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

Online Supplementary Materials

Contents

1. Choosing Grouping Variables

We consider five criteria in choosing demographic grouping variables. First, we searched for variables which were available in both our survey and census data. We were interested in including both design-based variables, relating to the methods by which the survey samples were selected (such as household size and geographic region), and model-based variables, associated either with survey non-response or with respondents' selection of a religious minority identity (such as age and ethnicity). Second, we examine the degree to which the survey data are unbalanced, compared to the census data, on each of the demographic factors (Table A1). Third, we examine the bivariate associations between demographic variables and each religious identity (Table A2). Fourth, we consider the extent of missing data (Table A1). 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.

In addition to religious identity, we were able to extract identically-coded data for age, gender, education, ethnicity and household size from the three survey projects and the census. Age, gender, education, ethnicity are factors that might be expected to shape the way in which respondents answered the religion survey question, or indeed, whether they responded at all. As such, these can be thought of as variables that would assist in a model-based approach to estimating religious demography. Household size, in contrast, is strongly related to the probability of selection into the sample. All three survey projects select households, and then select a particular respondent from each selected household. As is well known, respondents living in larger households thus have a lower probability of being selected into the sample. We include household size as a design-based factor that is additionally likely to be related to particular religious identities.

Two further design-based factors that we considered for inclusion were region and urban-rural area of residence. We were unable to include either due to inconsistencies or incompatibilities in the definitions of these across our four sources of data. Different categories for rural-urban status were used across our four sources. Official definitions of UK regions have changed considerably over time. Before 1996, there were 8 Standard Statistical Regions (SSR) in England, plus Scotland and Wales. In 1996 Government Offices for the Regions (GOR) became the primary regional classification system. There were originally 10 GORs in England. Although some of the GORs covered the same areas as SSRs, others did not. In addition, there were several changes to the GOR definitions in the following years (e.g. Merseyside was included with Yorkshire and The Humber in 1998.)[1]

We are left with five demographic factors, plus religious identity. Table A1 compares the distributions of these demographic factors in the census to the survey data. Census data is available for the years 2001 and 2011, so we split the survey data to roughly correspond to these two time points. We compare the 1995-2005 survey data to the 2001 census, and the 2006-2014 survey data to the 2011 census. We also report the proportion of missing data for each of the demographic factors.

Beginning with religious identity as the variable of prime interest, we see that although

[1] See `http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/government-office-regions/index.html` for further details.

Table A1. Survey and Census Marginals on all Demographic Variables

	Census		Survey		
	2001	2011	1995-2005	2006-2014	1995-2014
	Marginals	Marginals	Marginals	Marginals	Missing
Religion					0.0
Muslim	2.3	3.8	1.5	2.6	
Jewish	0.5	0.4	0.5	0.4	
Hindu	1.0	1.4	0.6	1.0	
Gender					0.0
Female	52.0	51.3	55.8	55.8	
Male	48.0	48.7	44.2	44.2	
HH size					0.0
0-1	16.2	16.3	27.3	29.0	
2-4	73.0	72.0	65.4	64.2	
5+	10.8	11.7	7.3	6.8	
Age					0.4
16-29	21.5	22.6	17.1	14.7	
30-49	36.7	34.7	37.9	34.7	
50-64	22.3	22.7	22.0	24.3	
65+	19.5	19.9	23.0	26.4	
Education					5.3
Degree	18.0	27.5	13.5	22.3	
No degree	82.0	72.5	86.5	77.7	
Ethnicity					17.3
Non-white	7.0	11.6	5.8	8.2	
White	93.0	88.4	94.2	91.8	

Cell entries are percentages. Survey $N = 91{,}862$.

the pooled survey data produce an accurate estimate of the relative size of the Jewish population (0.5 percent in both 2001 and 2011), they tend to underestimate the size of Muslims and Hindus. Whereas UK census reports that in 2001 2.3 percent and in 2011 3.8 percent of the population were Muslim, survey data around the same time periods estimate only 1.5 and 2.6 percent, respectively. These survey estimates are 32 to 35 percent below the census results. And according to the census Hindus comprised 1 percent of the population in 2001 and 1.4 percent in 2011, but survey based estimates are 29 to 39 percent below these proportions. This finding underscores a key assumption and motivation for our work, i.e. that survey samples of religious minorities tend to be unrepresentative of the true population.

