

# Improving multilevel regression and poststratification with structured priors

Yuxiang Gao<sup>1</sup>

Department of Statistical Sciences, University of Toronto

Lauren Kennedy<sup>2</sup>

Columbia Population Research Center and Department of Statistics, Columbia University

Daniel Simpson<sup>3</sup>

Department of Statistical Sciences, University of Toronto

Andrew Gelman<sup>4</sup>

Department of Statistics and Department of Political Science, Columbia University

## Abstract

A central theme in the field of survey statistics is estimating population-level quantities through data coming from potentially non-representative samples of the population. Multilevel Regression and Poststratification (MRP), a model-based approach, is gaining traction against the traditional weighted approach for survey estimates. MRP estimates are susceptible to bias if there is an underlying structure that the methodology does not capture. This work aims to provide a new framework for specifying structured prior distributions that lead to bias reduction in MRP estimates. We use simulation studies to explore the benefit of these prior distributions and demonstrate their efficacy on non-representative US survey data. We show that structured prior distributions offer absolute bias reduction and variance reduction for posterior MRP estimates, regardless of data regime.

## 1 Introduction

Multilevel regression and poststratification (MRP) is an increasingly popular tool for a non-representative sample to a larger population. In particular, MRP appears to

---

Daniel Simpson and Yuxiang Gao was funded by the Canadian Natural Sciences and Engineering Research Council and the Canadian Research Chairs program. Andrew Gelman was supported by the U.S. Office of Naval Research.

<sup>1</sup>ygao@utstat.toronto.edu

<sup>2</sup>lak2183@columbia.edu

<sup>3</sup>simpson@utstat.toronto.edu

<sup>4</sup>gelman@stat.columbia.edu

be effective in areas where traditional design-based survey approaches have traditionally struggled, notably small-area estimation (Pfeffermann et al., 2013; Rao, 2014; Zhang et al., 2014) and with convenience sampling (Wang, Rothschild, Goel, & Gelman, 2015).

One difference between MRP and traditional poststratified design-based weights is that MRP uses partial pooling. Simple poststratification has difficulties with empty cells, in which case the usual practice is to poststratify only on marginals (thus ignoring interactions), or pool cells together. In contrast, the partial pooling of multilevel modeling automatically regularizes group estimates.

Although other options for regularization with MRP have been explored (Bisbee, 2019; Gelman, 2018), applications of MRP typically assume independent group-level errors, for example in a political poll modeling varying intercepts for states using a regression on region indicators, state-level predictors such as previous voting patterns in the state, plus independent errors at the state level. In some applications, though, there is potential benefit from including underlying structure not captured by regression predictors. We demonstrate that this structure can be captured through more complex prior specifications. For example, instead of independent errors for an ordered categorical predictor, we specify an autoregressive structure instead. Ordered predictors are just one example where we can introduce *structured prior distributions*.

### 1.1 Post-sampling adjustments for non-representativeness and MRP

Post-sampling adjustments aim to correct for differences between a potentially biased sample and a target population. Poststratification is a commonly used weighting procedure for nonresponse in model-based survey estimates (Little, 1993). It can improve accuracy of estimates but is no silver bullet, since the quality of poststratified estimates depends on the quality of the known information about the population sizes of the strata, along with the assumption that the sample is representative of the population within each poststratification cell. An approximation to poststratification is raking, which is an iterative algorithm using marginal totals (Deming & Stephan, 1940; Lohr, 2009; Skinner, Wakefield, et al., 2017). When adjusting for many factors, raking can yield unstable estimates caused by high variability of the adjusted weights (Izrael, Battaglia, & Frankel, 2009). For a modern overview of current methods of inference and post-sampling adjustments for nonprobability samples, see Elliott, Valliant, et al. (2017). As the demands for small area estimation increase, so too should the utility of MRP. We use structured priors in our proposed improvement for MRP, with the aim of more sensible shrinkage of posterior estimates that should ultimately reduce estimation bias.

MRP has been used in a broad range of applied problems ranging from epidemiology (Downes et al., 2018; Zhang et al., 2014) to social science (Lax & Phillips, 2009; Trangucci, Ali, Gelman, & Rivers, 2018; Wang et al., 2015). MRP's beginnings saw applications in political science (Gelman & Little, 1997; Park, Gelman, & Bafumi, 2004) for the estimation of state-level opinions from national polls. The breadth of its applications has since matured substantially, even to the extent of being used by data journalists (Morris, 2019). One of MRP's appeals to applied researchers is the ability to produce reliable estimates for small areas in the population and simultaneously adjust for non-representativeness.

On the methodology front, Gelman, Lax, Phillips, Gabry, and Trangucci (2016); Ghitza and Gelman (2013) extended MRP to include varying intercepts and slopes for

interactions, along with inference for time series of polls.

## 1.2 Outline for this paper

This work explores alternative regularization techniques with structured prior distributions that lead to absolute bias reduction in MRP estimates. Our methodology of structured priors should not be confused with that of Si, Trangucci, Gabry, and Gelman (2017), who define structured priors as a way to perform variable selection for higher-order interaction terms of independent random effects. Our improvements on estimation precision come from replacing independent distributions of varying coefficients with Gaussian Markov random fields (Rue & Held, 2005).

This paper is structured as follows: Section 2 gives a concise overview of MRP and what's required for the methodology. Section 3 describes our structured priors framework in detail, along with motivation for their use in MRP. Section 4 and 5 presents simulation studies of structured priors across various regimes of non-representative survey data. Section 4 explains the simulation setup and section 5 interprets the simulation results. Bias and variance comparisons are made between structured priors and the classical independent random effects in MRP in section 5. Section 6 contains the application of structured priors in MRP to a real survey data set that's non-representative. Section 7 is the conclusion.

## 2 Overview of MRP

Multilevel regression and poststratification (Gelman & Little, 1997) proceeds by fitting a hierarchical regression model to survey data, and then using the population size of each poststratification cell to construct weighted survey estimates. More formally, suppose that the population can be split into  $K$  categorical variables and that the  $k^{\text{th}}$  categorical variable has  $J_k$  categories. Hence the population can be represented by  $J = \prod_{k=1}^K J_k$  cells. Usually the population contains continuous variables, and in that case these variables will be discretized to form categorical variables. For example, age in a demographic study can be discretized into a finite number of categories. For every cell, there is a known population size  $N_j$ . Increasing the number of groups for a continuous variable will increase the number of cells  $J$  and correspondingly decrease the individual cell population sizes  $N_j$ . Choosing the optimal group size for continuous variables is a difficult model selection problem, involving tradeoffs between accuracy and computational load, and this is something that we do not address in this paper.

Suppose that the response variable of individual  $i$  is  $y_i \in \{0, 1\}$ . MRP for binary survey responses is summarized by the two steps below:

**Multilevel regression step.** Fit the hierarchical logistic regression model below to get estimated population averages  $\theta_j$  for every cell  $j \in \{1, \dots, J\}$ . The hierarchical logistic regression portion of MRP has a set of varying intercepts  $\{\alpha_j^k\}_{j=1}^{J_k}$  for each categorical covariate  $k$ , which have the effect of partially pooling each  $\theta_j$  towards a globally-fitted regression model,  $X_j\beta$ , with sparse cells benefiting the most from this regularization. We

follow a notation consistent with Gelman and Hill (2006).

$$\begin{aligned} \Pr(y_i = 1) &= \text{logit}^{-1} \left( X_i \beta + \sum_{k=1}^K \alpha_j^k \right), \text{ for } i = 1, \dots, n \\ \alpha_j^k \mid \sigma^k &\stackrel{\text{ind}}{\sim} N(0, (\sigma^k)^2), \text{ for } k = 1, \dots, K, j = 1, \dots, J_k \\ \sigma^k &\sim N_+(0, 1), \text{ for } k = 1, \dots, K \\ \beta &\sim N(0, 1), \end{aligned}$$

where we are giving default weakly informative priors to the non-varying regression coefficients  $\beta$ .

**Poststratification step.** Using the known population sizes  $N_j$  of each cell  $j$ , poststratify to get posterior preference probabilities at the subpopulation level. The poststratification portion of MRP adjusts for nonresponse in the population by taking into account the sizes of every cell  $l$  relative to the total population size  $N = \sum_{j=1}^J N_j$ . Another way to interpret poststratification is as a weighted average of cell-wise posterior preferences, where the weighting scheme is determined by the size of each cell in the population. Smaller cells get downweighted and larger cells get upweighted. The final result is a more accurate estimate in the presence of non-representative data.

Let  $S$  be some subset of the population defined based on the poststratification matrix. Then the poststratified estimate for  $S$  is:

$$\theta_S = \frac{\sum_{j \in S} N_j \theta_j}{\sum_{j \in S} N_j}$$

For example,  $S$  could correspond to the oldest age category in the lowest income bracket. Then  $\theta_S$  would correspond to the proportion of people in this sub-population that would respond yes to the survey question of interest.

### 3 Proposed Approach and Motivation

We consider structured prior distributions for MRP taking the form of Gaussian Markov random fields (GMRF), modeling certain structure of the underlying categorical covariate in the hierarchical regression. We proceed as follows for a covariate in the population of interest:

**Case 1:** If we do not want to model any structure in a categorical covariate, we model its varying intercepts as independently normally distributed.

**Case 2:** If there is underlying structure we would like to model in a covariate, and spatial smoothing using this structure seems sensible for the outcome of interest, then we use an appropriate GMRF as a prior distribution for this batch of varying intercepts.

We will specify informative hyperpriors when possible and model via a full Bayesian approach. For a detailed overview of principled hyperprior specification in GMRF models, we refer the reader to Simpson et al. (2017). As well, we do not restrict structured priors to have directed or undirected conditional distributions. Some examples of directed conditional distributions include the autoregressive and random walk processes with discrete

time indices, which are frequently used in time series analysis. The CAR and ICAR processes (Besag, 1975) are common undirected conditional distributions and are often used in specifying priors in spatial models.

More complex prior structure allows for nonuniform information-borrowing in the presence of non-representative surveys from a population. For example, it makes sense to partially pool inferences for the oldest age group toward data from the second-oldest group. An autoregressive prior placed on the ordinal variable age achieves this effect, without making the strong global assumptions involved in simply including age as a linear or quadratic predictor in the regression. The proposal of using structured priors aims to reduce bias for MRP estimates in extremely non-representative data regimes.

Structured priors improve upon the multilevel aspect of MRP while maintaining the regression structure. Because MRP is a model-based survey estimation approach, the multilevel regression component can be replaced with other forms of regression modelling, for example sparse hierarchical regression (Goplerud, Kuriwaki, Ratkovic, & Tingley, 2018), Bayesian additive regression trees (Bisbee, 2019). It is important, though, that the regression step be regularized in some way to preserve the ability of the method to account for a potentially large number of adjustment factors and their interactions (Gelman, 2018).

## 4 Simulation Studies

### 4.1 Models for partial pooling of group-level errors

For the purpose of explaining our proposed method of MRP using structured priors, we work with a simple model of three poststratification categories—51 states, age in years ranging from 21–80, and income in 4 categories—and no other predictors. Age is further categorized into 12 groups. We define  $\alpha_{j[i]}^{\text{Age Cat.}}$ ,  $\alpha_{j[i]}^{\text{Income}}$  and  $\alpha_{j[i]}^{\text{Region}}$  to be the varying intercepts for age category, income category and region respectively for the  $i^{\text{th}}$  survey respondent. For all three prior specifications of MRP, we use the link function,

$$\Pr(y_i = 1) = \text{logit}^{-1} \left( \beta^0 + \alpha_{j[i]}^{\text{State}} + \alpha_{j[i]}^{\text{Age Cat.}} + \alpha_{j[i]}^{\text{Income}} \right), \text{ for } i = 1, \dots, n. \quad (1)$$

For all three prior specifications we assume independent mean-zero normal distributions for the  $\alpha_{j[i]}^{\text{Region}}$ 's,  $\alpha_{j[i]}^{\text{Age Cat.}}$ 's and  $\alpha_{j[i]}^{\text{Income}}$ 's along with a weakly informative half-normal distribution for the corresponding scale parameter:

$$\begin{aligned} \alpha_j^{\text{State}} | \beta^{\text{State-VS}}, \beta^{\text{Relig.}}, \{\alpha_{j*}^{\text{Region}}\}_{j*=1}^5, \sigma^{\text{State}} &\stackrel{\text{ind.}}{\sim} N(\alpha_{m[j]}^{\text{Region}} + \beta^{\text{Relig.}} X_{\text{Relig.},j} \\ &+ \beta^{\text{State-VS}} X_{\text{State-VS},j}, (\sigma^{\text{State}})^2), \\ &\text{for } j = 1, \dots, 51 \\ \alpha_m^{\text{Region}} | \sigma^{\text{Region}} &\stackrel{\text{ind.}}{\sim} N(0, (\sigma^{\text{Region}})^2), \text{ for } m = 1, \dots, 5 \quad (2) \\ \alpha_j^{\text{Income}} | \sigma^{\text{Income}} &\stackrel{\text{ind.}}{\sim} N(0, (\sigma^{\text{Income}})^2), \text{ for } j = 1, \dots, 4 \\ \sigma^{\text{Income}}, \sigma^{\text{State}}, \sigma^{\text{Region}} &\sim N_+(0, 1) \\ \beta^{\text{State-VS}}, \beta^{\text{Relig.}}, \beta^0 &\sim N(0, 1) \end{aligned}$$

where  $X_{\text{State-VS},j} \in [0, 1]$  is the covariate that corresponds to the 2004 Democratic vote share for state  $j$  and  $X_{\text{Relig.,}j} \in [0, 1]$  is the percentage of conservative religion in state  $j$ , which is defined as the sum of the percentage of Mormons and percentage of Evangelicals in state  $j$ . The term  $\alpha_{m[j]}^{\text{Region}} + \beta^{\text{Relig.}} X_{\text{Relig.,}j} + \beta^{\text{State-VS}} X_{\text{State-VS},j}$  are state-level predictors that utilize auxiliary data accounting for structured differences among the states.

