Surveys designed to collect data on similar variables using samples representing the same population may still result in different estimates due, for example, to differences in sample designs or modes of data collection. Considered in this paper is the case where two surveys were conducted concurrently, with one using the same methodology as used in prior rounds of the survey and the other using an updated methodology, resulting in substantial differences in several key estimates. Due to differences in sample size, only the latter survey was detailed enough for disaggregated-level estimates of publishable quality. We propose a hierarchical model to account for discrepancies in the estimates from the two surveys and a Bayesian approach for producing reliable estimates at various levels of aggregation. The model relies on a common latent structure at the disaggregated level to allow “bridging” between the two surveys. The methodology is applied to the 2016 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation and the 2016 50-State Surveys of Fishing, Hunting, and Wildlife-Associated Recreation. Aligning these two surveys is critical to extend the series of related statistics that have been published since 1955, allowing for meaningful comparisons over time despite the change in survey methodology.

1. Introduction. Many surveys are conducted repeatedly over time, with the goal of gaining understanding of the current state of the population at each survey time point as well as the population’s changes over time. The National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (FHWAR) is a prime example of such a survey, going back to 1955 and with directly comparable results since 1991; see U.S. Department of the Interior, Fish and Wildlife Service and U.S. Department of Commerce, Census Bureau (2018).

An important consideration in ensuring that results remain comparable over time is that consistent methodologies are applied at all steps of the survey process each time: sampling, data collection and processing, and statistical adjustments such as imputation and weighting. However, underlying changes in the population can cause methodologies that were previously appropriate and effective to gradually degrade over time. Two such changes affecting the National Survey of FHWAR (and numerous other surveys) are the increasing nonresponse rates and growing complexities of reaching the general public through traditional survey modes. Hence, ensuring that surveys continue to be conducted in a rigorous manner requires that methodologies occasionally be adapted or replaced. The introduction of such changes often has significant impact on the comparability of new and historical estimates, with changes in data collection mode found to be particularly problematic (Olson et al., 2020). However, if side-by-side data are collected, as was done for the 2016 National Survey of FHWAR, it is...
possible to develop bridging procedures that restore the ability to make meaningful comparisons over time.

Recently, van den Brakel, Zhang and Tam (2020) published a review of statistical methods on measuring discontinuities in time series obtained with repeated sample surveys. In this paper, we are concerned with reducing such discontinuities. Particularly, we address one critical step in bridging two 2016 national surveys of FHWAR. For this, we construct survey estimates using data publicly available from the two surveys. We propose a hierarchical model to account for various sources of error in the data available from the two surveys and a Bayesian approach for producing reliable estimates at various levels of aggregation. The model relies on a common latent structure at the disaggregated level to allow for formal bridging between the two surveys and to provide appropriate measures of uncertainty. The resulting bridged model predictions reduce the discontinuity observed in the survey estimates over time.

While our approach is innovative for the given problem, it has similarities with studies concerned with estimation or forecasting. For example, Raghunathan et al. (2007) combine data from two surveys using a hierarchical Bayes approach to construct county-level estimates of prevalence of cancer risk factors and screening. The authors had sufficient information to produce survey-based estimates at the county level using data from both sources, one of which was considered the gold standard. Also for estimation purposes, Lohr and Brick (2012) study blending methods for domain estimates from two victimization surveys when one survey may be biased relative to the other. A crop yield forecasting problem has been addressed in Balgobin, Berg and Barboza (2013), also from the point of view of combining data from multiple surveys. A hierarchical Bayes approach was developed to produce state-level survey forecasts by reconciling state-level forecasts available from two surveys at multiple time points to state-level estimates based on a third survey available at the end of the crop season; the latter survey was considered to be the gold standard. In contrast to these studies, we do not assume either of the two surveys is a gold standard and we overcome the challenge of reconciling survey-based estimates produced using data from two surveys at distinct levels of aggregation.

For the current application, the survey-based estimators constructed using data from the two surveys each contain various sources of non-sampling error. Because of biases due to the non-sampling errors and because of the different survey designs, the pairs of survey estimators have different expectations for the true, underlying, target population quantity, though they are unbiased with respect to the corresponding sample designs. For the purpose of ensuring consistency with FHWAR estimates from previous years, we assume that the statistics from the 2016 National Survey of FHWAR are unbiased for the population target of interest. Informally, we say this target is in the “currency” of the 2016 National Survey, and our inferential goal is to combine the data from both surveys to produce disaggregated-level statistics in the 2016 National Survey currency. The symmetry of our proposed bridge allows for either survey to serve as the reference; disaggregated-level model estimates of improved precision, compared to the survey estimates, are produced in the currency of the other survey, too.