Survey samples and true population also differ with regard to other standard demographics. They tend to overrepresent women (by 3.8 percentage points in 2001 and 4.5 points in 2011), respondents living in smaller households (e.g. compare 16.2 percent and 16.3 percent single households in the census to 27.3 percent and 29.0 percent in the survey data) and older people (e.g. 65 year olds and older are overrepresented by 3.5 and 6.5 points). At the same time, respondents with an educational degree tend to be underrepresented, as are mem-

Table A2. Distribution of the Survey Data Across Demographic Categories and Religious Minority Identities

	Muslim	Jewish	Hindu	Total
Gender				
Female	911	224	353	51,242
Male	958	186	372	40,610
Pearson χ^2 test	38.1*	0.2	14.6*	
Household size				
0-1	230	130	83	25,877
2-4	1,012	244	516	59,490
5+	626	36	125	6,480
Pearson χ^2 test	2,095.0*	5.3	184.9*	
Age				
16-29	656	39	167	14,533
30-49	920	129	370	33,155
50-64	201	99	123	21,201
65+	79	141	63	22,591
Pearson χ^2 test	989.4*	29.1*	156.0*	
Education				
Degree	452	134	304	15,584
No degree	1,373	258	414	71,405
Pearson χ^2 test	59.0*	69.8*	292.0*	
Ethnicity				
Non-white	1,312	20	558	5,383
White	202	343	34	70,577
Pearson χ^2 test	14,842.4*	1.1	6,872.2*	
Project				
BSAS	1,345	339	535	69,285
EB	339	46	131	15,517
ESS	185	25	59	7,060
Pearson χ^2 test	16.9*	12.1*	1.1	
Total	1,869	410	725	91,862

Cells either report the sample size for that combination of demographic group and religious identity, or the Pearson χ^2 statistic of the test of difference across a contingency table. $^*p < .05$.

bers of ethnic minorities. These five demographic categories are thus possible candidates for modeling and post-stratification variables. Using a model to predict (for example) Muslim religious identity across the household size categories, post-stratifying those estimates by the population proportions within those household size categories, and then aggregating across the three categories may well improve our estimates of Muslim group size.

However, when choosing demographic grouping factors, we should also take account of the degree to which these factors predict Muslim (for example) identity in the survey data. We examine the evidence in Table A2, which presents the distributions of demographic categories across the three religious identities, as well as Pearson χ^2 tests of the association between each factor and each religious identity. Jewish identity is not strongly related to any of the demographics, although age, education and ethnicity show significant associations. Thus, although our survey data undersamples men, those in large households, the young, non-white, and educated, this does not significantly affect the representativity of those holding a Jewish identity because these factors are not strongly related to holding a Jewish identity.

The situation is quite different for Muslims and Hindus. Holding either identity is significantly related to all the demographic factors we consider, and quite substantially so for household size, age, and especially, ethnicity. Part of the reason that the raw survey data under-represent Muslims and Hindus is that these religious minorities are more likely to be ethnic minorities, live in larger households, and younger in age. These are also groups that public opinion surveys undersample: larger households, by design, and age and ethnic minority status because of sampling difficulties.

The essence of our method is to correct for this imbalance between survey and population data by using a model to predict religious minority identity within each age by gender by age by household size demographic sub-category, and then weighting these by their population proportions to arrive at a post-stratified and hopefully more accurate estimate of group size. We expect that this method will have little effect on the Jewish estimates, because the raw survey data is already quite accurate. But we expect that this method will lead to significantly improved estimates of Muslim and Hindu group size.

Although ethnicity appear to be a crucial variable to include in the modeling and post-stratification steps, due to its association with both Muslim and Hindu identity, data for this variable is not collected by the EB.