The *baseline specification* is the classical prior distribution used in MRP with independent normal distributions for the varying intercepts for age categories:

$$\begin{aligned} \alpha_j^{\text{Age Cat.}} &| \sigma^{\text{Age Cat. ind.}} \sim N(0, (\sigma^{\text{Age Cat.}})^2), \text{ for } j = 1, \dots, 12 \\ \sigma^{\text{Age Cat.}} &\sim N_+(0, 1) \end{aligned} \quad (3)$$

The *autoregressive specification* models the ordinal structure of age category as a first-order autoregression (Rue & Held, 2005). The prior distribution imposed on  $\rho$  is restricted to the range  $(-1, 1)$ , enforcing stationary for the autoregressive process.

$$\begin{aligned} \alpha_1^{\text{Age Cat.}} &| \rho, \sigma^{\text{Age Cat.}} \sim N(0, \frac{1}{1 - \rho^2} (\sigma^{\text{Age Cat.}})^2) \\ \alpha_j^{\text{Age Cat.}} &| \alpha_{j-1}^{\text{Age Cat.}}, \dots, \alpha_1^{\text{Age Cat.}}, \rho, \sigma^{\text{Age Cat.}} \sim N(\rho \alpha_{j-1}^{\text{Age Cat.}}, (\sigma^{\text{Age Cat.}})^2), \text{ for } j = 2, \dots, 12 \\ \sigma^{\text{Age Cat.}} &\sim N_+(0, 1) \\ (\rho + 1)/2 &\sim \text{Beta}(0.5, 0.5). \end{aligned} \quad (4)$$

Finally, we consider the *random walk specification*, which is a special case of first-order autoregression with  $\rho$  fixed as 1, although with a different parameterization to avoid the division by  $1 - \rho^2$  above. In addition, we introduce the constraint  $\sum_{j=1}^J \alpha_j^{\text{Age Cat.}} = 0$  to ensure that the joint distribution for the first-order random walk process is identifiable.

$$\begin{aligned} \alpha_j^{\text{Age Cat.}} &| \alpha_{j-1}^{\text{Age Cat.}}, \dots, \alpha_1^{\text{Age Cat.}}, \sigma^{\text{Age Cat.}} \sim N(\alpha_{j-1}^{\text{Age Cat.}}, (\sigma^{\text{Age Cat.}})^2), \text{ for } j = 2, \dots, J \\ \sigma^{\text{Age Cat.}} &\sim N_+(0, 1) \\ \sum_{j=1}^J \alpha_j^{\text{Age Cat.}} &= 0. \end{aligned} \quad (5)$$

The three prior specifications differ in the amount of information shared between neighbors in the age category random effect. In the baseline specification, no information is shared between  $\alpha_j^{\text{Age Cat.}}$  and  $\alpha_{j-1}^{\text{Age Cat.}}$  for  $j \in \{2, \dots, 12\}$ . In the autoregressive specification, partial information is shared and in the random walk specification the full amount of information is shared. In the simulation studies below, we empirically show that the property of shrinking towards the previous neighboring variable in the autoregressive and random walk specifications result in decreased posterior bias of MRP estimates for every cell in the population.

## 4.2 Simulated data

**4.2.1 The sample.** We consider three scenarios of true  $E(y)$  as a function of age: U-shaped, cap-shaped, or monotonically increasing.

We investigate the effects of non-representative data amongst elderly individuals (ages 61–80) in the simulation samples, and show that the random walk specification provides the lowest absolute bias in subpopulation level estimates when compared to the other two specifications. The likelihood of sampling from a given subpopulation cell is dependent on the size of the subpopulation group along with the response probability of an individual in that group.

The probability vector of sampling is defined as:

$$\frac{(\text{Probability of response}) \odot (\text{Population})}{\sum_j (\text{Probability of response})_j \cdot (\text{Population})_j}$$

where  $\odot$  is the Hadamard product. This probability vector is in reference to the post-stratification matrix defined for this simulation study. A special case for the probability vector of sampling is when the probability of response is equal for all cells in the population, resulting in a probability vector of sampling that's fully representative of the population. The probability vector of sampling is used to generate a sample of binary responses along with covariates. Through this probability vector, one can augment it to get highly non-representative samples for certain subpopulation groups. In the case of a completely random sample for subpopulation groups of interest, all subpopulation groups of interest have the same probability of sampling. As an example, all 12 age categories would have equal probability of being sampled from in the scenario of completely random sampling for age categories.

**4.2.2 Assumed sample and population.** In the following simulation study we will assume that the population is sufficiently large so that sampling with replacement is equivalent to retrieving a random sample from the population.

To empirically validate the improvements that structured priors have on posterior MRP estimates, we construct various data regimes for age categories 9–12. More specifically, let  $S$  be the index set corresponding to age categories 9–12. Summing the probability of sampling over  $S$  will return the expected proportion of the sample who are older adults. We perturb this probability through 9 scenarios, ranging from 0.05 (under-representing older adults) to 0.82 (over-representing older adults). This section contains plots for the U-shaped true preference curve, with the appendix containing plots for the increasing-shaped true preference curve and the cap-shaped true preference curve.

These three true preference curves capture the rough structure of the unseen truths in real survey data. Let  $x$  represent age of an individual, and let  $f(x)$  represent the probability that such an individual will vote yes to the survey question of interest. The three different preference curves with respect to age are defined as:

**Cap-shaped preference:**

$$f(x) = \frac{\Gamma(4)}{\Gamma(2)\Gamma(2)}x(1-x), \quad x \in [21, 80]$$

**U-shaped preference:**

$$f(x) = 1 - \frac{\Gamma(4)}{\Gamma(2)\Gamma(2)}x(1-x), \quad x \in [21, 80]$$

### Increasing-shaped preference:

$$0.7 - 3\exp\left(-\frac{x}{0.2}\right), \quad x \in [21, 80]$$

True preferences for every poststratification cell  $j \in \{1, \dots, J\}$  in the population are then generated with the following formula:

$$\theta_j = \text{logit}^{-1} \left( \beta^0 + f(X_{\text{Age}[j]}) + X_{\text{Income}[j]} + \beta^{\text{State}} X_{\text{State}[j]} + \beta^{\text{Relig.}} X_{\text{Relig.}[j]} \right)$$

where  $X_{\text{Age}[j]}$ ,  $X_{\text{Income}[j]}$ ,  $X_{\text{State}[j]}$ ,  $X_{\text{Relig.}[j]}$  correspond to the age, income effect, state effect and religion effect respectively of poststratification cell  $j$ .  $X_{\text{Income}[j]}$ ,  $X_{\text{State}[j]}$ ,  $X_{\text{Relig.}[j]}$  along with  $\beta^0$ ,  $\beta^{\text{State}}$  and  $\beta^{\text{Relig.}}$  are defined in the appendix.

## 4.3 Results

We fit all models using the probabilistic programming language Stan (Carpenter et al., 2017) to perform full Bayesian inference, using the default settings of 2000 iterations on 4 chains run in parallel, with half the iterations in each chain used for warmup.

**4.3.1 Impact of prior choice on bias of posterior preferences.** The first way we evaluate the impact of prior specification is by considering the impact of bias when we manipulate the expected proportion of the sample that are older adults. In Figure 1 below, we plot the results for a sample size of 100 and 500.

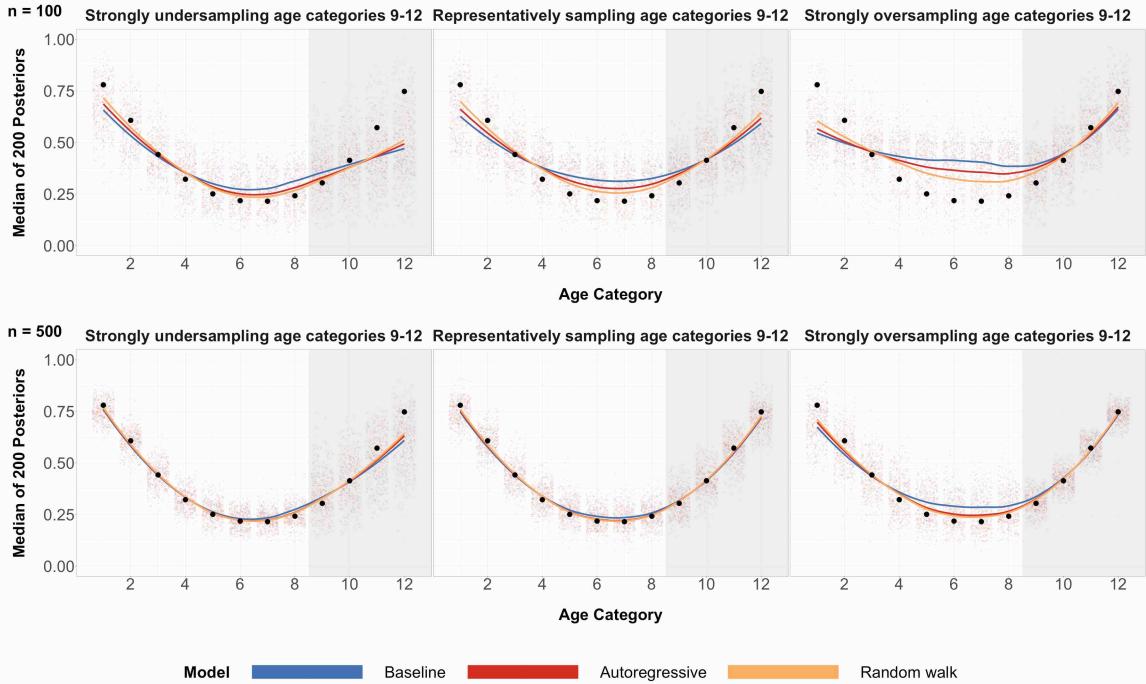
When the expected proportion of the sample that are older adults is equal to 0.33, this corresponds to a completely random sample for age categories (probability of sampling every age category is the same) *and* a fully representative sample for age categories (probability of sampling every age category is proportional to the population sizes for every age category). In certain scenarios, a completely random sample may be more desirable than a fully representative sample of the population for modeling purposes. Certainly, oversampling a sparse subpopulation group in the population will return lower variance model estimates for that specific subpopulation group.

We can see from Figure 1 that the two structured prior specifications outperform the baseline prior specification by a few percentage points for almost all 12 age categories, and achieving the same performance for the remaining age categories.

When elderly individuals are undersampled relative to the rest of the population, the random walk prior specification outperforms the baseline prior specification in lower absolute bias by a few percentage points across all the age categories.

When elderly individuals are oversampled relative to the rest of the population, the random walk prior specification outperforms the baseline prior specification in lower absolute bias by close to 10 percentage points for mid-aged individuals when sample size is 100. As expected, the three prior specifications produce essentially the same posterior estimates in the bottom row of Figure 1, due to the sample size being large in each of these age categories – Increasing  $n$  will increase the weight of the likelihood on the posterior in a statistical model. Regardless, absolute bias is reduced or stays the same for all age categories and all data regimes for the two structured priors specifications, as seen in Figure 1.

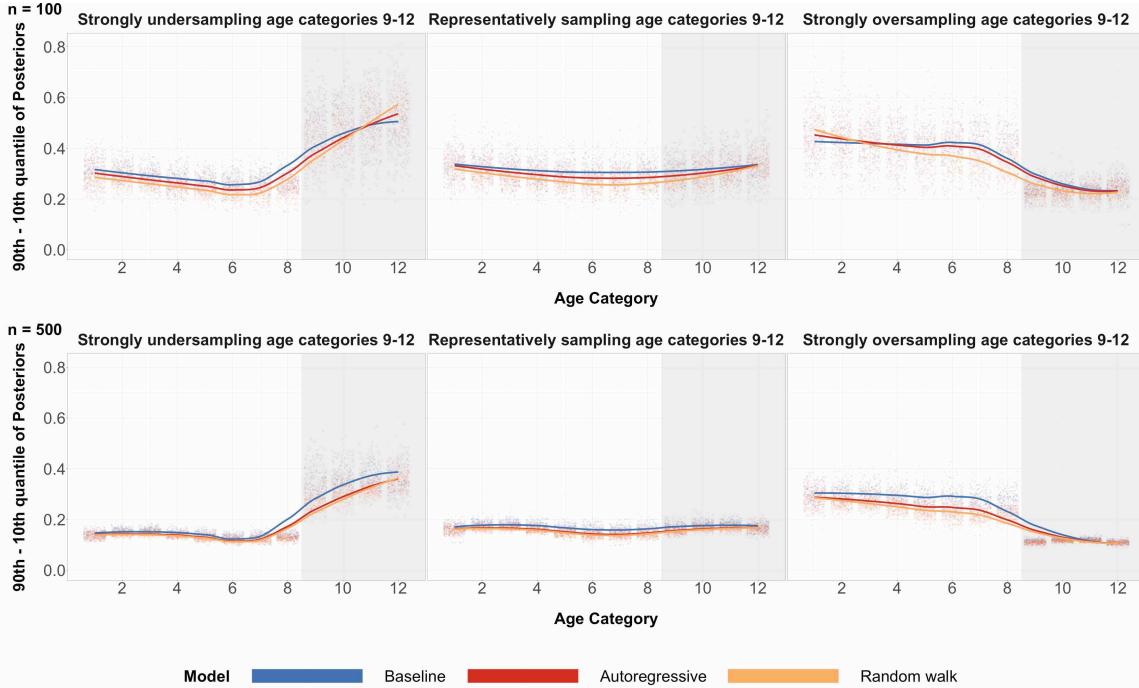
Another visualization of bias reduction is based on Figure 3. It shows bias of posterior preferences for each cell in the population, the finest granularity, as the expected



*Figure 1.* Posterior medians for 200 simulations for each age group under three different regimes of data, where true age preference is U-shaped. The top row corresponds to a sample size of 100 and the bottom row corresponds to a sample size of 500. Black circles are true preferences for each age group. The shaded grey region corresponds to the age categories of older individuals for which we over/undersample. The left column has a probability of sampling age categories 9-12 equal to 0.05. The middle column has a probability of sampling age categories 9-12 equal to 0.33, which is completely random sampling and representative sampling for all age categories. The right column has a probability of sampling age categories 9-12 equal to 0.82. Local regression is used for the smoothed estimates amongst the three prior specifications. For the same plots involving different probabilities of sampling, refer to Table 4 in the appendix.

proportion of the sample that are older adults is perturbed. Absolute bias is significantly decreased when switching from the baseline specification to the random walk specification. The autoregressive specification also reduces absolute bias, but not as much when compared to the random walk specification. This is due to the prior  $\rho$  defining before inference that the information being borrowed from the neighboring age category posteriors should be a value in  $[-1, 1]$ .