The rest of the paper is as follows. In Section 2, we introduce the motivation for this work, describe the data available for the application study, and present selected results motivating the need for bridging. Survey estimation steps and initial data investigation results are provided in Section 3. The hierarchical bridging model is presented in Section 4, along with a Bayesian framework for model fit and prediction. Model validation methods are described in Section 5. Results for bridging fishing participation totals (of interest in the motivating study) are provided throughout the paper. Alternative specifications for the hierarchical bridging model are provided in Section 6. A discussion of practical challenges, model extensions, and concluding remarks is given in Section 7.
2. Application Background. In 2016, national and 50-state versions of the FHWAR study were fielded concurrently, specifically to allow for comparison between the two and for subsequent reconciliation of the estimates. The national version was conducted by the U.S. Census Bureau, using methods employed in previous FHWAR studies. The 50-state version was conducted by the Rockville Institute (RI), the nonprofit affiliate of Westat. For brevity, we will refer to the Census study as the “National Survey” and the RI study as the “50-State Survey.” Both versions used generally approved survey approaches of high quality, including probability sampling of addresses, prescreening, multiple waves of data collection, etc. Their samples were designed to be representative of the U.S. population living in households. They identified likely sportspersons and wildlife watchers within sampled households, and individual participation surveys covered all of calendar year 2016. Design-based weights were constructed by both Census Bureau and RI for the respective surveys, together with associated replicate weights for variance estimation.

On the other hand, the two FHWAR surveys differed in several aspects of the survey implementations. This included mode of data collection (the National Survey data were collected in person and by telephone and the 50-State Survey data were collected by mail), sample design and allocation (for example, the National Survey was geographically clustered and the 50-State Survey was not), and sample selection process (the 50-State Survey included a sample of likely non-participants for follow-up, the National Survey did not). The two FHWAR surveys used questionnaires as nearly identical as possible, given this difference in mode. See U.S. Department of the Interior, Fish and Wildlife Service and U.S. Department of Commerce, Census Bureau (2018) and Rockville Institute (2018) for more details about the two surveys.

Because of these differences between the two surveys, their estimates differ despite nominally targeting the same population quantities. The 50-State Survey estimates almost invariably turned out larger, often considerably larger than the National Survey estimates. It is not possible to determine which of the two targets is closer to the true measurement of the population. Therefore, the best one can do is to treat both population targets as imperfect proxies, which are nevertheless useful for evaluating time trends and relative differences between estimates within a survey.

The goal of this paper is two-fold. First, we consider the problem of converting the participation estimates from the 50-State Survey into the “currency” of the National Survey, allowing comparisons with prior rounds of the survey. Second, we address the need for reliable statistics at a finer level of geography (states) than the one at which National Survey estimates are available (census divisions). We will focus here on the estimation of fishing participation, one of the key quantities of interest in the FHWAR surveys.

The data sources for the application study are the public use files containing record-level National Survey data and record-level 50-State Survey data. Final and replicate survey weights are available in these files, constructed under two distinct variance replication methods adopted by Census and RI, for the respective surveys. Using these data, we construct fishing participation estimates and associated standard errors, measured in the two currencies. State-level, census division-level, and nation-level estimates are constructed using the 50-State Survey data. The National Survey was not designed to support state-level estimates and the public use files do not contain state indicators, so only census division-level and nation-level estimates are constructed.

3. Survey Estimation. Let \(i = 1, \ldots, 9\) index census divisions and \(j = 1, \ldots, n_i\) index the \(n_i\) states within census division \(i\). For a given geography \(g\) (state, census division, or nation), let \(s_g\) be the survey respondents residing in \(g\). Then, the National Survey estimators
where the indicator $I(y_k^C = 1) = 1$ if respondent $k$ reports being a participant in fishing activities in the Census survey and 0 otherwise; $Y_{gi}$ is constructed similarly to $Y_g^C$, but with replicate weights $w_{k,r}^C$, replacing $w_k^C$; and $r = 1, \ldots, 160$ indexes the replicate constructed using the successive difference replication method; see U.S. Department of the Interior, Fish and Wildlife Service and U.S. Department of Commerce, Census Bureau (2018) and Ash (2014).

