2016). `Stan` was called using the `stan_lmer` function from the `RStanARM` library in the `R` statistical environment. Four parallel chains were run for 500 iterations each, with the first half of the samples in each chain treated as warm ups and discarded, and the remaining 1,000 samples saved and analyzed further. This number of iterations proved to be more than sufficient for convergence, with the Gelman-Rubin diagnostic (Gelman and Rubin 1992) reaching a value of between 0.95 and 1.05 for all parameters.

The classical logit regression models are estimated using the `svyglm` function from the `survey` library in `R`, which fits the classical logit model using inverse-probability weights and design-based standard errors. Two vectors of survey weights are used: the first uses the joint population age-gender-education-household size distributions; the second adds ethnicity. As such, the estimates obtained from the classical logit models are weighed *before* the subgroup predictions are made.

For each of our 12 models and three religious groups, the heart of our analysis is a comparison between the accuracy of the estimates of religious group prevalence within the 32 demographic subgroups with the corresponding census estimates.

## 7. Results

We test the fit of our 12 models by comparing their post-stratified estimates (e.g. $\pi_k^{MRP1}$) to the "true" values as reported in the 2001 and 2011 censuses, a test of out-of-sample predictive error. In particular, we compare the estimated and "true" proportions of British adults holding each of the three religious identities, in 2001 and 2011, within gender by age by household size groups. Using several metrics, we select a best-fitting model for each of the three religious group. We then examine the estimates of these best-fitting models in more detail.

### 7.1. Testing the accuracy of the estimates

We now compare our estimates of the religious identity within socio-demographic subgroups to the "true" values reported in the 2001 and 2011 censuses. In particular, we use our 12 models to predict the size of each of the three religious identities within 32 age by gender by education by census year subgroups. Such estimates of religious-demographic size are interesting in their own right, but also provide us with a more rigorous test of out-of-sample error, afforded by having 32 estimate-true values comparisons for each religious group and model.

We use two metrics to evaluate the out-of-sample predictive error of our models. First, we calculate the average percent error. For example, for our first set of estimates, from model MRP1:

$$APE^{MRP1} = \frac{\sum_{k=1} \frac{\left|\pi_k^{MRP1} - \pi_k^{Census}\right|}{\pi_k^{Census}}}{k}$$

For each of the $k = 32$ estimates, we find the absolute value of the error compared with the census value, divide by the census value to convert to the percentage scale, and then take the mean. Second, we calculate Pearson's correlation coefficient between the census and estimated proportions. These two metrics are presented for the 12 models and three religious groups in Figure 1.

A comparison of the results shows that the better performing models – those near the top of each plot – achieve roughly comparable accuracy across the three religious groups: a rate of error of around 20 percent, and correlations with the corresponding census values of between .92 and .98. These deviations are modest for the estimation of such small subgroups.[19] The observed rates of average percent error correspond to average deviations from the census values (in percentage points) of .45 for the estimates of Muslim identity within the 32 subgroups, .20 for the Hindu estimates, and .11 for Jewish estimates.

Both the MRP and CRSW estimates substantially improve on the estimates obtained by disaggregating the pooled survey data. As the figures clearly show, even the worst performing model offers a substantial improvement over the disaggregated results. The modeled results are up to 62 percent more accurate for Muslims, 77 percent for Hindus, and 82 percent for Jews.

There is, however, considerable variation in the performance of particular models. The two best estimates for each religious group (focusing on average percent error) are those derived using MRP, rather than CRSW. Yet, the worst-performing estimates for each religious group are also those based on MRP. MRP holds promise but perhaps peril as well. Model choice nevertheless clearly matters for improving survey estimates of religious demography.

Which modeling strategy is most accurate? It depends on the religious group in question. For the newer religious minorities – Hindus and Muslims – the three MRP estimates including ethnicity are the three most accurate. Despite the loss of data that follows from incorporating ethnicity in models of religious identity, doing so clearly helps improve the

[19]For example, only 0.26 percent of men aged 30-49 without a degree reported a Jewish identity in 2001.

accuracy of estimates of the demography of newer religious minorities.

There is little to choose among the three estimates that do incorporate ethnicity. However, the MRP4 estimates, which include ethnicity and all two-way interactions, are perhaps best as they rank either first or second according to the average percent error and correlation tests for both Muslims and Hindus.