As a secondary benefit of structured priors, averaging over all 200 runs, simulation studies had shown the difference of the 90<sup>th</sup> and 10<sup>th</sup> posterior quantiles for almost all age categories to be smaller when  $n = 100$ . This is shown in Figure 2. This difference can be interpreted as a measure of posterior standard deviation. When  $n = 500$ , reduction in posterior quantiles difference is even more apparent. Reduction in posterior standard deviations may not be ideal for estimators when the tradeoff is higher absolute bias, but for the case of structured priors, we see a reduction in both for every age category implying a



*Figure 2.* Differences in the 90<sup>th</sup> and 10<sup>th</sup> posterior quantiles for every age category when true preference is U-shaped for 200 simulations. The top row corresponds to a sample size of 100 and the bottom row corresponds to a sample size of 500. The shaded grey region corresponds to the age categories of older individuals for which we over/undersample. The left column has a probability of sampling age categories 9-12 equal to 0.05. The middle column has a probability of sampling age categories 9-12 equal to 0.33, which is completely random sampling *and* representative sampling for all age categories. The right column has a probability of sampling age categories 9-12 equal to 0.82. Local regression is used for the smoothed estimates amongst the three prior specifications. For the same plots involving different probabilities of sampling, refer to Table 5 in the appendix.

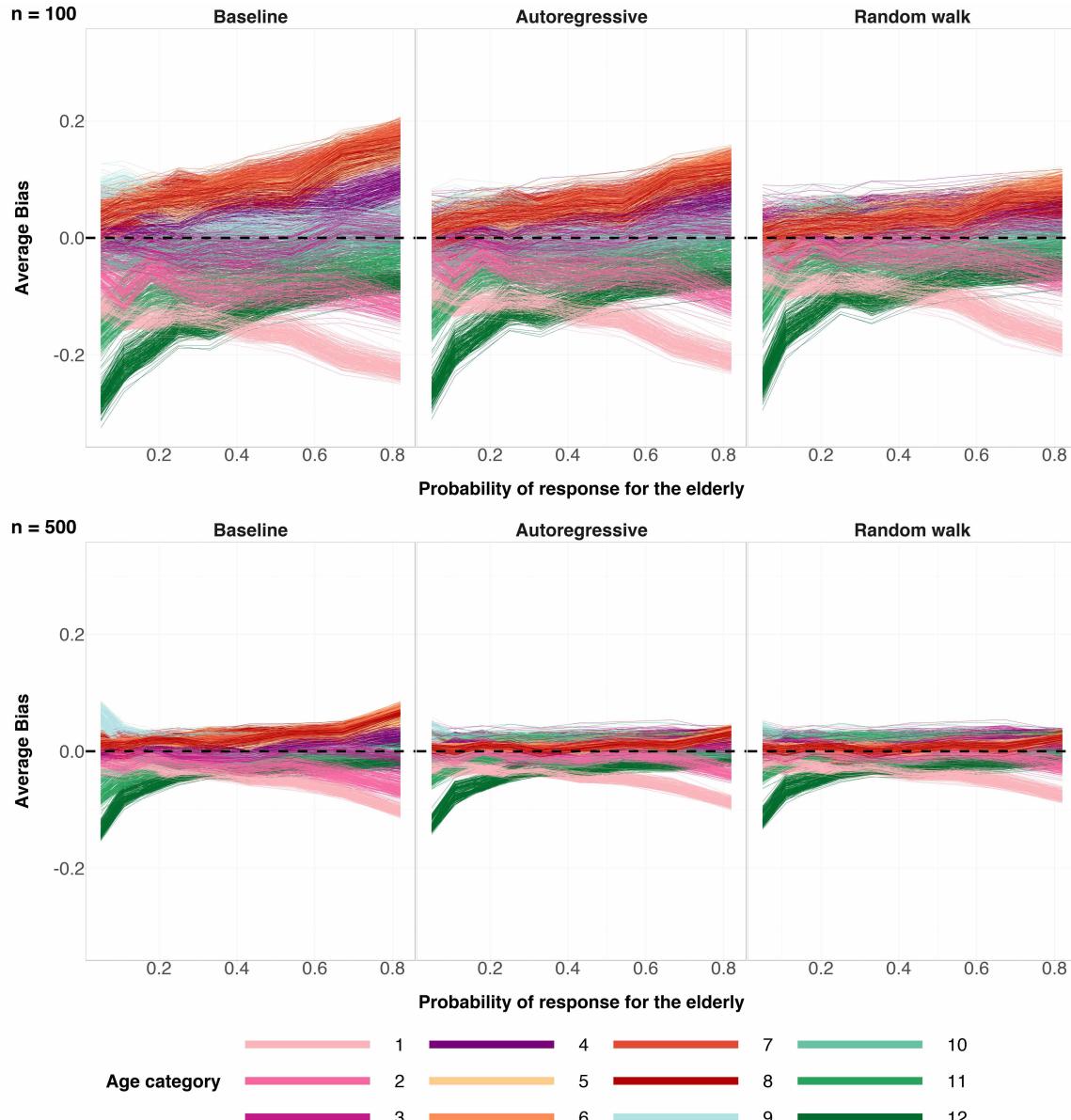
decrease in  $L_2$  risk for posterior estimates of every age category.

The population preference estimates for the three prior specifications remain nearly the same across all probability of sampling indices when the true preference curve is U-shaped or cap-shaped. When the true preference curve is increasing-shaped, the population preference remains nearly the same for all probability of sampling indices except 0.05 and 0.82. In those cases, the first-order random walk prior produces less unbiased population estimates by a few percentage points. The advantage of structured priors appear to be more drastic when reducing to more granular sub-population levels. For additional bias plots on all three true preference curves, the reader can refer to the appendix.

In summary, based on the simulation studies on the U-shaped true preference along with the two other true preference curves, we see that structured priors decrease absolute bias for posterior MRP estimates more than the classical specification of priors in MRP, regardless of how representative the survey data are to the population of interest. This implies that posterior MRP estimates coming from structured priors are much more invariant

to differential nonresponse and biased sampling when compared to the classical priors used in MRP. The main goal of this paper is to argue that structured priors offer an improvement to MRP, even in extremely non-representative data regimes. We indeed see that in the simulation studies as large decreases in absolute bias are seen when the probability of sampling age categories 9-12 are 0.05 and 0.82. A secondary benefit of structured priors is variance reduction on posterior estimates of the structured covariates.

Structured priors start to have a beneficial effect on posterior MRP estimates when the number of categories for the structured covariates of interest is sufficiently large. To quantify "sufficiently large" is problem-dependent as every structured prior will be different depending on the covariates of the data set. Furthermore, there are multiple structured priors one can choose from for a covariate. This is something we will not address here. We previously ran the same set of experiments in this results section for 3 and 6 age categories and did not observe a significant difference in posterior estimates for all three prior specifications. 12 age categories and more for our simulation studies are when the beneficial effects of structured priors become obvious.



*Figure 3.* The average of bias values coming from 200 simulations of posterior medians of the 2448 poststratification cells. Sample size  $n = 100$  (top) and  $n = 500$  (bottom). The true preference curve for age is U-shaped. The horizontal dashed line at  $y = 0$  represents zero bias.

## 5 Analysis on U.S. Survey Data

Along with simulation studies that validate the benefit of structured priors, we further apply our approach to the National Annenberg Election Survey 2008 Phone Edition (NAES08-Phone) (The Annenberg Public Policy Center of the University of Pennsylvania, 2008). NAES08-Phone was a phone survey conducted over the course of the 2008 US Presidential Election and the sampling methodology was based on random telephone number generation. NAES08-Phone observed a response rate of 23 percent. The population comes from the 2006–2010 5-year American Community Survey (ACS, United States Census Bureau / American FactFinder (2010)). The response variable of interest is whether an individual favors gay marriage or not. In 2008, this question was discussed heavily in the political landscape, as some states had not legalized same-sex marriage yet. The covariates used in the Annenberg survey sample are sex, race/ethnicity, household income, state of residence, age, education. The same covariates in the 5-year ACS are used so that poststratification and more specifically MRP can be performed.

### 5.1 National Annenberg Election Survey 2008

Table 1 contains the percentages of each factor for four of the covariates in the 2008 Annenberg phone survey (excluding age and state of residence). A histogram summarizing the age covariate in the Annenberg phone survey is shown in the bottom plot of Figure 5. The size of the Annenberg phone survey is 24,387 respondents.

Sex	Education	Race/Eth.	Income
Male: 43.28	Bachelor: 19.63	American Indian: 1.18	Less than 10000: 4.83
Female: 56.72	Bachelor+: 19.01	Asian : 1.45	[10000, 15000]: 5.14
	Highschool: 28.52	Black : 8.62	(15000, 25000]: 8.75
	No highschool: 7.13	Hispanic : 1.55	(25000, 35000]: 10.35
	Some college: 16.65	Other : 7.03	(35000, 50000]: 15.32
	2 year college: 9.07	White : 80.16	(50000, 75000]: 19.05
			(75000, 100000]: 14.36
			(100000, 150000]: 12.33
			More than 150000: 9.87

Table 1

Percentage of each factor in the Annenberg phone survey for sex, education, race/ethnicity, household income.

### 5.2 Poststratifying to the US population

The 2006–2010 5-year ACS is a weighted probability survey, with weights assigned to every individual in the sample. Based on the weights of individuals in the 5-year ACS, we form a 929,082-row poststratification matrix as seen in Table 3, which we will assume to be representative of the overall population for the 2008 Annenberg phone survey. We will use Table 3 to poststratify the 2008 Annenberg survey estimates to the US population. The ACS is conducted by The Census Bureau and aggregates monthly probabilistic samples to form 1, 3, and 5-year ACS data sets. It aims to capture the most current demographic

information annually, and answering the survey is mandatory according to US Federal Law. For these reasons, we believe that it's the most accurate representation of the US population every year.

Table 2 contains the percentages of each factor for four of the covariates in the 2006–2010 5-year ACS (excluding age and state of residence). A smoothed density summarizing the age covariate in the ACS is shown in the bottom plot of Figure 5. The continuous age covariate in both the 5-year ACS and the Annenberg survey is discretized into either 12, 48, or 72 age categories in our analysis. In theory, the number of poststratification cells for Table 3 is  $2 \times 6 \times 6 \times 9 \times 51 \times 78 = 2,577,744$ . The cells left out by Table 3 are assumed have a population size of 0.

Sex	Education completed	Race/Eth.	Income
Male: 48.07	Bachelor: 16.86	American Indian: 0.50	Less than 10000: 5.09
Female: 51.93	Bachelor+: 9.28	Asian : 4.80	[10000, 15000]: 4.07
	Highschool: 29.17	Black: 11.24	(15000, 25000]: 9.11
	No highschool: 14.87	Hispanic: 5.15	(25000, 35000]: 9.65
	Some college: 22.53	Other: 10.25	(35000, 50000]: 14.04
	2 year college: 7.30	White: 68.06	(50000, 75000]: 20.04 (75000, 100000]: 13.90 (100000, 150000]: 14.21 More than 150000: 9.89

Table 2

Percentage of each factor from the 2006–2010 5-year American Community Survey for sex, education completed, race/ethnicity, household income.

Age	Sex	Race/Eth.	Education	State	Income	N
18	Male	American Indian	Highschool	Alabama	(15000, 25000]	35
18	Female	American Indian	Highschool	Arizona	Less than 10000	124
:	:	:	:	:	:	:
95	Female	White	2 year college	Rhode Island	(100000, 150000]	15

Table 3

Full poststratification matrix for the 5-year American Community Survey.

### 5.3 The models for the 2008 Annenberg phone survey

Figure 4 shows proportions of every state in the Annenberg survey and the differences in proportions between Annenberg survey and ACS. Ideally, there shouldn't be differences in the proportions for both surveys if both ACS and Annenberg survey are equally representative of the US population. Based on the bottom heatmap which corresponds to the Annenberg proportions subtracted by the ACS proportions, California and Texas are shown to be the most under-represented by the Annenberg survey whereas New Hampshire and Missouri are the most over-represented by the Annenberg survey.

For ages 18–40, the smooth ACS density in Figure 5 is higher than the Annenberg histogram, implying the that Annenberg survey underrepresents younger individuals. For

the other demographic traits, relative to the 5-year ACS, Tables 1 and 2 show that the Annenberg survey overrepresents whites and women.