Similarly, let the 50-State Survey estimators be

$$Y_{gR} = \frac{\sum_{k \in s_g} w_{kR} I(y_k^{\text{RI}} = 1)}{\sum_{k \in s_g} w_{kR}}$$

and

$$\text{Var}(Y_{gR}) = \sum_{r=1}^{160} (Y_{gR} - Y_{gR}^r)^2 \frac{159}{160},$$

where $Y_{gR}$ is constructed similarly to $Y_{gR}^C$, but with replicate weights $w_{kR}^C$ replacing $w_k^C$; and $r = 1, \ldots, 160$ indexes the replicate constructed using the delete-1 Jackknife replication method; see Rockville Institute (2018) and Kott (2001).

Using the National Survey data, we obtain the participation rate estimates and their estimated variances for census division $i$, denoted $(Y_i^C, \text{Var}(Y_i^C))$, where the participation rate is defined as the percentage of participants in the population (people age 16 and older). Similarly, using the 50-State Survey data, we obtain the participation rate estimates and their estimated variances for state $j$ of census division $i$, denoted $(Y_{ij}^C, \text{Var}(Y_{ij}^C))$. The sampling variances will be fixed at their estimated values and treated as known in this analysis, as is common in model-based small area estimation literature, specifically when area-level models are developed; see, for example, the pioneering paper for area-level small area estimation models, Fay and Herriot (1979). For fishing participation estimates, the realized sample sizes in the National Survey survey were 1257 at the nation level and ranged from 35 to 327 at the census division level.

Using the 50-State Survey data, we construct state population totals as the sum of the final weights for records in state $j$ of census division $i$, $N_{ij} := \sum_{k \in s_g} w_{kR}$, where $g$ is the state and $k$ is an index for all responding units in a state. The 50-State Survey weights were calibrated to official Census totals for the population of people age 16 and older. However, the National Survey weights were calibrated to Census totals internally available to the Census Bureau for the population of people age 16 and older. As a consequence, the census division population totals constructed using the 50-State Survey weights differ from the corresponding totals constructed using the National Survey weights, i.e. $\sum_{k \in s_g} w_{kR} \neq \sum_{k \in s_g} w_k^C$, where $g$ is the census division. Nevertheless, lacking state-level information from the National Survey, we assume that the state population totals $N_{ij}$, as constructed using the 50-State Survey data, are fixed and known. For fishing participation estimates, the realized sample sizes in the 50-State Survey survey were 3709 at the nation level, ranging from 250 to 626 at the census division level, and ranging from 5 to 139 at the state level.

The discrepancies between the fishing participation rates from the National Survey and the 50-State Survey are illustrated in Figure 1. The plot on the left shows point estimates and uncertainty intervals from both surveys on the vertical axis versus census divisions on the horizontal axis. The plot on the right shows 50-State Survey estimates on the vertical axis versus National Survey estimates on the horizontal axis, with a one-to-one line and a fitted simple linear regression line included for reference. It is clear that the 50-State Survey participation rate estimates are larger than the corresponding National Survey estimates. Moreover, there
is a reasonably strong linear relationship between the estimates in the two currencies. The ob-
servation farthest away from the regression line corresponds to census division 7, for which
both the National Survey and the 50-State Survey estimates have high estimated variances.
To mitigate the discrepancies in these estimates, we developed a bridging model assuming
a linear relationship between the true state-level participation rates for the two surveys; see
Section 4.

![Census Division-Level Statistics for Fishing Participation Rates](image)

**Fig 1.** Census division-level survey estimates for fishing participation rates. Left: point estimates and uncertainty intervals for both surveys versus census division. Right: 50-State Survey estimates versus National Survey estimates, with one-to-one line and fitted simple linear regression line for reference.

4. **Bridging Model.** We propose a hierarchical model in which the National Survey cen-
sus division estimators are unbiased for the underlying census division population targets in
“National Survey currency” while the 50-State Survey state estimators are unbiased for the
underlying state population targets in “50-State Survey currency”; this is reflected in the first
two hierarchical levels corresponding to the sampling models for both sets of survey esti-
mates. The quantities of interest are the state population targets (fishing participation rates)
in “National Survey currency.” For these, we assume a bridging model that links the (latent)
state population targets in “National Survey currency” to the state population targets in “50-
State Survey currency” via a linear “currency conversion.” The final level of the hierarchical
model is a smoothing model for the state population targets in “50-State Survey currency”
that borrows information across two nested geographical levels, states and census divisions.
Therefore, the model is specified for the data at the lowest level of geographic availability:
state for the 50-State Survey data and census division for the National Survey data.