Figures 2 and 3 then display the estimated and census prevalence of Muslim and Hindu identity within each of the 32 demographic subgroups. We have already confirmed, using the average percent error and correlations metrics, that these estimates are generally quite accurate. These figures provide a visual confirmation of this accuracy. In particular, the Muslim estimates are close to the census values, with the partial exception of the young (those aged 16-29) in 2011. As already noted, our Hindu estimates are even more accurate than the Muslim ones, but perhaps underestimate Hindu prevalence among those without a degree in 2001.

The best model for measuring the demography of British Jews differs the best model in the Muslim and Hindu cases. Here, the MRP3 estimates, which do not include ethnicity, but do include two-way interactions, are the most accurate and the most strongly correlated with the census values. The other models that exclude ethnicity also perform well for estimating Jewish demography, including the simplest model of all, a classical logit model with four categorical predictors, a linear time effect, and survey weights.

Figures 4 then plots the MRP3 estimates of Jewish identity along with the census values. The figure again confirms the extraordinary accuracy of this model. With the interesting exception of women with a degree, aged 65 and over, in 2001, our estimates hew closely to the census values.

It is also worth considering the overall performance of the MRP models relative to the CRSW models. The latter are considerably easier to fit and use than the method of MRP, which requires both computationally expensive Bayesian MCMC methods and manual processing of model parameter estimates in the post-stratification step. However, the best

CRSW estimates are somewhat less accurate than the best MRP estimates. For Muslims, the CRSW1 estimates are 15.5 percent less accurate than the MRP4 estimates that we prefer; for Hindus, the CRSW3 estimates are 25.5 percent less accurate; while for Jews, the CRSW1 estimates are 11 percent less accurate than the MRP 3 results. Thus, the time and effort required to implement the MRP method does appear to pay off in a noticeable improvement in accuracy.

In sum, we find that applying regression models and post-stratification to survey data can provide accurate estimates of small religious-demographic subgroups. The method of MRP, although more onerous than using classical regressions with survey weights, produces estimates that are 10-25 percent more accurate. We also find that MRP models with two-way demographic interactions appear a good general choice for modeling religious demography. Varying the effects of time by demographic category produces uneven results and is probably best avoided. Finally, for newer religious minorities, such as Hindus and Muslims – but not more settled minorities, such as Jews – including ethnicity can improve estimates despite some ensuing loss of data.

### 7.2. Examining religious minority population dynamics

Although our empirical strategy called for estimating the size of three religious groups within 32 demographic subgroups, analysts and practitioners might perhaps be more interested in our estimates of the overall size of these religious groups over time. Figure 5 thus plots our estimates of the prevalence of the three religious identities over the 20 years for which we have data. For each plot, we use the results that we have designated our most-preferred: the MRP estimates with two-way interactions, and – for Muslims and Hindus – ethnicity included.

Figure 5 contrasts the MRP estimates, in orange, with the disaggregated estimates, in grey. Both sets of estimates are shown with 80% uncertainty intervals. The actual census data for 2001 and 2011 are then presented using red dots. At the level of overall population size, this figure shows that the method of MRP applied to existing survey data produces very accurate estimates of Jewish population size. Our 2001 and 2011 estimates deviate from the census estimates by a tiny 0.02 percentage points, or 4.5 percent when we adjust for the baseline prevalence. Jewish demographics are perhaps fairly easy to estimate, as this group is long-established in Britain. Yet our estimates of Hindu and Muslim population size remain very accurate, with average absolute errors of 0.12 and 0.37 percentage points respectively, which translates into percentage error rates of 11.5 and 10.4 percent.

These results show that our method is even more accurate when it comes to measuring overall prevalence of religious minority identities, than for measuring the prevalence of these identities within demographic subgroups. Indeed, results as accurate as these confirm that MRP can indeed be used, together with existing survey data, to validly measure the size of religious minorities.

## 8. Conclusion

This paper proposes and tests a method for estimating the demographic size, composition, and dynamics of religious minority groups. The method is MRP, the application of multilevel regression modeling to survey data, with post stratification of the estimates. We compare the accuracy of six MRP models to two simpler methods for estimating demographics from survey data: first, a corresponding set of six classical logit regressions with simple survey weighting; second, the raw, disaggregated proportions within the pooled survey data.

We focus on three religious minority groups (Muslim, Jewish, and Hindu) in Great Britain. Although all three are minorities, these groups differ in important ways. Jewish Britons are an older and more stable group, while Muslims and Hindus are younger, more rapidly growing, and much more likely to be immigrants. These latter groups thus provide a particularly stern test for estimating group size using survey data.