Let  $y_i = 1$  if respondent  $i$  favors same-sex marriage. Then we model,

$$\begin{aligned} \Pr(y_i = 1) = \text{logit}^{-1} & \left( \beta^0 + \alpha_{j[i]}^{\text{Age Cat.}} + \beta^{\text{Sex}} X_{\text{Sex},i} + \alpha_{j[i]}^{\text{Ethnicity}} \right. \\ & \left. + \alpha_{j[i]}^{\text{Education}} + \alpha_{j[i]}^{\text{State}} + \alpha_{j[i]}^{\text{Income}} \right) \end{aligned} \quad (6)$$

We define the *baseline, autoregressive and random walk specifications* to have these prior distributions in common:

$$\begin{aligned} \alpha_j^{\text{Ethnicity}} \mid \sigma^{\text{Ethnicity}} & \stackrel{\text{ind.}}{\sim} N(0, (\sigma^{\text{Ethnicity}})^2), \text{ for } j = 1, \dots, 6 \\ \alpha_j^{\text{Education}} \mid \sigma^{\text{Education}} & \stackrel{\text{ind.}}{\sim} N(0, (\sigma^{\text{Education}})^2) \text{ for } j = 1, \dots, 6 \\ \alpha_j^{\text{State}} \mid \beta^{\text{State-VS}}, \beta^{\text{Relig.}}, \{\alpha_{j^*}^{\text{Region}}\}_{j^*=1}^5, \sigma^{\text{State}} & \stackrel{\text{ind.}}{\sim} N(\alpha_{m[j]}^{\text{Region}} + \beta^{\text{Relig.}} X_{\text{Relig.},j} \\ & + \beta^{\text{State-VS}} X_{\text{State-VS},j}, (\sigma^{\text{State}})^2), \\ & \text{for } j = 1, \dots, 51 \\ \alpha_m^{\text{Region}} \mid \sigma^{\text{Region}} & \stackrel{\text{ind.}}{\sim} N(0, (\sigma^{\text{Region}})^2), \text{ for } m = 1, \dots, 5 \\ \alpha_j^{\text{Income}} \mid \sigma^{\text{Income}} & \stackrel{\text{ind.}}{\sim} N(0, (\sigma^{\text{Income}})^2), \text{ for } j = 1, \dots, 9 \\ \sigma^{\text{Ethnicity}}, \sigma^{\text{Education}}, \sigma^{\text{State}}, \sigma^{\text{Region}}, \sigma^{\text{Income}} & \sim N_+(0, 1) \\ \beta^{\text{Sex}}, \beta^{\text{State-VS}}, \beta^{\text{Relig.}}, \beta^0 & \sim N(0, 1) \end{aligned} \quad (7)$$

Let  $J^{\text{Age Cat.}}$  be the number of categories for the continuous covariate age. The *baseline specification* has the prior distributions:

$$\begin{aligned} \alpha_j^{\text{Age Cat.}} \mid \sigma^{\text{Age Cat.}} & \stackrel{\text{ind.}}{\sim} N(0, (\sigma^{\text{Age Cat.}})^2), \text{ for } j = 1, \dots, J^{\text{Age Cat.}} \\ \sigma^{\text{Age Cat.}} & \sim N_+(0, 1) \end{aligned} \quad (8)$$

The *autoregressive specification* has the prior distributions:

$$\begin{aligned} \alpha_1^{\text{Age Cat.}} \mid \rho, \sigma^{\text{Age Cat.}} & \sim N(0, \frac{1}{1-\rho^2} (\sigma^{\text{Age Cat.}})^2) \\ \alpha_j^{\text{Age Cat.}} \mid \alpha_{j-1}^{\text{Age Cat.}}, \dots, \alpha_1^{\text{Age Cat.}}, \rho, \sigma^{\text{Age Cat.}} & \sim N(\rho \alpha_{j-1}^{\text{Age Cat.}}, (\sigma^{\text{Age Cat.}})^2), \text{ for } j = 2, \dots, J^{\text{Age Cat.}} \\ \sigma^{\text{Age Cat.}} & \sim N_+(0, 1) \\ (\rho + 1)/2 & \sim \text{Beta}(0.5, 0.5). \end{aligned} \quad (9)$$

The *random walk specification* has the prior distributions:

$$\begin{aligned} \alpha_j^{\text{Age Cat.}} \mid \alpha_{j-1}^{\text{Age Cat.}}, \dots, \alpha_1^{\text{Age Cat.}}, \sigma^{\text{Age Cat.}} & \sim N(\alpha_{j-1}^{\text{Age Cat.}}, (\sigma^{\text{Age Cat.}})^2), \text{ for } j = 2, \dots, J^{\text{Age Cat.}} \\ \sum_{j=1}^{J^{\text{Age Cat.}}} \alpha_j^{\text{Age Cat.}} & = 0. \end{aligned} \quad (10)$$

We treat age as an ordered categorical predictor. It is reasonable to believe that people of similar ages will have similar attitudes on same-sex marriage. Hence we propose autoregressive and random walk structures as the prior distributions for age category.

#### 5.4 Performing MRP with structured priors for the 2008 Annenberg phone survey

Hierarchical logistic regression with the two structured prior specifications and the baseline specification described previously are fit to the 2008 Annenberg phone survey. The poststratification matrix formed by the 5-year ACS is then used to poststratify posterior estimates for every age category. This is shown in Figure 5.

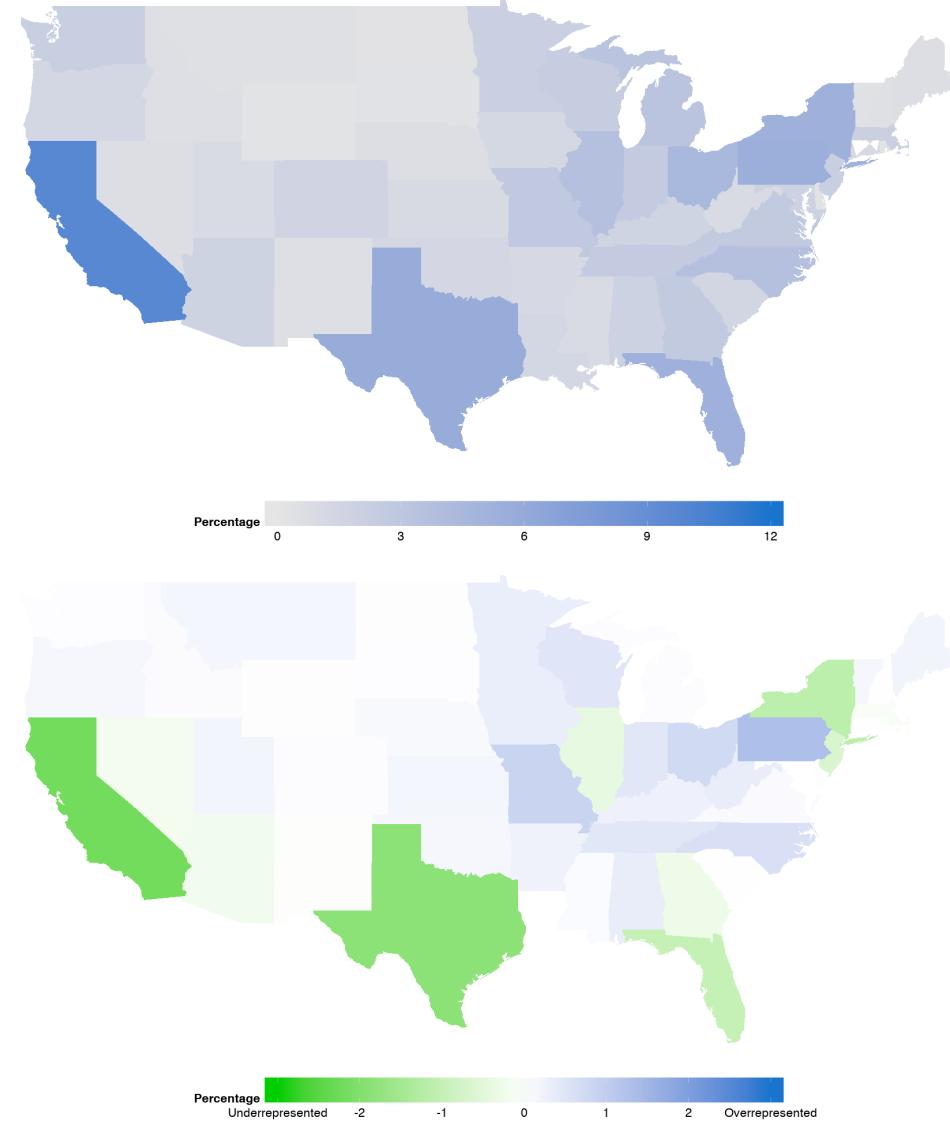
When age is discretized into 12 categories, there are no noticeable differences among the three prior specifications for age categories 1–11. Only at age category 12 do we start seeing a difference between the baseline specification and the two structured prior specifications. As expected, this difference in posteriors is observed when the underlying age category is a sparse cell for the survey data set.

When age is discretized into 48 and 72 categories, one starts to see differences between the structured prior specifications and the baseline specification in terms of posterior variance for every age category. Posterior variances for the baseline specification are wider based on the 5–95 percent quantiles, and they expand a significant amount for the oldest age categories. The baseline prior specification's posteriors become contracted towards their respective empirical means, which is not ideal since the empirical means swing more wildly for the older age categories. On the other hand, the autoregressive and random walk specifications are more smooth due to their property of having neighboring posterior random effects for age categories sharing information, and this is most noticeable when the number of age categories is 72. This smoothing effect is desirable for ordinal data as one may be interested in capturing a long-term trend when age increases.

What's also worth noting is that the baseline specification drastically changes the posterior variances when the number of categories for age changes from 12 to 48 to 72. Structured priors provide some stability in posterior variances despite how the input survey data is preprocessed through discretization of continuous variables.

The posterior population preferences for all three prior specifications remain nearly identical across the three age categories. This remains consistent with population preference results based the simulation studies.

Based on the simulation studies, we had shown that structured priors reduce absolute bias and posterior variances of structured covariates. In our application of structured priors to the non-representative 2008 Annenberg phone survey, we see that structured priors reduce posterior variance on the structured covariate age as well.



*Figure 4.* Percentage proportions of data in every state for the Annenberg survey (top) and Annenberg percentage proportions - ACS percentage proportions (bottom). The proportions in both surveys were rounded to two decimal places. A state with green hues in the bottom heatmap corresponds to the Annenberg survey underrepresenting that particular state. A state with blue hues in the bottom heatmap corresponds to the Annenberg survey overrepresenting that particular state.

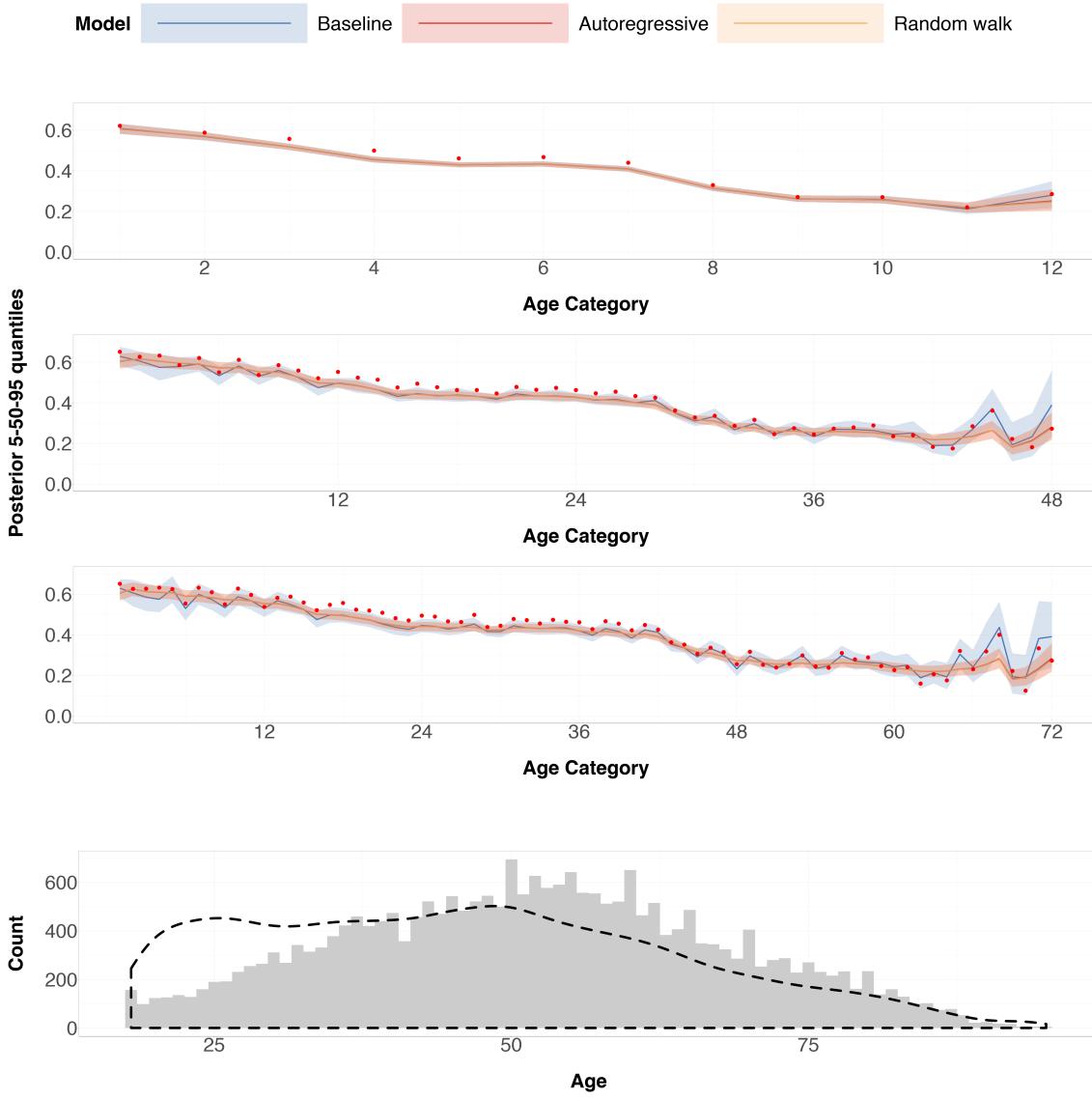


Figure 5. 12 (top), 48 (middle) and 72 (bottom) age categories. Red points in the top three plots are the empirical mean. The upper and lower bands in the top three plots correspond to the 95-percent and 5-percent posterior quantiles for every age category, and the middle solid line contains the posterior median for every age category. The density plot of ages in the ACS are coming from a random sample based off the 5-year ACS, where sampling is conducted with replacement using person weights given by the ACS. This random sample size is the same size as the 2008 Annenberg phone survey, and is assumed to be representative of the overall population defined the the 5-year ACS. 2000 iterations for 4 chains were run, for each prior specification and for age discretized into 12, 48 and 72 categories. The burn-in was set to 50 percent.

## 6 Conclusion

We proposed using priors that exploit underlying structure in the covariates of multilevel regression and poststratification. Defined as structured prior distributions, they aim to introduce more intelligent shrinkage of posterior estimates.

We show through simulation studies that structured priors, when compared to independent random effects reduce posterior MRP bias regardless of nonresponse pattern if there is an underlying pattern. A secondary benefit of structured priors when compared to independent random effects is that they reduce posterior variances for MRP estimates at the subpopulation levels corresponding to structured covariates of interest. We show that structured priors weather even extreme nonresponse patterns when compared to traditional random effects used in MRP. This is as expected since structured priors enable intelligent information-borrowing and shrinkage in posterior MRP estimates. Our modeling strategy of using structured priors was also applied to the non-representative 2008 Annenberg phone survey. The structured priors we describe here have similar smoothing properties to nonparametric regression methods such as GP regression and kernel smoothing (Gelman et al., 2013; Rasmussen, 2003).

Our investigations of using MRP for the Annenberg survey had used ACS data to its full capacity through the usage of a 5-year ACS that covered the year 2008. Using a 1-year or a 3-year ACS which would have resulted in rougher information about the population. Indeed, the information used to build the poststratification can be a limiting factor for MRP. The accuracy of poststratification in MRP is dependent on whether the poststratification matrix used is a true representation of the target population or not. Based on both simulation studies and analysis on the Annenberg survey, we saw that more age categories resulted in lower posterior variance and bias for age category estimates. This comes at the tradeoff of coarser information about  $N_j$ , the size of the poststratification cell  $l$ . Another limitation one may have is deciding covariates to impose structured priors on. This choice is dependent on the modeller's knowledge of the problem and the data used.