Using the sampling levels, we account for the different state-level and census division-
level sampling errors in the survey estimates, as a result of the two complex survey designs
adopted for the National Survey and the 50-State Survey, respectively. We do not assume one
of the surveys provides gold standard estimates. Rather we acknowledge that, at the census
division level, we have two independent measurements of the same quantity, each subject to
different error, and we account for it using the two independent sampling levels.
The bridging level is assumed at the state level, hence allowing us to construct state-level estimates based on the 50-State Survey data but expressed in the same currency as the estimates from the National Survey. The two independent measurements we observe for the same quantity of interest, in different units of measurement, are linked via the bridging model. In contrast, for an application with the same units of measurement, this bridging model would be simply the identity link. The specification of the bridging model is symmetric, allowing for either specifying the National Survey quantity of interest as a function of the 50-State Survey quantity of interest, or vice-versa.

The sampling and bridging levels provide the solution to our two-fold problem: different currencies and different levels of aggregation. To further improve the precision of the state-level estimates, a smoothing level is included to borrow information across all the states with available 50-State Survey estimates. This last level is similar to the second level in a hierarchical small area estimation model, where it is often referred to as the “linking level” because it is used to link the underlying true quantities of interest with covariates. In the smoothing model, we use a common grand mean for all the states, and state-level and census division-level effects; see, for example, Erciulescu, Cruze and Nandram (2020). The resulting smoothing model, we use a common grand mean for all the states, and state-level and census division-level effects; see, for example, Erciulescu, Cruze and Nandram (2020). The resulting smoothed, state-level estimates from the smoothing model represent the 50-State Survey data used in the bridging model; in other words, we improve the state-level estimates and bridge the 50-State Survey estimates to the National Survey estimates simultaneously.

The proposed hierarchical model is as follows:

\[ Y_i^C(\mu_i^C, \text{Var}(Y_i^C)) \sim N(\mu_i^C, \text{Var}(Y_i^C)) \]  
(National Survey census division-level sampling model)

\[ Y_{ij}^{RI}(\mu_{ij}^{RI}, \text{Var}(Y_{ij}^{RI})) \sim N(\mu_{ij}^{RI}, \text{Var}(Y_{ij}^{RI})) \]  
(50-State Survey state-level sampling model)

\[ \mu_i^C = \gamma_0 + \gamma_1 \mu_{ij}^{RI} \]  
(bridging state-level model)

\[ \mu_i^C = \left( \sum_{j=1}^{n_i} N_{ij} \right)^{-1} \sum_{j=1}^{n_i} N_{ij} \mu_{ij}^C \]  
(identity between census division-level and state-level, quantities of interest; i.e. census division and state participation rates in the National Survey currency)

\[ \mu_{ij}^{RI} = \mu + u_i + s_{ij} \]  
(50-State Survey smoothing state-level model)

\[ u_i \sim N(0, \sigma_u^2), s_{ij} \sim N(0, \sigma_s^2). \]

The true state-level participation rates in the 50-State Survey currency are denoted by \( \mu_{ij}^{RI} \), the true state-level participation rates in the National Survey currency are denoted by \( \mu_{ij}^C \) and the true census division-level participation rates in the National Survey currency are denoted by \( \mu_i^C \). From the surveys, we observe imprecise measurements of \( \mu_{ij}^{RI} \) and \( \mu_i^C \), as denoted by \( Y_{ij}^{RI} \) and \( Y_i^C \), with associated variances \( \text{Var}(Y_{ij}^{RI}) \) and \( \text{Var}(Y_i^C) \), respectively. However, the interest is in \( \mu_i^C \) and \( \mu_{ij}^C \), given that the National Survey currency allows for comparability with prior National Survey estimates, maintaining the time trend. As previously specified, we further improve the precision of the state-level estimates by borrowing information from neighboring states within the census division (via \( s_{ij} \)) and from the census divisions across the nation (via \( u_i \)).