Our method leverages this crucial information and corrects for these biases by including it in the statistical model and the post-stratification weights. As Table A1 shows, including ethnicity means that we lose 17 percent of our survey sample. Whether the improvement in estimation afforded by including ethnicity outweighs the loss of data is an empirical question that we address by specifying models both with and without ethnicity and comparing their predictive accuracy.

2. Indicative R Code

```
### Multilevel model estimation
#################################################

## load relevant libraries

library(arm)
library(dplyr)
library(parallel)
library(rstanarm)

## set options

options(mc.cores = parallel::detectCores())

## fit muslim model with ethnicity and 2-way interactions
## survey data have already been pooled, recoded and saved as 'survdat' dataframe

mus_mod_4 = stan_glmer(Muslim ~ YrStd + (1|Sex) + (1|HHsize) + (1|Educ) + (1|AgeCat)
                       + (1|Ethnic) + (1|Project) + (1|AgeCat_Sex) + (1|AgeCat_Educ)
                       + (1|AgeCat_HHsize) + (1|AgeCat_Ethnic) + (1|Sex_Educ)
                       + (1|Sex_HHsize) + (1|Sex_Ethnic) + (1|Educ_HHsize)
                       + (1|Educ_Ethnic) + (1|Ethnic_HHsize),
                       data = survdat,
                       family = binomial,
                       chains = 4,
                       cores = 4,
                       iter = 400,
                       warmup = 150,
                       prior_intercept = normal(-4, 1, autoscale = FALSE),
                       prior = normal(0, 2, autoscale = FALSE),
                       adapt_delta = 0.98)

## save summary and MCMC samples for later analysis

write.csv(as.data.frame(mus_mod_4), "Mus_RStanARM_4_samples.csv")
write.csv(as.data.frame(summary(mus_mod_4)), "Mus_RStanARM_4_summary.csv")

...

### Post-stratification
#################################################

## rearrange census data to obtain population proportions for each of the 1,920 cells
## census data have already been recoded, aligned with survey data and saved as 'cendat':
## a data frame with 96 rows (for each permutation of demographic subgroup) and 25 columns
## (the 20 years of data plus 5 variables indicating age, gender, ethnicity, etc.)

pop.vector = stack(cendat, names(cendat)[6:25])
dem.inds = apply(cendat[,1:5], 2, rep, 20)
pop.data = data.frame(dem.inds, pop.vector[,2], pop.vector[,1])
colnames(pop.data)[6:7] = c("Year", "PopProp")
```

```
## create design matrix to calculate predicted effects from MCMC

sex.vec = matrix(rep(c(rep(1, 48), rep(2, 48)), 20), nrow=1)
hhs.vec = matrix(rep(c(rep(1, 16), rep(2, 16), rep(3, 16)), 2*20), nrow=1)