The choice of the UK allows us to test the accuracy of our estimates using official data on religious group size and demography, which was collected in the 2001 and 2011 censuses.

Because the UK is included in many cross-national survey projects, such as the GSS/ESS and ISSP, this case also potentially permits our results to be of use for other scholars interested in estimating the size of religious or, indeed, other demographic groups.

With only two "true" data points for each group – the 2001 and 2011 censuses – we focus instead on estimating the prevalence of each religious group within 32 demographic subgroups formed by the intersection of age, sex, education and census year categories. Our analysis then focuses on measuring the out-of-sample prediction error across these 32 demographic subgroups, for each of the three religious groups and each of the 12 models.

We find, firstly, that survey data can indeed be used to accurately measure the size of small minority groups, and even the joint distributions of these minorities within other demographic subgroups, such as men aged 30-49 without degree in a particular year. We have over 90 thousand survey respondents in our pooled dataset, but similar rich troves of survey data are likely to exist for other North American and European countries and, indeed, for minority groups other than religious ones.

In addition, we find a similar accuracy (in average percentage error terms) when measuring the prevalence of established minorities, British Jews in our case, and newer, more rapidly growing minorities (British Muslims and Hindus), whose survey samples may be more suspect. We thus conclude that even rapidly growing groups can be reliably estimated using a model with a simple linear time trend.

We find that MRP models with two-way demographic interactions appear to be a good general choice for modeling religious demography. Varying the effects of time by demographic category produces uneven results and is probably best avoided. Finally, for newer religious minorities, such as Hindus and Muslims – but not more settled minorities, such as Jews – including ethnicity can improve estimates despite some ensuing loss of data.

We find that the method of MRP – Bayesian multilevel regression models to predict group prevalence in the joint distribution of these demographic categories, coupled with post-stratification of the cells using population weights allows for the most accurate estimates.

However, a simple classical logit regression with a vector of survey weights is also a reasonable option, especially if one if only interested in the marginal distributions of overall group size.

In future research, scholars might use these methods to estimate the size of religious minorities in settings other than the UK, especially those countries where official religious data are not collected. Another potential application would be to measure the size of racial or ethnic minorities in cases where survey data exist but census data is lacking. In either scenario, our results show that analysts could estimate both the marginal distributions, or overall size, of these groups or their distributions joint with other demographic factors.

## References

Adida, Claire L., David D. Laitin and Marie-Anne Valfort. 2016. *Why Muslim Integration Fails in Christian-Heritage Societies.* Harvard University Press.

Brown, Davis and James, Patrick. 2015. Religious Characteristics of States Data Set. Maryville University.

Buttice, Matthew K. and Highton, Benjamin. 2013. "How Does Multilevel Regression and Poststratification Perform with Conventional National Surveys?" *Political Analysis* 21: 449-467.

DiPrete, Thomas, Andrew Gelman, Tyler McCormick, Julien Teitler, and Tian Zheng. 2011. "Seggregation in Social Networks Based on Acquaintance and Trust." *American Journal of Sociology* 116(4): 1234–1283.

Enns, Peter K. and Julianna Koch. 2013. "Public Opinion in the U.S. States: 1956 to 2010." *State Politics and Policy Quarterly.* 13(3): 349–372.

Fearon, James, and David Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97(1): 228-90.

Font, Joan, and Mónice Méndez. (2013). "Introduction: The methodological challenges of surveying populations of immigrant origin." In *Surveying Ethnic Minorities and Immigrant Populations: Methodological Challenges and Research Strategies*, Joan Font and Mónice Méndez, eds. Amsterdam: Amsterdam University Press. pp. 11–30.

Geddes, Andrew. 2003. *The Politics of Migration and Immigration in Europe.* London: Sage.

Gelman, Andrew. 2007. "Struggles with Survey Weighting and Regression Modeling" *Statistical Science* 22(5):153–164.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel / Hierarchical Models.* New York: Cambridge University Press.

Gelman, Andrew and Thomas C. Little. 1997. "Poststratification Into Many Categories using Hierarchical Logistic Regression." *Survey Methodology* 23:12735.

Gelman, Andrew, David Park, Boris Shor, and Jeronimo Cortina. 2008. *Red State, Blue State, Rich State, Poor State. Why Americans Vote the Way They Do.* Princeton: Princeton University Press.