The specified bridging model is an error-free linear transformation between the underlying state-level quantities of interest, measured in the two currencies. Moreover, the smoothing model is a composition of shrinkage effects, to a grand mean (over all of the states), to an average of state-level quantities, and to an average of census division-level quantities. Hence, flipping the bridging and the smoothing models to \( \mu_{ij}^{RI} = \gamma_0^{*} + \gamma_1^{*} \mu_{ij}^C \) and \( \mu_{ij}^C = \mu^{*} + u_i^{*} + s_{ij}^{*} \), respectively, would result in an equivalent model specification.
To complete the model specification, we adopted the following independent, weakly-informative priors for the model parameters: $\gamma_0 \sim N(0, 10^4)$, $\gamma_1 \sim N(0, 10^4)$, $\mu \sim N(0, 10^4)$, $\sigma_d \sim \text{half-Cauchy}(0, 5)$, $\sigma_s \sim \text{half-Cauchy}(0, 5)$. We fit the model in R STAN (Stan Development Team, 2020) using Markov chain Monte Carlo (MCMC) with three chains. Each chain consists of 200,000 draws with the first 50,000 as burn-in and with the remaining 150,000 thinned to every 10th draw to reduce dependence. The remaining $R = 45,000$ draws from the three chains combined are used to approximate the joint posterior distribution of all unknown parameters.

Let $\zeta = 1, \ldots, R$ index the MC draws. Then, state-level bridged predictions may be constructed using the MC samples $\mu_{ijC,\zeta}$. Census division-level bridged predictions may be constructed using the MC samples $\mu_{iC,\zeta}$, where

$$
\mu_{iC,\zeta} = \left( \sum_{j=1}^{n_i} N_{ij} \right)^{-1} \sum_{j=1}^{n_i} N_{ij} \mu_{ijC,\zeta}.
$$

Nation-level bridged predictions may be similarly constructed using the MC samples $\mu_{C,\zeta}$, where $\mu_{C,\zeta}$ are defined as aggregations of the state-level MC samples $\mu_{ijC,\zeta}$ over all the states and census divisions in the nation. Similarly, other posterior statistics of interest, such as credible intervals, can be obtained from the MCMC samples.

5. Model Validation. For internal model validation, we considered mixing and convergence diagnostics for the MCMC sampler, residual diagnostics for the normality assumptions, and posterior predictive checks for the normality and linearity assumptions, and for other characteristics of the data such as the correlation between the two sets of census division-level estimates. For the posterior predictive checks, we used discrepancy measures comparing predictions simulated from the posterior predictive distribution for the participation rates against sample participation rates (50-State Survey or National Survey survey direct estimates) or against the model predictions:

- the indicator comparing the survey direct estimates and the corresponding predictions simulated from the posterior predictive distributions for the participation rates: state-level for 50-State Survey and census division-level for National Survey;
- the difference between the model predictions and the corresponding predictions simulated from the posterior predictive distributions for the participation rates: state-level for 50-State Survey and census division-level for National Survey;
- the difference between the survey direct estimates and the corresponding predictions simulated from the posterior predictive distributions for the participation rates: state-level for 50-State Survey and census division-level for National Survey;
- the correlation between the census division-level National Survey and 50-State Survey survey direct estimates and the corresponding correlation between the predictions simulated from the posterior predictive distributions for the participation rates.

Analysis results for the final model showed that convergence and mixing of the chains has been achieved, so that the resulting posterior draws are valid for inference. For example, the multiple potential scale reduction factors were all less than 1.1 and the autocorrelations at lag 1 and the cross-correlations were close to 0 (Gelman and Rubin, 1992). Moreover, the effective numbers of independent simulation draws were close to 45,000 and the ratios of MC standard error to posterior standard deviation were smaller than 0.53%, so that the posterior summaries are sufficiently accurate. Overall, there were no significant departures from normality for standardized residuals, and no substantial indication of model lack of fit using the posterior predictive checks listed above.
External model validation checks were based on comparisons of survey estimates $Y_{ij}^{RI}, Y_i^C$ to model predictions $\hat{\mu}_{ij}^{RI}, \hat{\mu}_{ij}^C$, and $\hat{\mu}_i^C$ (posterior means of $\mu_{ij}^{RI}, \mu_{ij}^C$, and $\mu_i^C$, respectively). In addition, state-level estimates $Y_{ij}^C$ were made available by the Census Bureau for four states with sufficient sample size to meet Census Bureau precision requirements. We compared the model predictions $\hat{\mu}_{ij}^C$ against these four National Survey estimates, which were not used as part of the model fitting.