age.vec = matrix(rep(c(rep(1, 4), rep(2, 4), rep(3, 4), rep(4, 4)), 6*20), nrow=1)
eth.vec = matrix(rep(c(rep(1, 2), rep(2, 2)), 24*20), nrow=1)
edu.vec = matrix(rep(c(1, 2), 48*20), nrow=1)
year.vec = matrix(rep(sort(unique(survdat$YrStd)), each=96), nrow=1)
age.sex.vec = interaction(age.vec, sex.vec)
age.edu.vec = interaction(age.vec, edu.vec)
age.hhs.vec = interaction(age.vec, hhs.vec)
age.eth.vec = interaction(age.vec, eth.vec)
sex.edu.vec = interaction(sex.vec, edu.vec)
sex.hhs.vec = interaction(sex.vec, hhs.vec)
sex.eth.vec = interaction(sex.vec, eth.vec)
edu.hhs.vec = interaction(edu.vec, hhs.vec)
edu.eth.vec = interaction(edu.vec, eth.vec)
eth.hhs.vec = interaction(eth.vec, hhs.vec)

## load MCMC samples

mus_4_mcmc = read.csv("Mus_RStanARM_4_samples.csv")
names(mus_4_mcmc)

## extract effects from model results

sex.int = mus_4_mcmc[, 82:83]
educ.int = mus_4_mcmc[, 80:81]
hhsize.int = mus_4_mcmc[, 73:75]
age.int = mus_4_mcmc[, 69:72]
project.int = mus_4_mcmc[, 76:77]
ethn.int = mus_4_mcmc[, 78:79]
int = mus_4_mcmc[, 1]
year.eff1 = mus_4_mcmc[, 2]
age.hhs.int = mus_4_mcmc[ ,3:14]
age.eth.int = mus_4_mcmc[ ,15:22]
age.edu.int = mus_4_mcmc[ ,23:30]
age.sex.int = mus_4_mcmc[ ,31:38]
eth.hhs.int = mus_4_mcmc[ ,39:44]
edu.hhs.int = mus_4_mcmc[ ,45:50]
sex.hhs.int = mus_4_mcmc[ ,51:56]
edu.eth.int = mus_4_mcmc[ ,57:60]
sex.eth.int = mus_4_mcmc[ ,61:64]
sex.edu.int = mus_4_mcmc[ ,65:68]

## calculate matrix of raw predictions for 1920 cells (columns) and MCMC draws (rows)

cell.pred = as.matrix(invlogit(
    int + year.eff1 %*% year.vec
    + sex.int[, sex.vec]
    + educ.int[, edu.vec]
    + age.int[, age.vec]
    + hhsize.int[, hhs.vec]
```

```
    + ethn.int[, eth.vec]
    + age.hhs.int[, age.hhs.vec]
    + age.eth.int[, age.eth.vec]
    + age.edu.int[, age.edu.vec]
    + age.sex.int[, age.sex.vec]
    + eth.hhs.int[, eth.hhs.vec]
    + edu.hhs.int[, edu.hhs.vec]
    + sex.hhs.int[, sex.hhs.vec]
    + edu.eth.int[, edu.eth.vec]
    + sex.eth.int[, sex.eth.vec]
    + sex.edu.int[, sex.edu.vec]
    ))

dimnames(cell.pred) = NULL

## Weigh predicted effects matrix by cell population proportions

year.pred = matrix(NA, ncol=length(unique(survdat$Year)),
                       nrow=dim(cell.pred)[1])

for (i in 1:dim(cell.pred)[1]) {
  cell.wt.temp = cell.pred[i,] * pop.data$PopProp
  year.pred[i,] = as.numeric(by(cell.wt.temp, pop.data$Year, sum))
}

colnames(year.pred) = unique(pop.data$Year)

## calculate point estimates for each year and 80% uncertainty intervals

mus_4_pred = data.frame( "Prop.mean" = colMeans(year.pred),
                         "Low.80" = apply(year.pred, 2, quantile, 0.1),
                         "Upp.80" = apply(year.pred, 2, quantile, 0.9) )
mus_4_pred
```