Gelman, Andrew and Donald B. Rubin. 1992. "Inference From Iterative Simulation Using Multiple Sequences." *Statistical Science* 7(4):457–72

Ghitza, Yair and Andrew Gelman. 2013. "Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups." *American Journal of Political Science* 57(3): 762-776.

Grim, Brian J and Becky Hsu. 2011. "Estimating the Global Muslim Population: Size and Distribution of the World's Muslim Population." *Interdisciplinary Journal of Research on Religion* 7: n/a.

Gill, Jeff. 2008. *Bayesian Methods. A Social and Behavioral Sciences Approach.* Chapman & Hall.

Hanretty, Chris, Benjamin E. Lauderdale, and Nick Vivyan. Forthcoming. "Comparing Strategies for Estimating Constituency Opinion from National Survey Samples." *Political Science Research and Methods* `doi:10.1017/psrm.2015.79`

Johnson, Todd M. and Brian J. Grim. 2013. *The World's Religions in Figures. An Introduction to International Religious Demography.* Chichester: Wiley-Blackwell.

Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best, and Peyton Craighill.

2006. "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70(5): 759–779.

Lax, Jeffrey R. and Justin H. Phillips. 2009. "How Should We Estimate Public Opinion in the States? *American Journal of Political Science* 53:107-21.

Leemann, Lucas and Fabio Wasserfallen. Forthcoming. "Extending the Use and Prediction Precision of Subnational Public Opinion Estimation." *American Journal of Political Science*

Levitt, Peggy. 2003. " 'You Know, Abraham Was Really the First Immigrant:' Religion and Transnational Migration." *International Migration Review.* 37(3): 847–873.

Lewandowski Daniel, Dorota Kurowicka, and Harry Joe. 2009. "Generating Random Correlation Matrices Based on Vines and Extended Onion Method." *Journal of Multivariate Analysis* 100(9): 1989–2001.

Lipset, Seymour M. and Stein Rokkan. 1967. *Party Systems and Voter Alignments. Cross-National Perspectives.* New York: Collier-Macmillan.

Maoz, Zeev and Erol A. Henderson. 2013. "The World Religion Dataset, 1945-2010: Logic, Estimates, and Trends." *International Interactions* 39(3): 265-291.

Office for National Statistics. 2009. "Final Recommended Questions For the 2011 Census in England and Wales: Religion." London: Office for National Statistics. Retrieved 28 November, 2016 (`http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/2011-census-questionnaire-content/final-recommended-questions-2011---religion.pdf`.

Pacheco, Julianna. 2011. "Using National Surveys to Measure Dynamic U.S. State Public Opinion: A Guideline for Scholars and an Application." *State Politics and Policy Quarterly* 11: 415-39.

Park, David K., Andrew Gelman and Joseph Bafumi. 2004. “Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls.” *Political Analysis* 12: 375–385.

Putnam, Robert D. and David E. Campbell. 2010. *American Grace: How Religion Divides and Unites Us.* New York: Simon and Schuster.

Rao, J. N. K. 2003. *Small Area Estimation*. Hoboken, NJ: Wiley

Selb, Peter and Simon Munzert. 2011. “Estimating Constituency Preferences From Sparse Survey Data using Auxiliary Geographic Information. *Political Analysis* 19: 455–70.

Selway, Joel Sawat. (2011). “Cross-Cuttingness, Cleavage Structures and Civil War Onset.” *British Journal of Political Science* 41(1): 111–138.

Smith, Thomas W. (2005). *Jewish Distinctiveness in America*. New York: The American Jewish Committee.

Stan Development Team (2016) *Stan Modeling Language: User's Guide and Reference Manual: Stan 2.10.0.* Stan Development Team

Tighe, Elizabeth, David Livert, Melissa Barnett, and Leonard Saxe. 2010. “Cross-Survey Analysis to Estimate Low-Incidence Religious Groups.” *Sociological Methods & Research* 39(1): 56–82.

Toft, Monica, Daniel Philpott, and Timothy S. Shah. 2011. *God's Century: Resurgent Religion and Global Politics*. New York: W.W. Norton.

Toshkov, Dimiter. 2015. “Exploring the Performance of Multilevel Modeling and Poststratification with Eurobarometer Data.” *Political Analysis* 23(3): 455–460.

Warshaw, Christopher and Jonathan Rodden. 2012. “How Should We Measure District-Level Public Opinion on Individual Issues?” *Journal of Politics* 74(1): 203–219.