The plots in Figure 2 illustrate the relationship between the model predictions and the survey fishing participation rate estimates, for both the 50-State and National Surveys. This figure comprises six plots arranged in three rows, corresponding to the model predictions $\hat{\mu}_{ij}^{RI}, \hat{\mu}_{ij}^C$, and $\hat{\mu}_i^C$ on the vertical axes, and two columns, corresponding to the point estimates and associated standard errors, respectively. Recall that $\hat{\mu}_{ij}^{RI}$ are the state-level model predictions in the 50-State Survey currency, and $\hat{\mu}_{ij}^C$ and $\hat{\mu}_i^C$ are the state-level and census division-level model predictions in the National Survey currency, respectively. The model predictions $\hat{\mu}_{ij}^{RI}$ represent a smoothed version of the state-level 50-State Survey estimates and are expected to be closer to the 50-State Survey estimates for states with small sampling variances $\text{Var}(Y_{ij}^{RI})$ than for states with large sampling variances $\text{Var}(Y_{ij}^{RI})$.

In Figure 2, the survey estimates are given on the horizontal axes: National Survey estimates in the first row of plots, and 50-State Survey estimates in the last two rows of plots. Because comparisons against the National Survey estimates are only possible at the census division level, the first row of plots shows census division-level comparisons and the last two rows of plots show state-level comparisons. The census division-level National Survey fishing participation rate estimates and corresponding model predictions in the National Survey currency are close to the 45-degree line, as illustrated in the plot in the first row and first column of Figure 2. Therefore, the large positive discrepancies observed in the survey participation data are now mitigated by the model, as the model predictions are produced in the National Survey currency.

The state-level 50-State Survey fishing participation rate estimates and corresponding model predictions in the 50-State Survey currency are close to the 45-degree line, as illustrated in the plot in the second row and first column of Figure 2. There is modest smoothing observed in the survey participation data. The state-level 50-State Survey fishing participation rate estimates and corresponding model predictions in the National Survey currency are below the 45-degree line, as illustrated in the plot in the third row and first column of Figure 2. Therefore, state-level 50-State Survey participation rate estimates are reconstructed in the National Survey currency; since large positive discrepancies were observed between the 50-State Survey participation estimates and the corresponding National Survey estimates, the observed positive discrepancies between the 50-State Survey participation estimates and the corresponding model predictions are in the expected direction.
Fig 2. Model predictions versus survey estimates (left column) and model posterior standard errors versus estimated survey standard errors (right column), for fishing participation rates.
Based on the observations from the three plots in the second column of Figure 2, we conclude that most of the model prediction standard errors are smaller than the corresponding survey standard errors. These results are the effects of overall model bridging and smoothing.

To compare the model predictions to the survey estimates, we re-construct the left-hand-side plot in Figure 1, and include the census-division model predictions in the National Survey currency, \( \hat{\mu}_C \), and in the 50-State Survey currency, \( \hat{\mu}_{RI} \), with their associated uncertainty intervals defined as plus or minus two posterior standard errors. The results are shown in Figure 3. Note the improved agreement between the National Survey estimates and the bridging model predictions in the nation currency, at the census division level. Roughly speaking, Figure 3 can be interpreted as follows: from 50-State Survey direct estimate (black dot) to 50-State Survey modeled estimate (purple square) is mostly the effect of the smoothing model, while from 50-State Survey modeled estimate (purple square) to National Survey modeled estimate (blue diamond) is mostly the effect of the bridging model, a linear transformation.

Comparisons between the state-level model predictions and the additional four National Survey state-level estimates (for Maine, Minnesota, Oklahoma, and Virginia) are presented in Table 1, for all fishing participation rates. The four National Survey state-level estimates are a special tabulation made available by the Census Bureau (no state-level estimates can be constructed using the Census public use files because there are no state indicators). These additional estimates are not part of the National Survey data used to fit the model, and we use them for external validation only.

The four numerical columns in Table 1 correspond to point estimates, associated standard errors and associated lower and upper bounds of the 95 percent confidence/credible intervals (CIs), as applicable: for the survey estimates, 95 percent confidence intervals were approximated by adding and subtracting 2 standard errors from the point estimates, and 95 percent...
credible intervals were constructed for the model predictions using the corresponding 2.5 percent and the 97.5 percent quantiles of the posterior distribution. The four rows for each of the four states correspond to the 50-State Survey estimate, the model prediction in the 50-State Survey currency, the model prediction in the National Survey currency, and the National Survey estimate.