There is usually more than one set of structured priors to propose, and this model selection and comparison problem is not addressed in this paper. The method in the paper could also be extended to using structured priors on interaction terms (Ghitza & Gelman, 2013). Furthermore, we do not analyze the scenario when a structured prior is used for a covariate with no apparent structure.

In this manuscript we demonstrate improvements to MRP estimates through the use of structured priors when justified to do so. We believe that this is a contribution to the wider field considering other forms of regularization with MRP, but rather than employing black box methods, using structured priors exploits methodologist and survey administrator knowledge.

## 7 References

- Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 24(3), 179–195.
- Bisbee, J. (2019). *BARP: MRP - Multilevel + BART*. Retrieved from <https://github.com/jbisbee1/BARP/blob/master/vignettes/BARP.pdf>

- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017, 1). Stan : A probabilistic programming language. *Journal of Statistical Software*, 76(1). doi: 10.18637/jss.v076.i01
- Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11(4), 427–444.
- Downes, M., Gurrin, L. C., English, D. R., Pirkis, J., Currier, D., Spittal, M. J., & Carlin, J. B. (2018). Multilevel regression and poststratification: A modeling approach to estimating population quantities from highly selected survey samples. *American journal of epidemiology*, 187(8), 1780–1790.
- Elliott, M. R., Valliant, R., et al. (2017). Inference for nonprobability samples. *Statistical Science*, 32(2), 249–264.
- Gelman, A. (2018). Regularized prediction and poststratification: A generalization of Mister P. *Statistical Modeling, Causal Inference, and Social Science Blog*. Retrieved from <https://statmodeling.stat.columbia.edu/2018/05/19/regularized-prediction-poststratification-generalization-mister-p/>
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
- Gelman, A., Lax, J., Phillips, J., Gabry, J., & Trangucci, R. (2016). Using multilevel regression and poststratification to estimate dynamic public opinion. *Unpublished manuscript, Columbia University*.
- Gelman, A., & Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression.
- Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.
- Ghitza, Y., & Gelman, A. (2013). Deep interactions with MRP: Election turnout and voting patterns among small electoral subgroups. *American Journal of Political Science*, 57(3), 762–776.
- Goplerud, M., Kuriwaki, S., Ratkovic, M., & Tingley, D. (2018). Sparse multilevel regression (and poststratification [smrp]). *Unpublished manuscript, Harvard University*.
- Izrael, D., Battaglia, M. P., & Frankel, M. R. (2009). Extreme survey weight adjustment as a component of sample balancing (aka raking). In *Proceedings from the thirty-fourth annual sas users group international conference*.
- Kastellec, J. P., Lax, J. R., & Phillips, J. H. (2010). Estimating state public opinion with multi-level regression and poststratification using R. *Unpublished manuscript, Princeton University*.
- Lax, J. R., & Phillips, J. H. (2009). Gay rights in the states: Public opinion and policy responsiveness. *American Political Science Review*, 103(3), 367–386.
- Little, R. J. (1993). Post-stratification: a modeler's perspective. *Journal of the American Statistical Association*, 88(423), 1001–1012.
- Lohr, S. L. (2009). *Sampling: design and analysis*. Nelson Education.
- Morris, G. E. (2019). If everyone had voted, hillary clinton would probably be president. *Economist*. Retrieved from <https://www.economist.com/graphic-detail/2019/07/06/if-everyone-had-voted-hillary-clinton-would-probably-be-president>
- Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: state-level estimates from national polls. *Political Analysis*, 12(4), 375–385.
- Pfeffermann, D., et al. (2013). New important developments in small area estimation. *Statistical Science*, 28(1), 40–68.
- Rao, J. N. (2014). Small-area estimation. *Wiley StatsRef: Statistics Reference Online*, 1–8.
- Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer school on machine learning* (pp. 63–71).
- Rue, H., & Held, L. (2005). *Gaussian markov random fields: theory and applications*. Chapman and Hall/CRC.

- Si, Y., Trangucci, R., Gabry, J. S., & Gelman, A. (2017). Bayesian hierarchical weighting adjustment and survey inference. *arXiv preprint arXiv:1707.08220*.
- Simpson, D., Rue, H., Riebler, A., Martins, T. G., Sørbye, S. H., et al. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1), 1–28.
- Skinner, C., Wakefield, J., et al. (2017). Introduction to the design and analysis of complex survey data. *Statistical Science*, 32(2), 165–175.
- The Annenberg Public Policy Center of the University of Pennsylvania. (2008). *The Annenberg Public Policy Center's National Annenberg Election Survey 2008 Phone Edition (NAES08-Phone)* [Data file and code book]. Available from <https://www.annenbergpublicpolicycenter.org/tag/data-sets/>.
- Trangucci, R., Ali, I., Gelman, A., & Rivers, D. (2018). Voting patterns in 2016: Exploration using multilevel regression and poststratification (mrp) on pre-election polls. *arXiv preprint arXiv:1802.00842*.
- United States Census Bureau / American FactFinder. (2010). *2006-2010 ACS 5-year Public Use Microdata samples (PUMS) - CSV format* [Data file and code book]. Available from <https://factfinder.census.gov/>.
- Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3), 980–991.
- Zhang, X., Holt, J. B., Lu, H., Wheaton, A. G., Ford, E. S., Greenlund, K. J., & Croft, J. B. (2014). Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. *American journal of epidemiology*, 179(8), 1025–1033.

## Appendix

The various simulation conditions based on sample size  $n$  and true preference curve based on age of an individual is given in Table 4. Table 5 below summarizes posterior quantile differences for the three true preference curves when the probability of sampling index is perturbed.

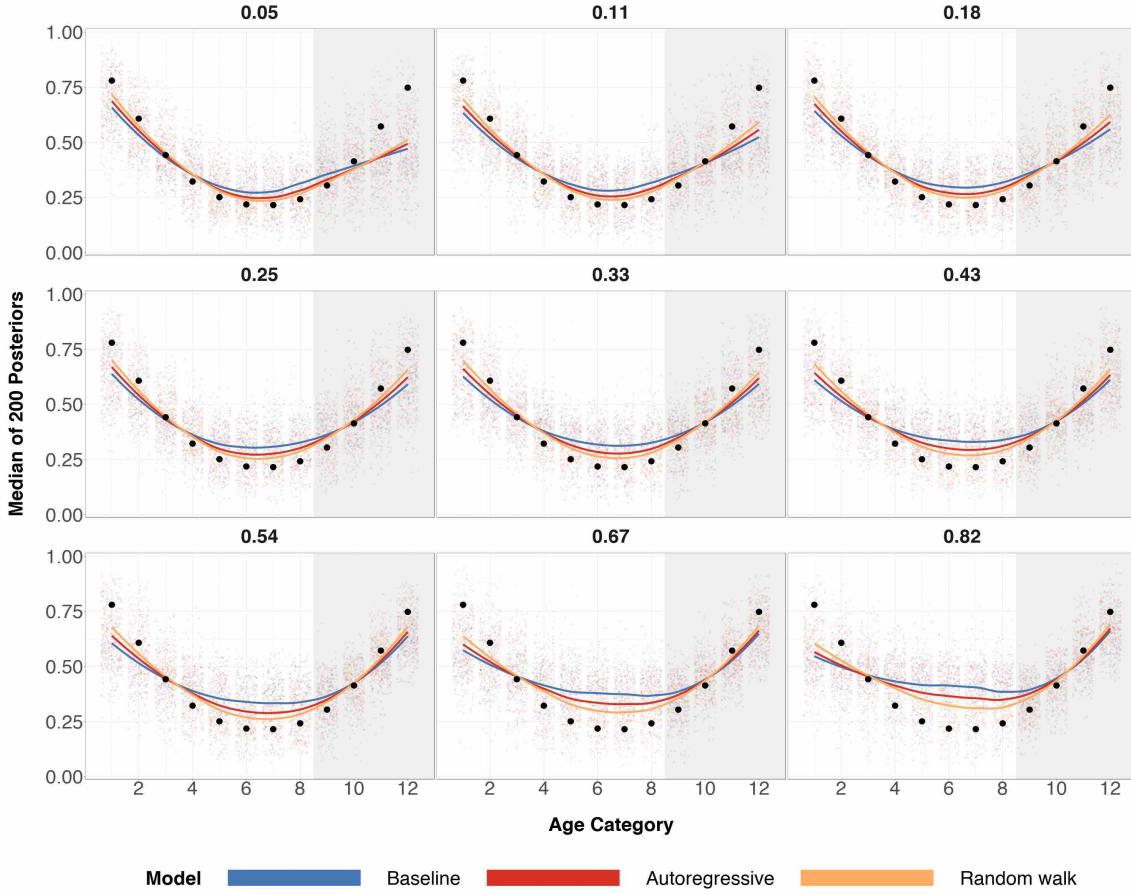
Sample size	Age preference curve	Poststrat. cell bias	Bias for each age category
100	U-shaped	Figure 6	Figure 3
500	U-shaped	Figure 7	Figure 3
100	Cap-shaped	Figure 10	Figure 12
500	Cap-shaped	Figure 11	Figure 12
100	Increasing	Figure 15	Figure 17
500	Increasing	Figure 16	Figure 17

Table 4  
Simulation scenarios for bias assessment

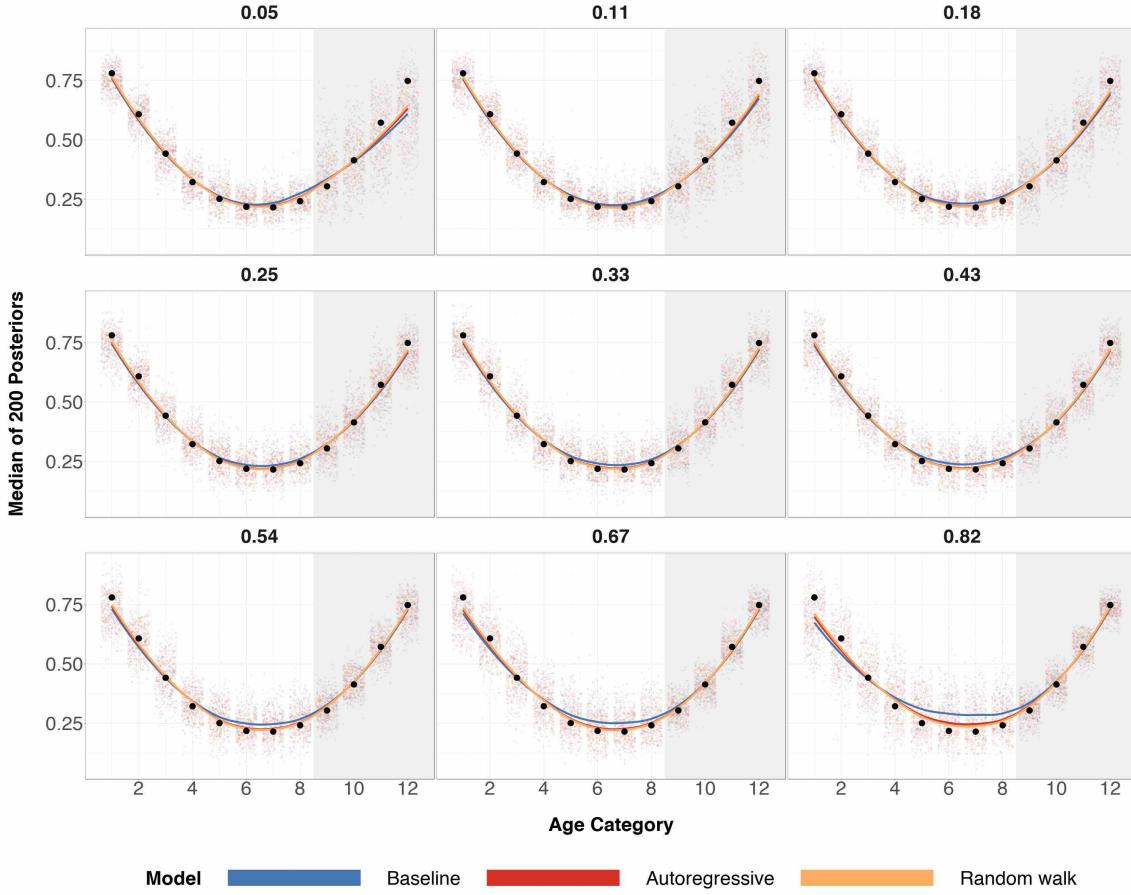
Sample size	Age preference curve	Difference in posterior quantiles
100	U-shaped	Figure 8
500	U-shaped	Figure 9
100	Cap-shaped	Figure 13
500	Cap-shaped	Figure 14
100	Increasing-shaped	Figure 18
500	Increasing-shaped	Figure 19

Table 5  
Simulation scenarios for posterior standard deviation assessment

In the simulation studies,  $X_{\text{Income}} = (0.1, 0, -0.2, 0.2)$ .  $X_{\text{State}}$  and  $X_{\text{Relig.}}$  are 51-length vectors that correspond to the 2004 Democratic vote share and 2004 percentage of Mormons + Evangelicals in every state respectively. These come from the data set used in Kastellec, Lax, and Phillips (2010). Finally,  $\beta^0 = 0$  if the true preference curve is increasing and  $-1.5$  otherwise.  $\beta^{\text{State}} = 0.5$  and  $\beta^{\text{Relig.}} = -0.5$ .



*Figure 6.* Posterior medians for 200 simulations for each age group, where true age preference is U-shaped and sample size  $n = 100$ . Black circles are true preference probabilities for each age group. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9-12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.



*Figure 7.* Posterior medians for 200 simulations for each age group, where true age preference is U-shaped and sample size  $n = 500$ . Black circles are true preference probabilities for each age group. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9-12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.