Voas, David, Alasdair Crockett and David V.A. Olson. 2002. "Religious Pluralism and Participation: Why Previous Research Is Wrong." *American Sociological Review* 67(2): 212–230.

Voas, David and Mark Chaves. 2016. "Is the United States a Counterexample to the Secularization Thesis?" *American Journal of Sociology* 121(5): 1517–1556.

**Table 1.** Survey Data by Year and Project

| Year | British Social Attitudes Survey | Eurobarometer | European Social Survey |
|---|---|---|---|
| 1995 | 3,633 | 2,154 | 0 |
| 1996 | 3,662 | 0 | 0 |
| 1997 | 1,355 | 2,183 | 0 |
| 1998 | 3,146 | 1,066 | 0 |
| 1999 | 3,143 | 0 | 0 |
| 2000 | 3,426 | 0 | 0 |
| 2001 | 3,287 | 0 | 0 |
| 2002 | 3,435 | 0 | 0 |
| 2003 | 4,432 | 0 | 0 |
| 2004 | 3,199 | 0 | 0 |
| 2005 | 4,268 | 3,063 | 0 |
| 2006 | 4,290 | 3,021 | 0 |
| 2007 | 4,124 | 0 | 0 |
| 2008 | 4,486 | 1,005 | 2,352 |
| 2009 | 3,421 | 1,015 | 0 |
| 2010 | 3,297 | 1,009 | 2,422 |
| 2011 | 3,311 | 0 | 0 |
| 2012 | 3,248 | 1,001 | 2,286 |
| 2013 | 3,244 | 0 | 0 |
| 2014 | 2,878 | 0 | 0 |

Cell entries are the number of respondents that were asked a survey question regarding their religious identity by year and survey project. In some years more than one Eurobarometer survey asked respondents about religious identity. Total $N$ = 91,862.

**Table 2.** Question Wording and Response Sets

**Census**[a]

*Question*: What is your religion?

*Response set*: 1) No religion; 2) Christian (Including Church of England, Catholic, Protestant and all other Christian denominations); 3) Buddhist; 4) Hindu; 5) Jewish; 6) Muslim; 7) Sikh; and 8) Any other religion or belief (WRITE IN).

**European Social Survey**

*Question*: Do you consider yourself as belonging to any particular religion or denomination? (IF YES) Which one?

*Response set*: 1) Yes, Roman Catholic; 2) Yes, Protestant; 3) Yes, Eastern Orthodox; 4) Yes, Other Christian denomination; 4) Yes, Jewish; 5) Yes, Islamic; 6) Yes, Eastern religions; 7) Yes, Other non-Christian religions; 8) No; 9) Don't Know

**Eurobarometer**

*Question*: Do you regard yourself as belonging to a religion? (IF YES) Which of them?

*Response set*: 1) Yes, Roman Catholic; 2) Yes, Protestant; 3) Yes, Orthodox; 4) Yes, Other Christian; 4) Yes, Jewish; 5) Yes, Muslim; 6) Yes, Buddhist; 7) Yes, Sikh; 8) Yes, Hindu; 9) Yes, Atheist; 10) Yes, Non-believer, agnostic; 11) Yes, Other (WRITE IN); 12) None

**British Social Attitudes Study**

*Question*: Do you regard yourself as belonging to any particular religion? (IF YES) Which?

*Response set*: 1) No religion; 2) Yes, Christian, no denomination; 3) Yes, Roman Catholic; 4) Yes, Church of England/Anglican; 5) Yes, Baptist; 6) Yes, Methodist; 7) Yes, Presbyterian/Church of Scotland; 8) Yes, Free Presbyterian; 9) Brethren; 10) United Reform Church (URC)/Congregational; 11) Other Protestant (WRITE IN); 12) Other Christian (WRITE IN); 13) Yes, Hindu; 14) Yes, Jewish; 15) Yes, Islam/Muslim; 16) Yes, Sikh; 17) Yes, Buddhist; 18) Yes, Other non-Christian (WRITE IN)