3. Additional Tables and Figures

We have demonstrated that our proposed method performs very well in estimating religious minority shares for the socio-demographic cells of main interest (i.e. by age × gender × eduction). Showing the performance for different socio-demographic cells including household size would further increase the confidence in our approach. Table A3 therefore presents the absolute estimation errors for different socio-demographic break-ups of varying granularity. These are based on the best fitting models for each religious minority group (MRP4 for Muslims and Hindus and MRP3 for Jews).

We find that the absolute error for the most detailed socio-demographic break-up – the *full joint distribution* defined by gender, age, education and household size – is roughly double the size of the age-gender-education specific cells: .20 percentage points for Jews, .51 for Hindus, and 1.12 for Muslims. While we would argue that these somewhat larger estimation errors give no reason for concern, they nonetheless caution us of possible limitations in the use of MRP for estimating religious demography at this fine level of aggregation. The performance in estimating religious minority shares for the remaining bi- and trivariate distributions falls in between the results discussed so far.

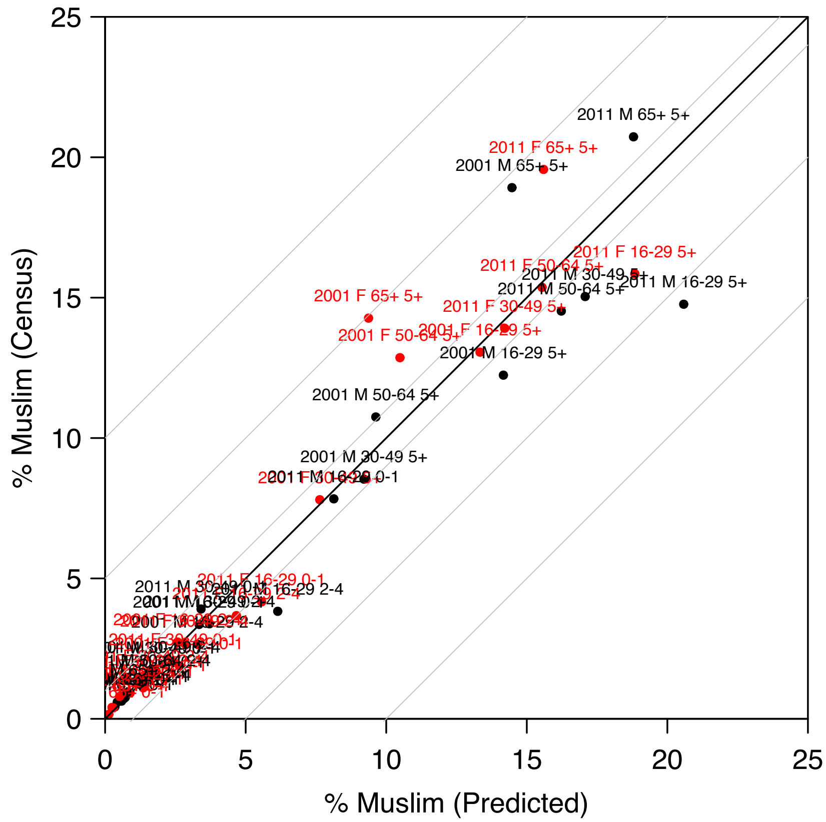
Table A3. Absolute Estimation Error for Different Socio-demographic Cells

	Muslim	Hindu	Jews
gender × age × education	0.58	0.22	0.11
gender × household size	0.60	0.19	0.05
age × household size	0.90	0.29	0.11
education × household size	0.68	0.26	0.07
gender × age × household size	0.93	0.33	0.11
gender × education × household size	0.77	0.30	0.08
age × education × household size	1.01	0.49	0.19
gender × age × education × household size	1.12	0.51	0.20

Estimates for Muslim and Hindu are based on model MRP4, for Jews, MRP 3.

Upon closer inspection, variable combinations including age and household size tend to exhibit relatively higher estimation errors. Figure A1 illustrates this pattern for Muslims. When compared to actual census data MRP tends to underestimate the share of older Muslims (65+ years of age) in large households (more than 5 members). For the year 2001, the estimate of old female Muslims in large households is especially poor and amounts to an underestimate of 5 percentage points. The performance is only slightly better in 2011 and for old male Muslims living in large households. Conversely, we find that at the same time our MRP approach tends to overestimate young Muslims (16 to 29 years of age) in large households. For the year 2011, this overestimate of young male Muslims in large households is more than 5 percentage points. One plausible interpretation is that older Muslims living in large households belong to a more traditional, less assimilated segment of the population which may lack the language skills required for answering surveys. Instead, younger members of the household step in. In any case, this pattern provides some evidence for survey selection issues that are not eliminated by the poststratification step and thus point to potential limitations of MRP in the estimation of religious demography.

Figure A1. Comparison of Estimated and Census Proportions For 48 Demographic Groups Defined by Year × Gender × Age × Household Size



Estimates based on Muslim model MRP4. Red indicates female subgroups, black indicates males. The model underestimates older Muslims in large households and overestimates young male Muslims in large households.