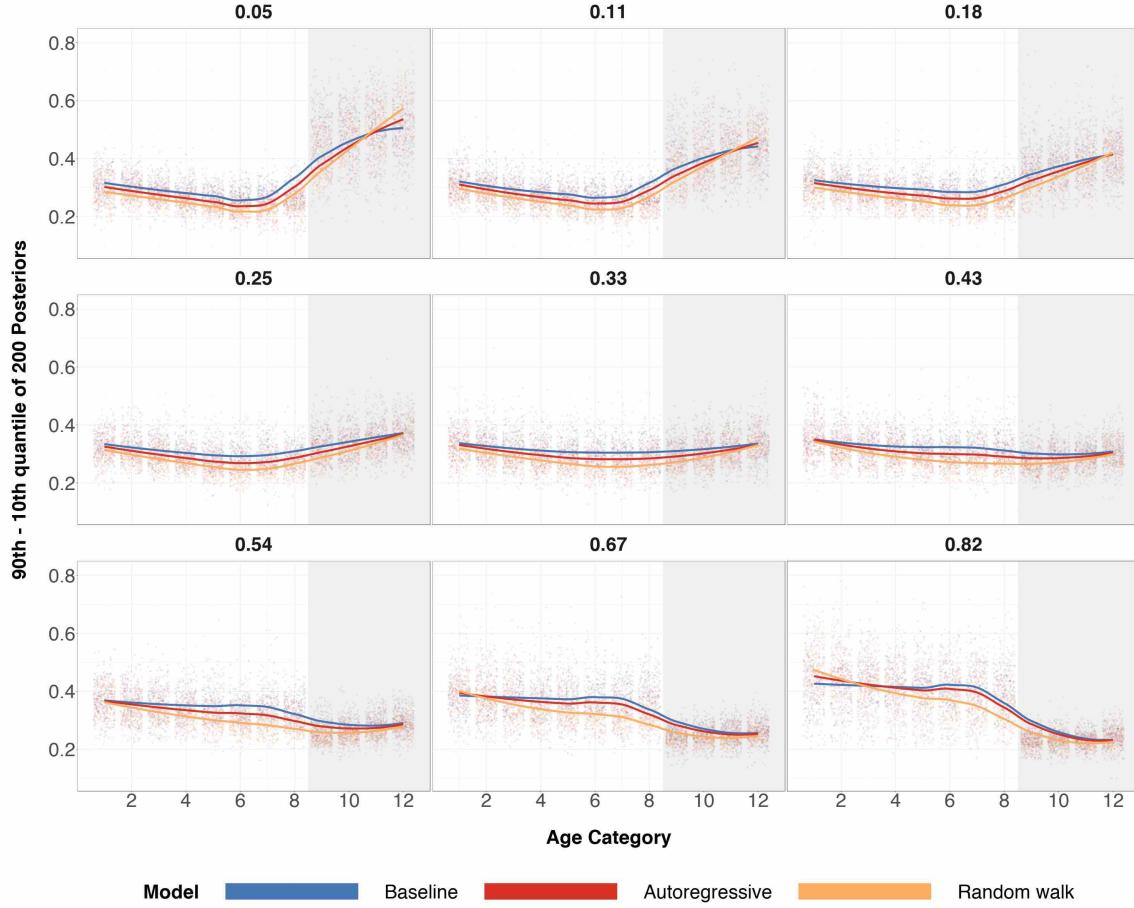


Figure 8. Differences in the 90<sup>th</sup> and 10<sup>th</sup> posterior quantiles for every age category when true age preference is U-shaped and  $n = 100$  for 200 simulations. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9-12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.

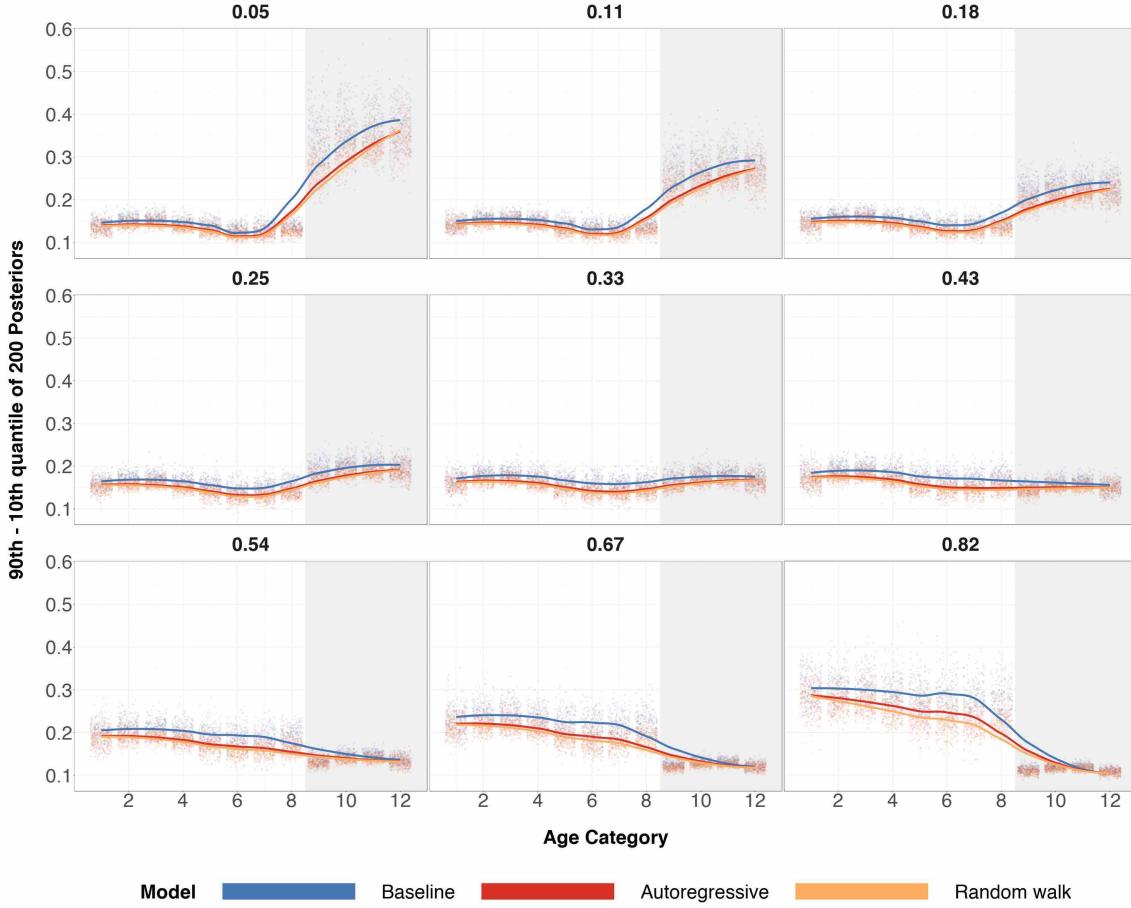
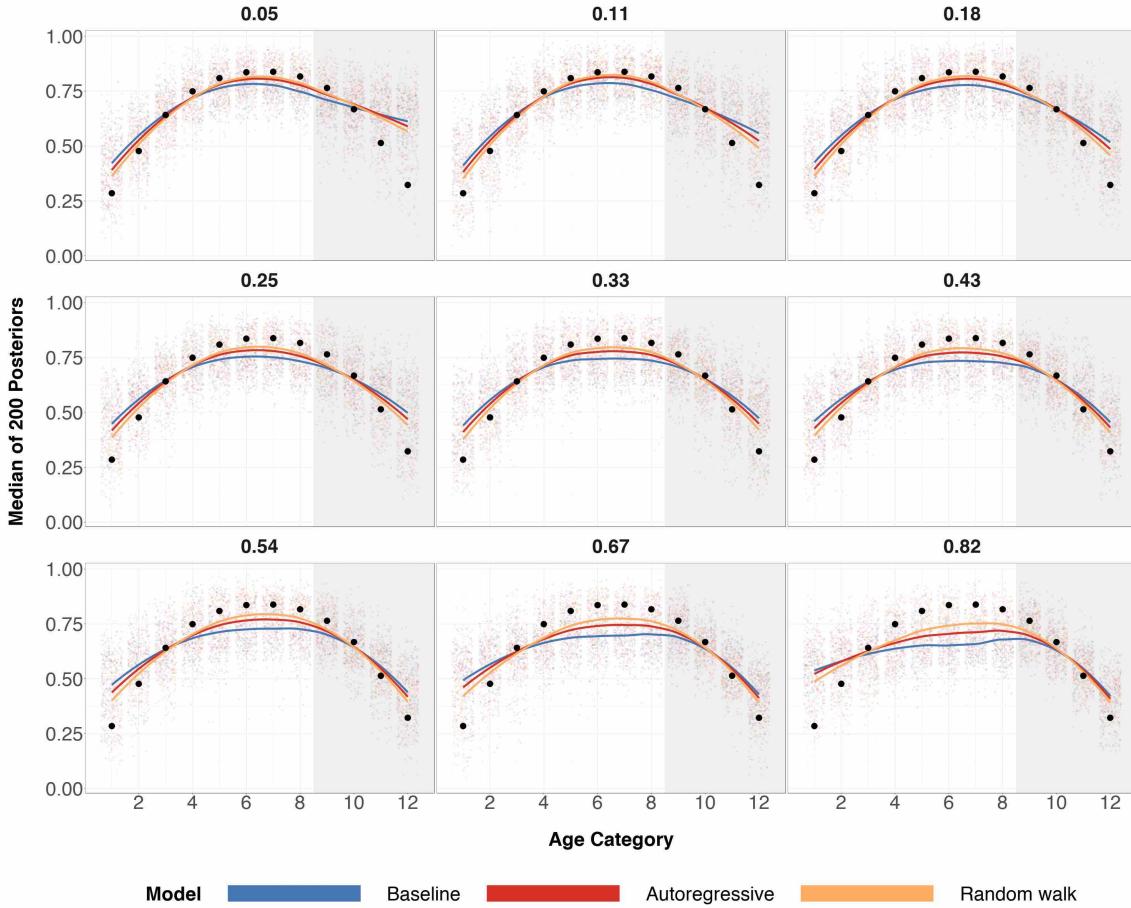
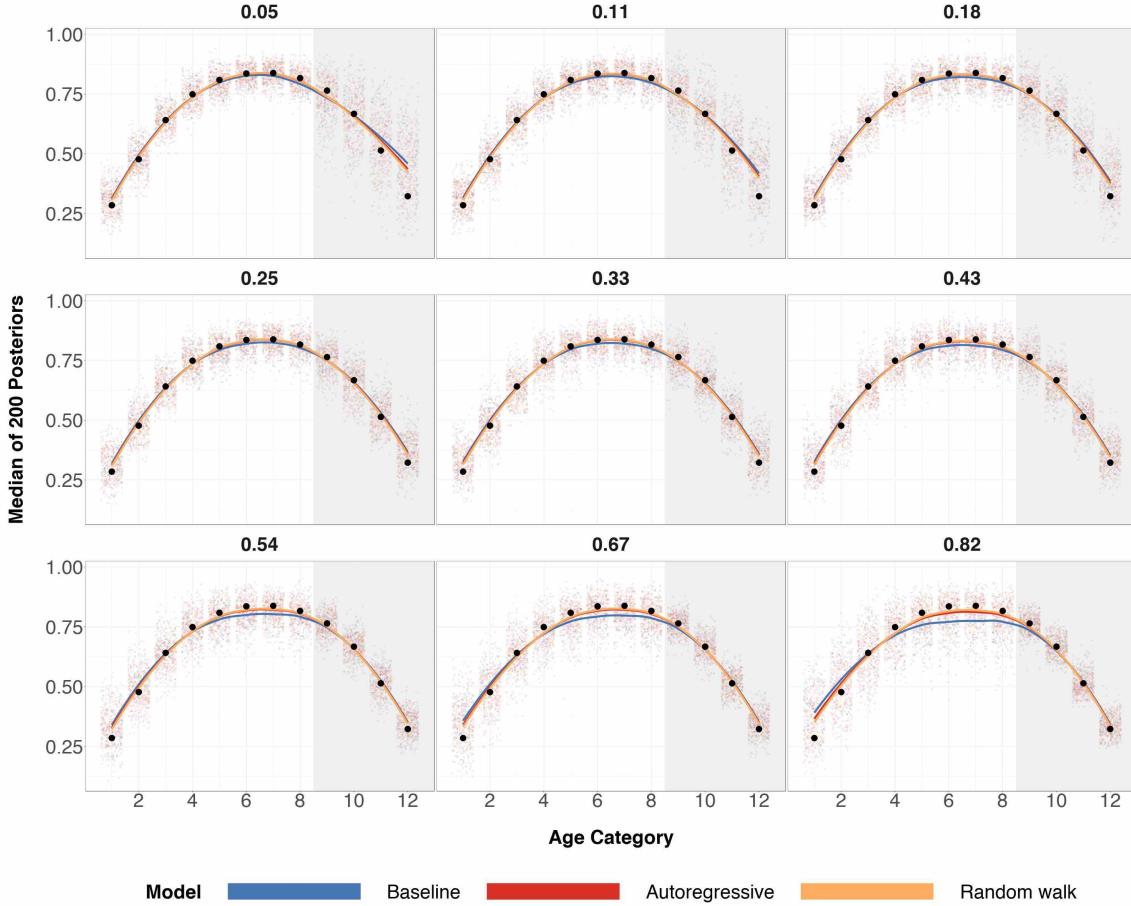


Figure 9. Differences in the 90<sup>th</sup> and 10<sup>th</sup> posterior quantiles for every age category when true age preference is U-shaped and  $n = 500$  for 200 simulations. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9-12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.



*Figure 10.* Posterior medians for 200 simulations for each age group, where true age preference is cap-shaped and sample size  $n = 100$ . Black circles are true preference probabilities for each age group. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9-12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.



*Figure 11.* Posterior medians for 200 simulations for each age group, where true age preference is cap-shaped and sample size  $n = 500$ . Black circles are true preference probabilities for each age group. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9–12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.