[a] England and Wales Census

**Table 3.** Models

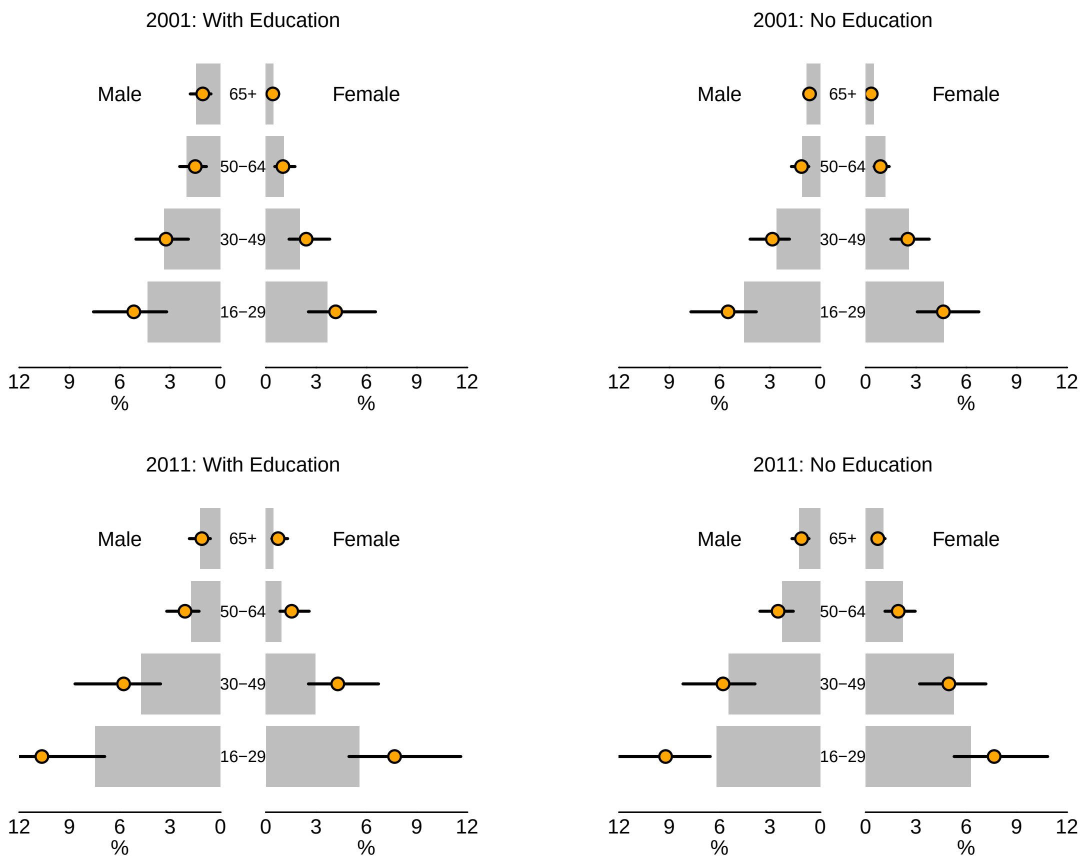
| Name | Ethnicity | Demographic interactions | Trend varies |
|---|---|---|---|
| MRP 1 | – | – | – |
| MRP 2 | Included | – | – |
| MRP 3 | – | All 2-way | – |
| MRP 4 | Included | All 2-way | – |
| MRP 5 | – | – | by demographics |
| MRP 6 | Included | – | by demographics |
| CRSW 1 | – | – | – |
| CRSW 2 | Included | – | – |
| CRSW 3 | – | All 2-way | – |
| CRSW 4 | Included | All 2-way | – |
| CRSW 5 | – | – | by demographics |
| CRSW 6 | Included | – | by demographics |

MRP: Multilevel (logit) Regression with Post-stratification.
CRSW: Classical (logit) Regression with Survey Weighting.

**Figure 1.** Testing the Predictive Accuracy of the Estimates

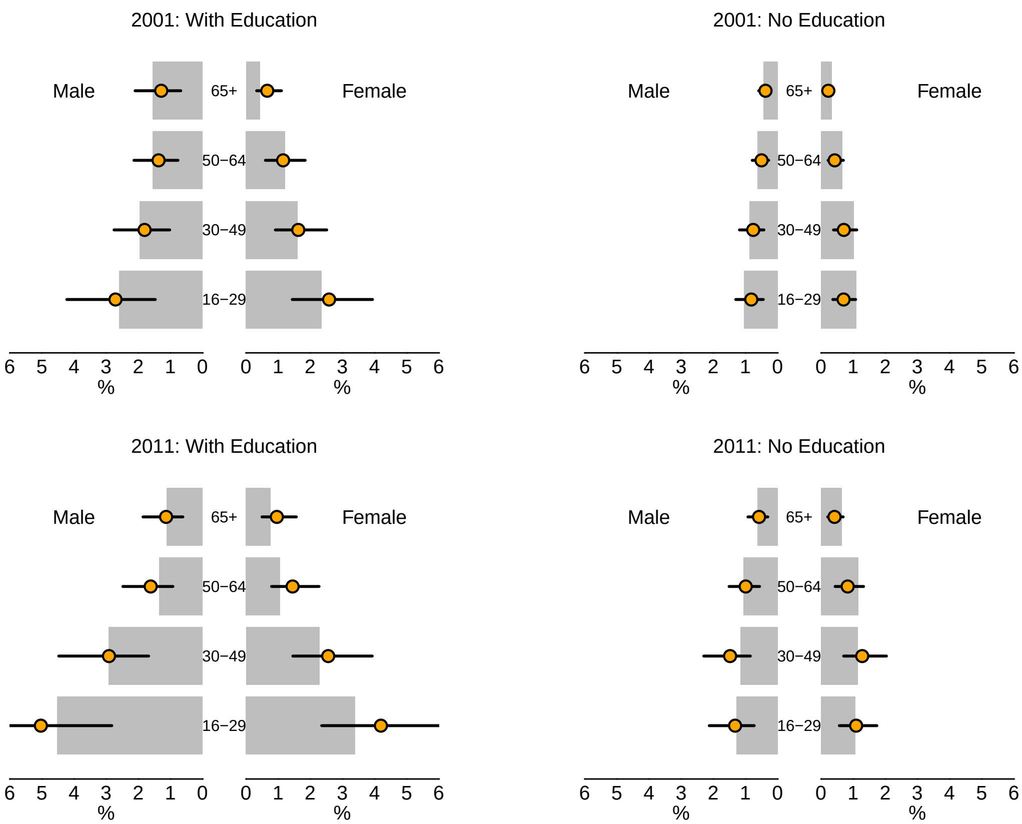

Average percent error and correlations calculated using 32 model and census estimates of religious group prevalence within gender by age by education by year (2001 and 2011) subgroups. Filled circles indicate Multilevel logit Regression with Post-stratified (MRP) estimates; hollow circles, Classical logit Regressions with Survey Weighted (CRSW) estimates; filled triangle: disaggregated estimates. Models are listed on the y-axes and are ranked in descending order of accuracy by average percent error.