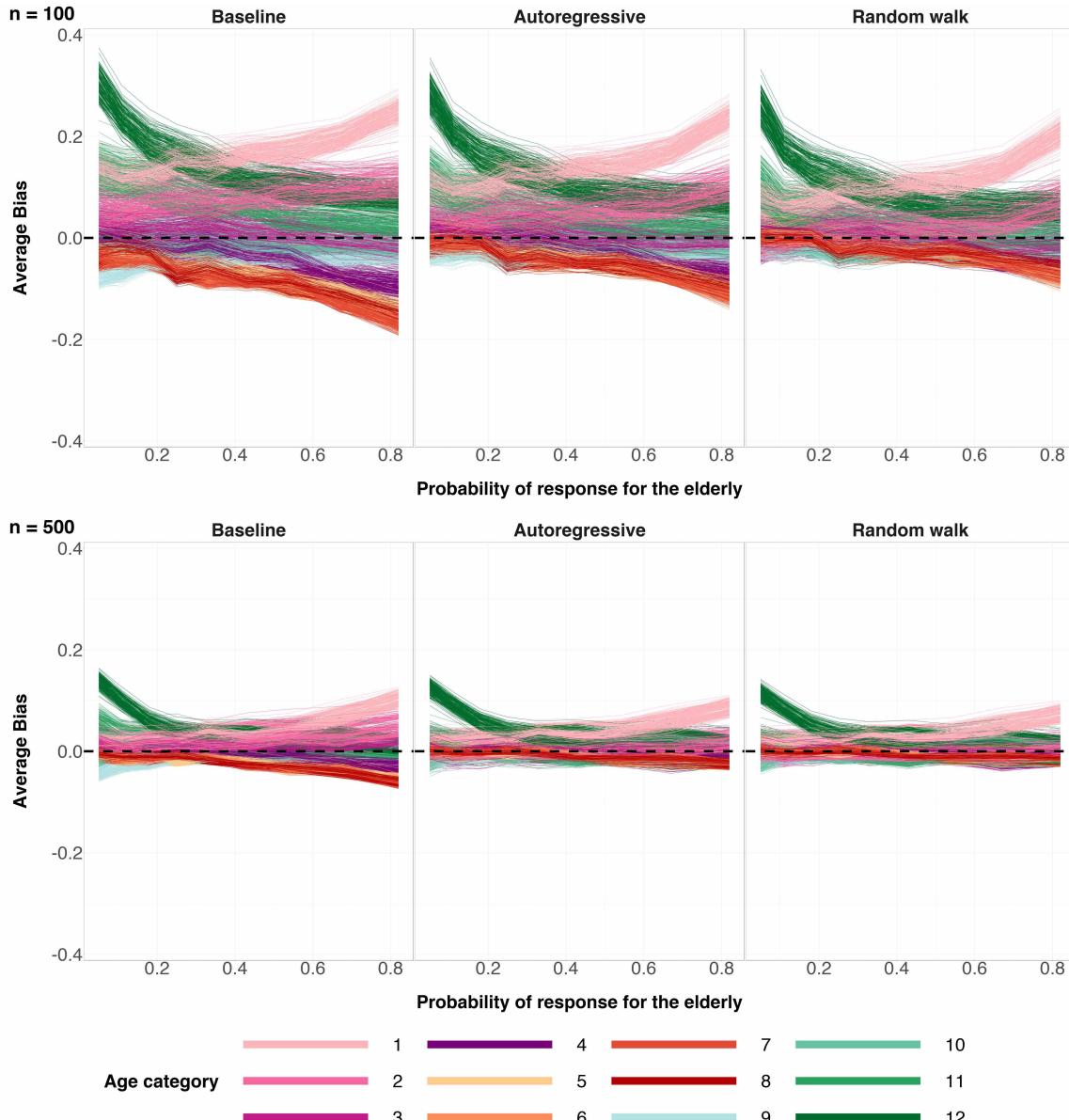


Figure 12. The average of bias values coming from 200 simulations of posterior medians of the 2448 poststratification cells. Sample size  $n = 100$  (top) and  $n = 500$  (bottom). The true preference curve for age is cap-shaped. The horizontal dashed line at  $y = 0$  represents zero bias.

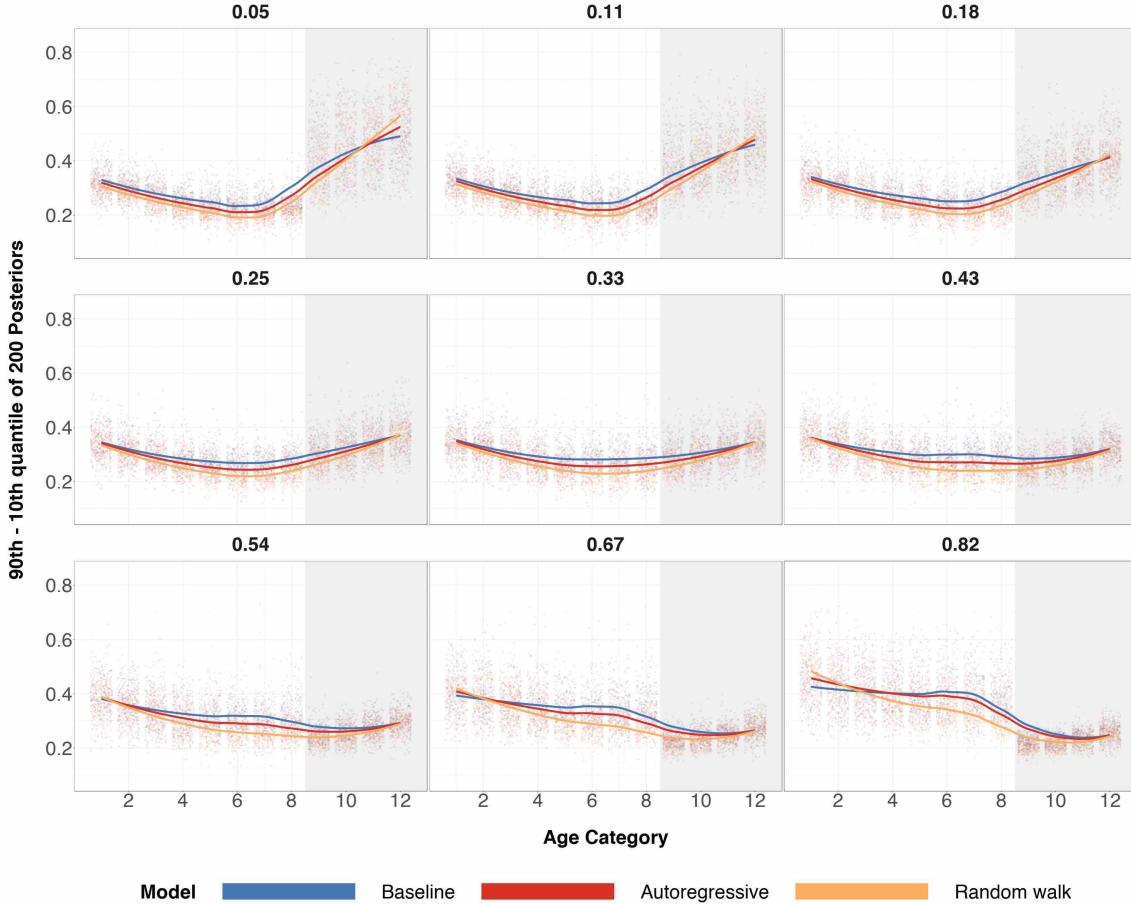
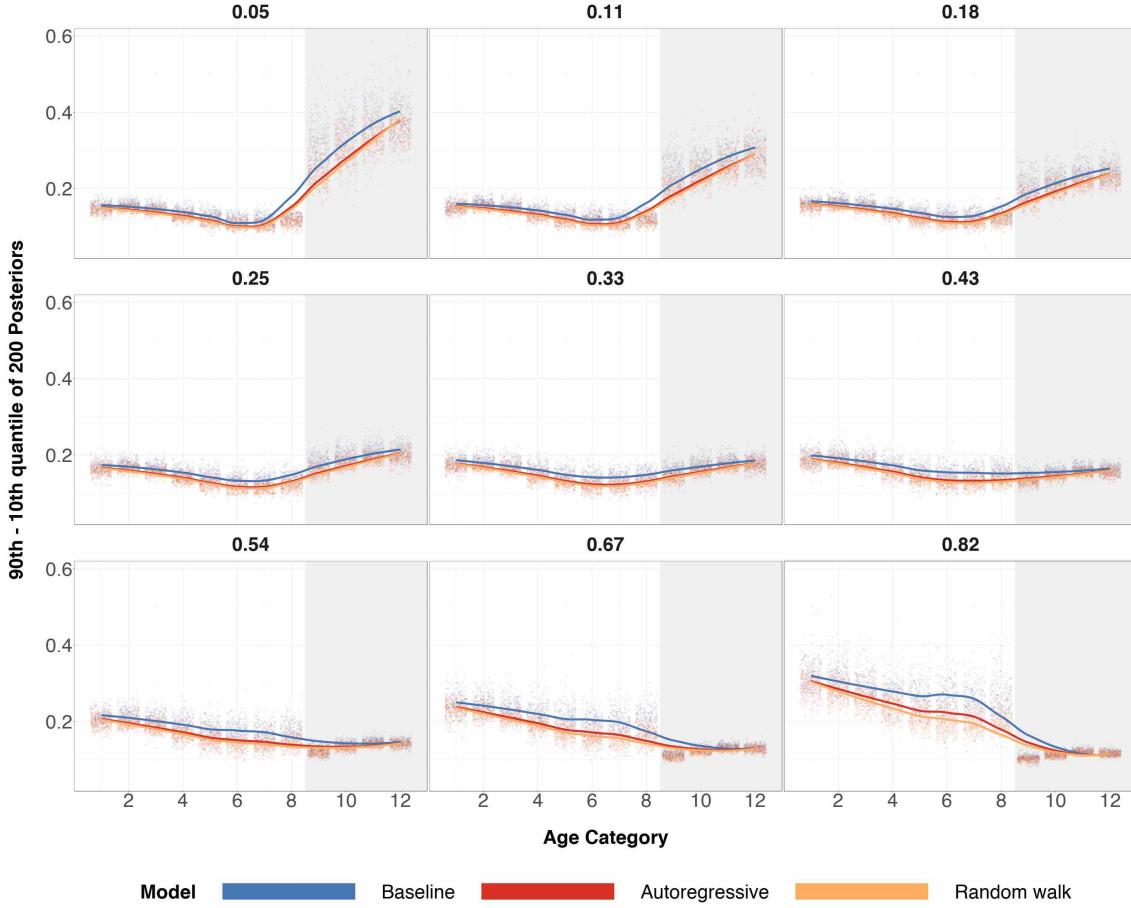


Figure 13. Differences in the 90<sup>th</sup> and 10<sup>th</sup> posterior quantiles for every age category when true age preference is cap-shaped and  $n = 100$  for 200 simulations. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9–12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.



*Figure 14.* Differences in the 90<sup>th</sup> and 10<sup>th</sup> posterior quantiles for every age category, when true age preference is cap-shaped and  $n = 500$  for 200 simulations. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9–12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.

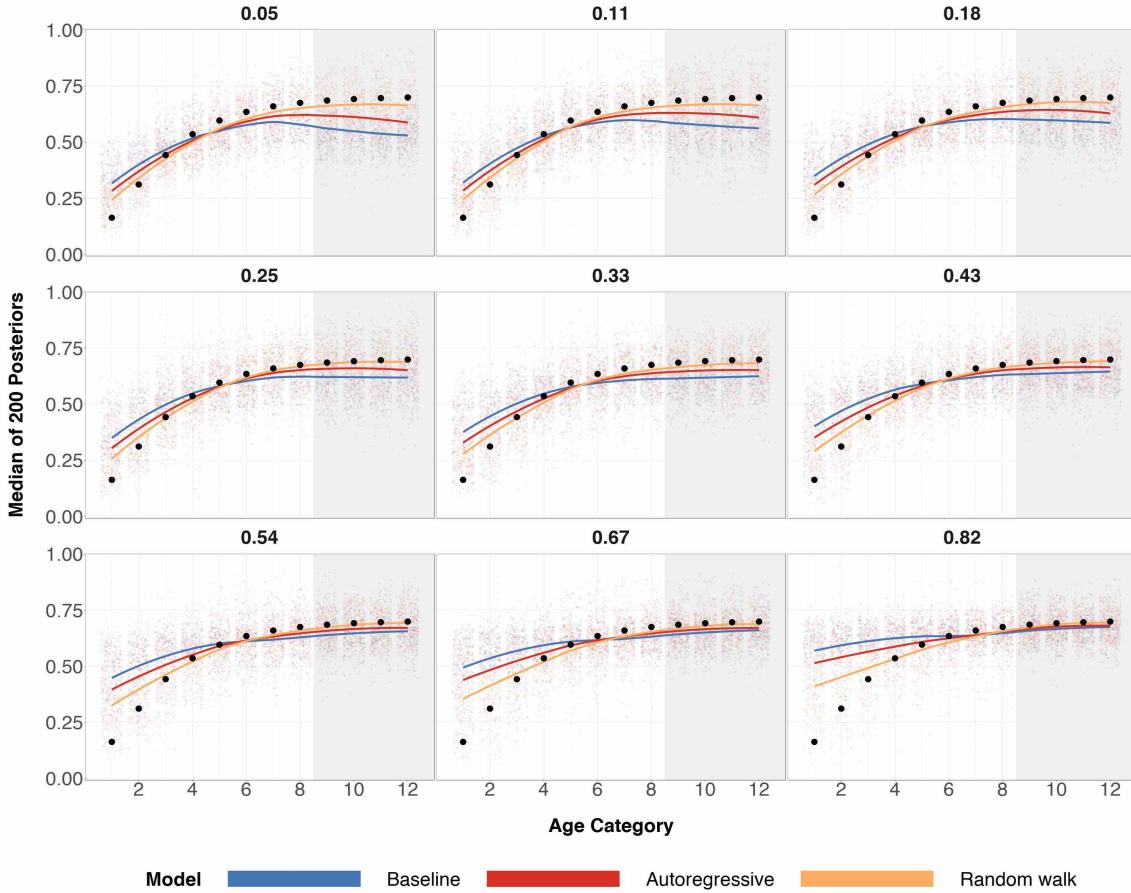
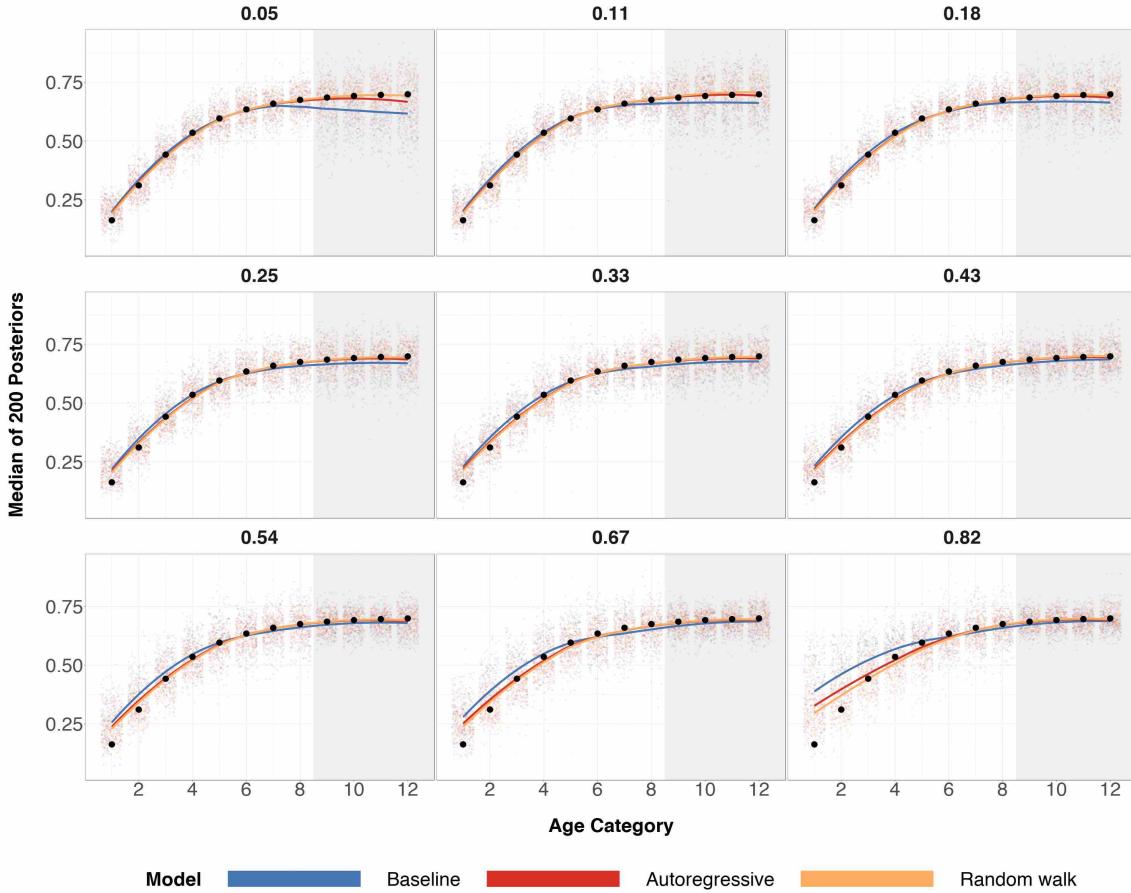
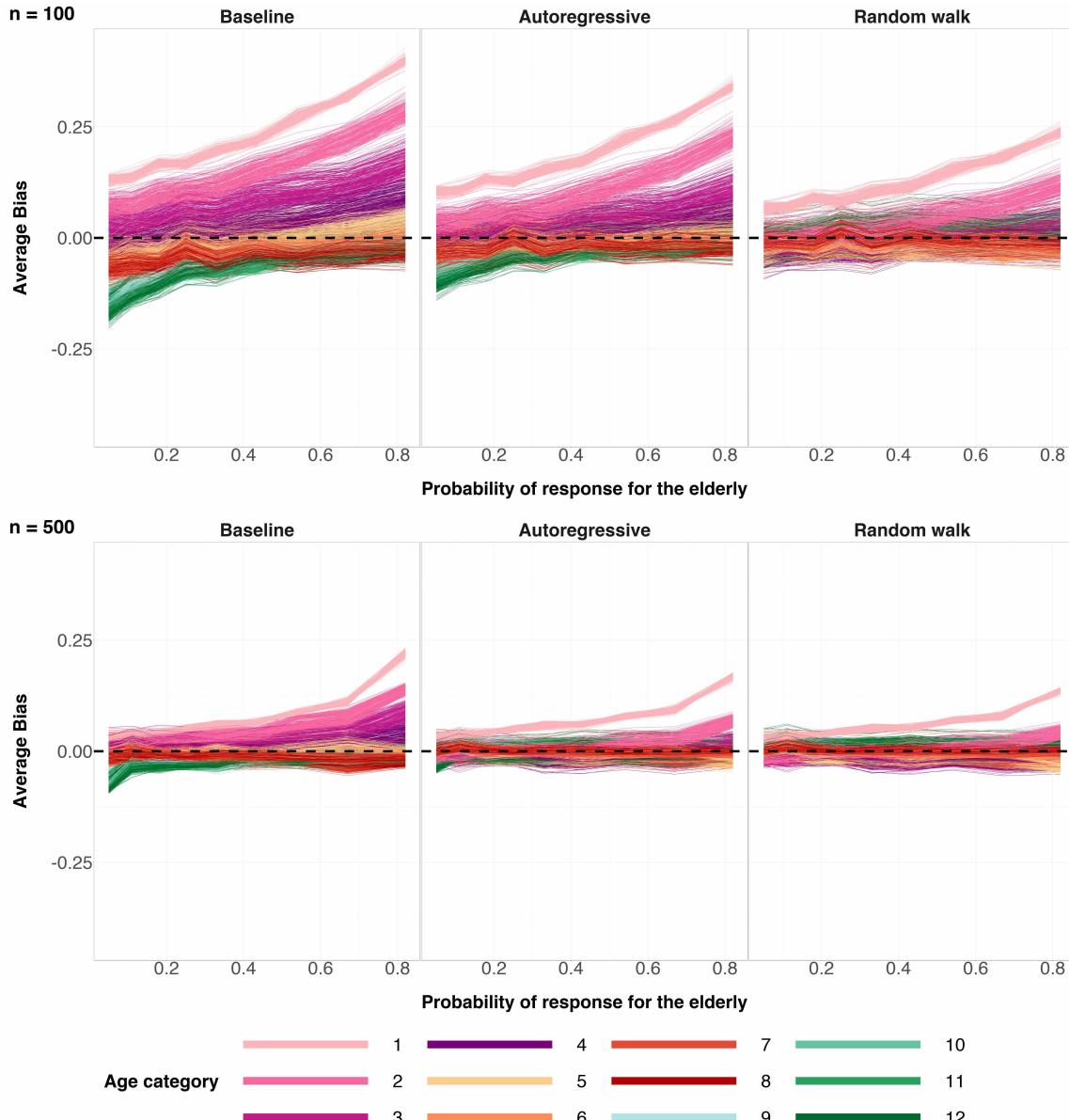


Figure 15. Posterior medians for 200 simulations for each age group, where true age preference is increasing-shaped and sample size  $n = 100$ . Black circles are true preference probabilities for each age group. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9–12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories 9–12. Local regression is used for the smoothed estimates amongst the three prior specifications.



*Figure 16.* Posterior medians for 200 simulations for each age group, where true age preference is increasing-shaped and sample size  $n = 500$ . Black circles are true preference probabilities for each age group. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9–12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling *and* representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.



*Figure 17.* The average of bias values coming from 200 simulations of posterior medians of the 2448 poststratification cells. Sample size  $n = 100$  (top) and  $n = 500$  (bottom). The true preference curve for age is increasing-shaped. The horizontal dashed line at  $y = 0$  represents zero bias.

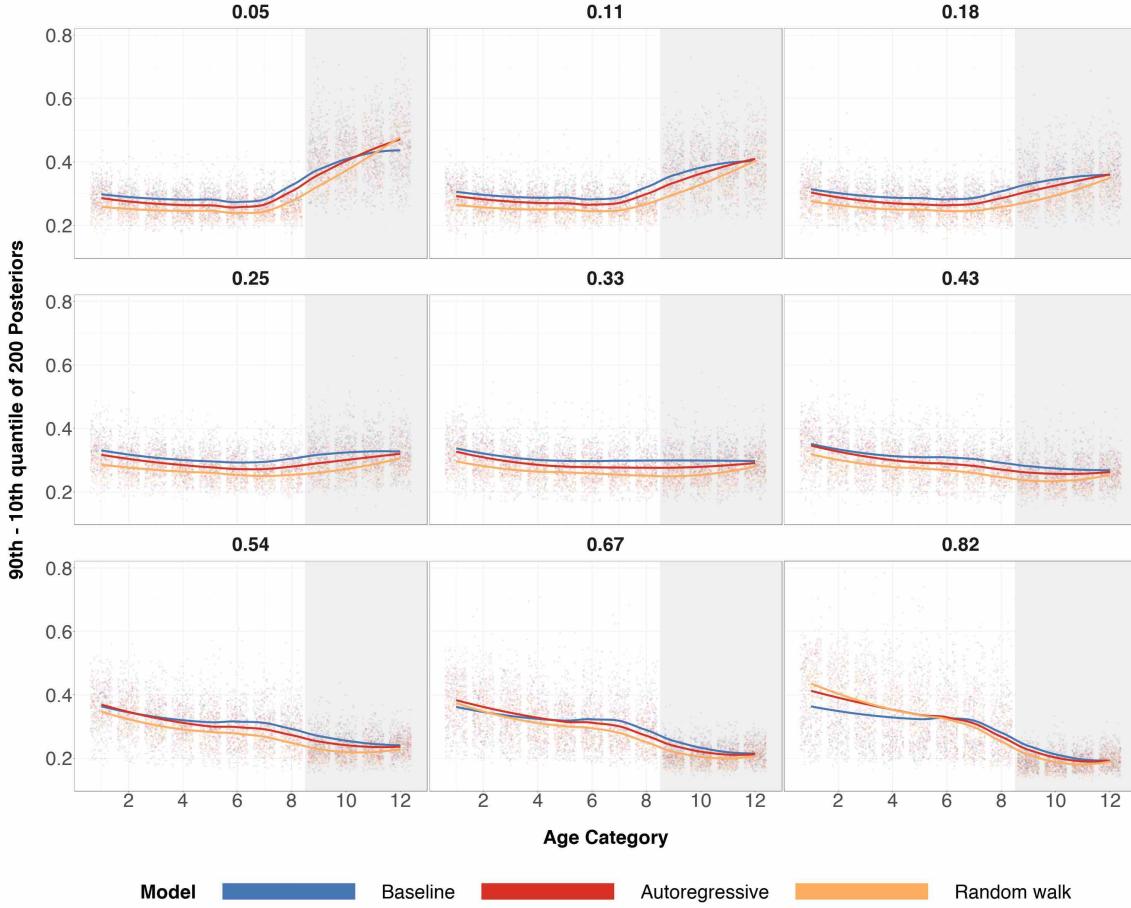


Figure 18. Differences in the 90<sup>th</sup> and 10<sup>th</sup> posterior quantiles for every age category when true age preference is increasing-shaped and  $n = 100$  for 200 simulations. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9–12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.

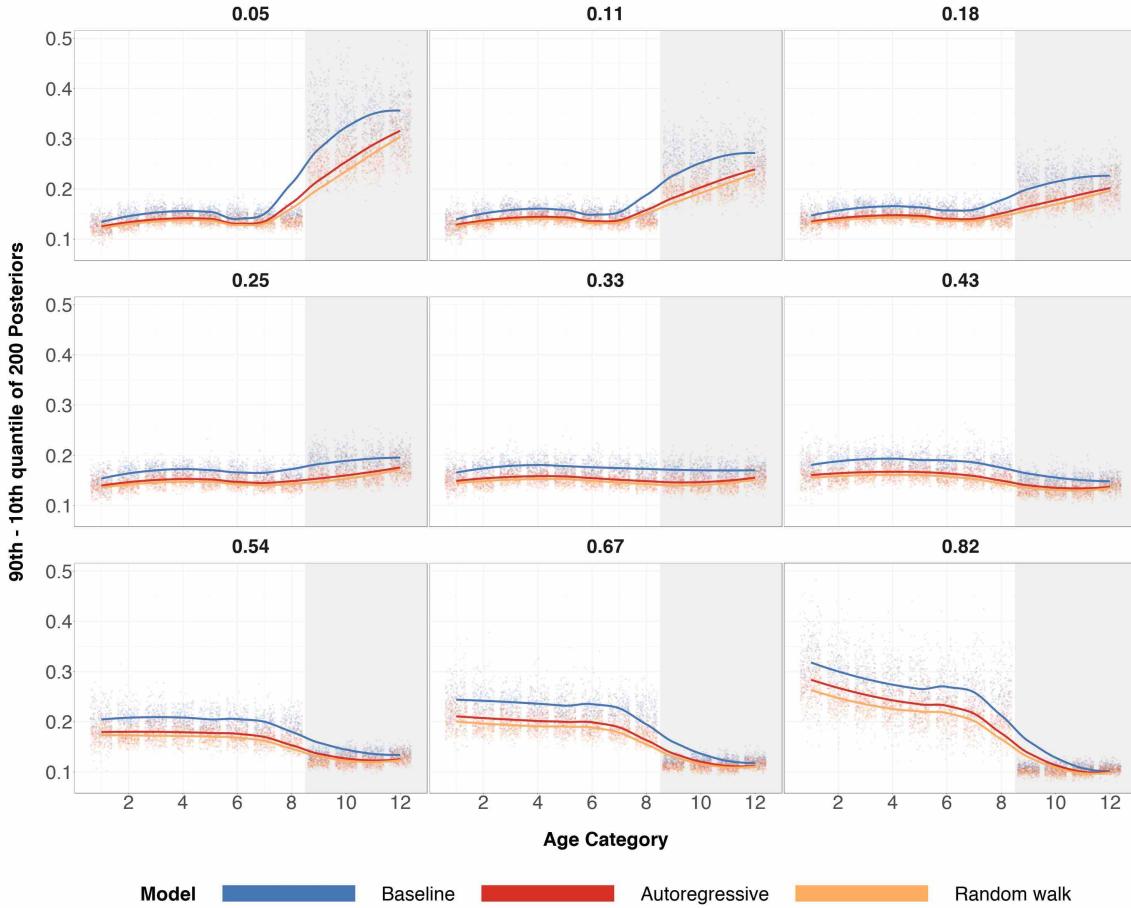


Figure 19. Differences in the 90<sup>th</sup> and 10<sup>th</sup> posterior quantiles for every age category when true age preference is increasing-shaped and  $n = 500$  for 200 simulations. The numerical index for the 9 plots correspond to the expected proportion of the sample that are older adults (also known as the probability of sampling the subpopulation group with age categories 9–12). The shaded gray region corresponds to the age categories of older individuals for which we over/under sample. The center of the grid represents completely random sampling and representative sampling for age categories. Local regression is used for the smoothed estimates amongst the three prior specifications.