**Figure 2.** Bayesian MRP Estimates of Muslim Demographic Subgroup Size



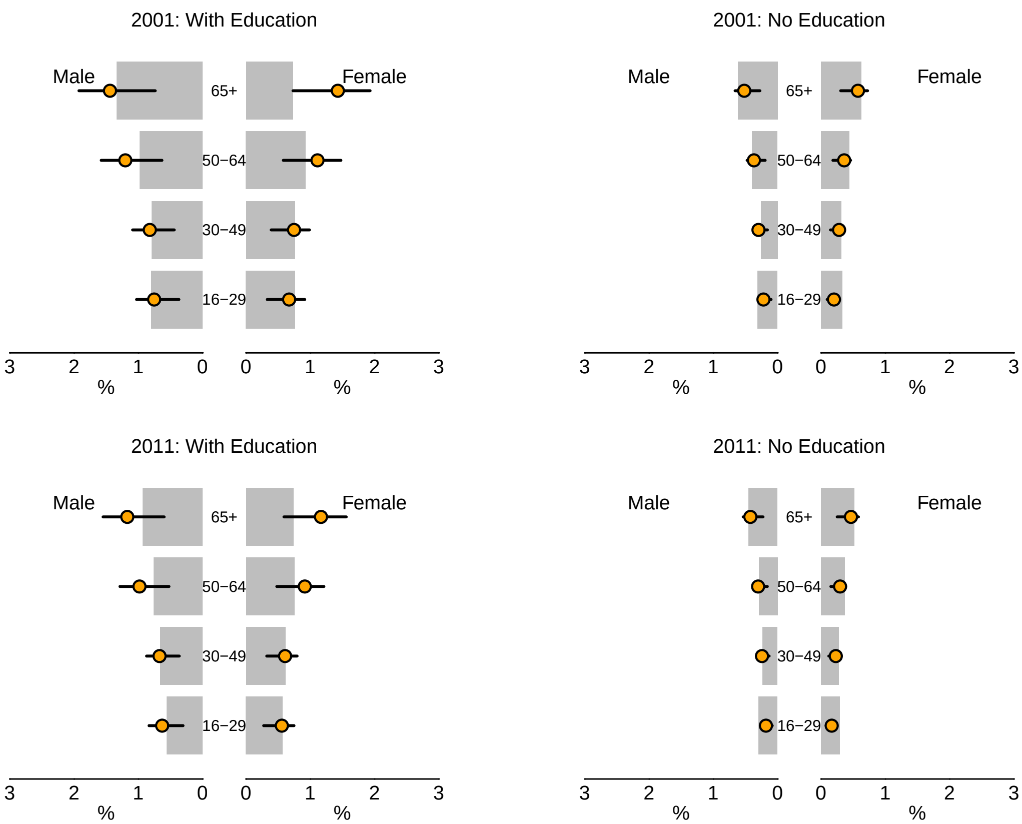

Plot indicates the estimated and actual (census) prevalence of Muslims within census year (2001 and 2011) by education by age by gender category. Estimates are indicated using orange circles, with 80 percent credible intervals shown using black lines. Census measures of Muslim prevalence within each subgroup are shown using grey bars. MRP4 estimates shown.

**Figure 3.** Bayesian MRP Estimates of Hindu Demographic Subgroup Size




Plot indicates the estimated and actual (census) prevalence of Hindus within census year (2001 and 2011) by education by age by gender category. MRP4 estimates shown.

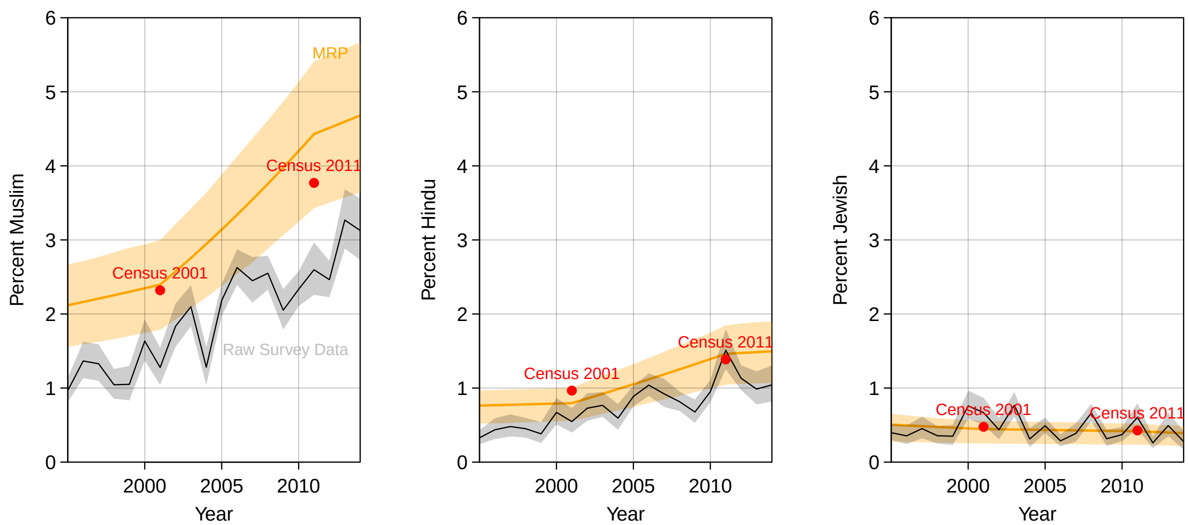
**Figure 4.** Bayesian MRP Estimates of Jewish Demographic Subgroup Size




Plot indicates the estimated and actual (census) prevalence of Jews within census year by education by age by gender category. MRP3 estimates shown.

**Figure 5.** Bayesian MRP Estimates of Religious Minority Identity in the UK, 1995–2014

**1.** Muslim **2.** Hindu **3.** Jewish




Black lines and grey regions show the yearly religious group size estimates obtained by disaggregating the survey dataset by year, along with 80% confidence intervals. Orange lines and regions show the MRP estimates and attendant 80% credible intervals. Red circles indicate the 2001 and 2011 census estimates. Best-fitting estimates displayed: Muslim and Hindu, MRP4; Jewish: MRP3. The steps in the Muslim and Hindu MRP estimates are due to the use of linearly interpolated census population estimates between the years of 2001 and 2011 and uninterpolated 2001 or 2011 estimates outside this window.