<table>
<thead>
<tr>
<th>State</th>
<th>Estimator</th>
<th>Point Estimate</th>
<th>Standard Error</th>
<th>95%CI_LB</th>
<th>95%CI_UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>$Y_{RI}^{ME}$</td>
<td>28.6</td>
<td>4.5</td>
<td>19.7</td>
<td>37.5</td>
</tr>
<tr>
<td></td>
<td>$\hat{\mu}_{RI}^{ME}$</td>
<td>24.1</td>
<td>3.4</td>
<td>17.6</td>
<td>31.0</td>
</tr>
<tr>
<td></td>
<td>$\hat{\mu}_{C}^{ME}$</td>
<td>16.3</td>
<td>3.0</td>
<td>11.0</td>
<td>23.0</td>
</tr>
<tr>
<td></td>
<td>$Y_{C}^{ME}$</td>
<td>21.1</td>
<td>2.9</td>
<td>15.4</td>
<td>26.8</td>
</tr>
<tr>
<td>MN</td>
<td>$Y_{RI}^{MN}$</td>
<td>25.4</td>
<td>3.3</td>
<td>19.0</td>
<td>31.8</td>
</tr>
<tr>
<td></td>
<td>$\hat{\mu}_{RI}^{MN}$</td>
<td>25.3</td>
<td>2.7</td>
<td>20.0</td>
<td>30.5</td>
</tr>
<tr>
<td></td>
<td>$\hat{\mu}_{C}^{MN}$</td>
<td>17.1</td>
<td>2.5</td>
<td>12.7</td>
<td>22.6</td>
</tr>
<tr>
<td></td>
<td>$Y_{C}^{MN}$</td>
<td>24.2</td>
<td>3.2</td>
<td>17.8</td>
<td>30.5</td>
</tr>
<tr>
<td>OK</td>
<td>$Y_{RI}^{OK}$</td>
<td>31.0</td>
<td>4.0</td>
<td>23.1</td>
<td>38.8</td>
</tr>
<tr>
<td></td>
<td>$\hat{\mu}_{RI}^{OK}$</td>
<td>28.9</td>
<td>3.3</td>
<td>22.5</td>
<td>35.3</td>
</tr>
<tr>
<td></td>
<td>$\hat{\mu}_{C}^{OK}$</td>
<td>20.0</td>
<td>3.2</td>
<td>14.5</td>
<td>27.2</td>
</tr>
<tr>
<td></td>
<td>$Y_{C}^{OK}$</td>
<td>16.9</td>
<td>1.9</td>
<td>13.2</td>
<td>20.6</td>
</tr>
<tr>
<td>VA</td>
<td>$Y_{RI}^{VA}$</td>
<td>24.1</td>
<td>3.5</td>
<td>17.2</td>
<td>30.9</td>
</tr>
<tr>
<td></td>
<td>$\hat{\mu}_{RI}^{VA}$</td>
<td>22.8</td>
<td>2.8</td>
<td>17.4</td>
<td>28.4</td>
</tr>
<tr>
<td></td>
<td>$\hat{\mu}_{C}^{VA}$</td>
<td>15.2</td>
<td>2.5</td>
<td>10.6</td>
<td>20.7</td>
</tr>
<tr>
<td></td>
<td>$Y_{C}^{VA}$</td>
<td>12.3</td>
<td>0.9</td>
<td>10.5</td>
<td>14.1</td>
</tr>
</tbody>
</table>

Based on the results in Table 1, we conclude that more precise versions of the 50-State Survey estimates were indeed successfully adjusted towards the National Survey estimates, resulting in model predictions in the National Survey currency. The model predictions in the 50-State Survey currency have lower standard errors and narrower uncertainty intervals than the corresponding 50-State Survey estimates. Our model-based estimates in the alternative currencies and their corresponding measures of uncertainty rely critically on our modeling assumptions. However, our diagnostics did not indicate departures from the assumptions, and the resulting model-based estimates are consistent with the reliable estimates from external data, giving some reassurance that our model specification is appropriate. As an informal comparison, we conducted two sample t-tests (not shown here) to compare model predictions in the National Survey currency and the National Survey estimates, finding no significant differences in any of the four states at level $\alpha = 0.05$.

6. Alternative Model Specifications. For the smoothing level, we assumed state-level effects nested within census division-level effects. Replacing the grand mean and the division-level effect by a division-level mean would provide an alternative model specification. Specifically, the smoothing level would be $\mu_{RI}^{AL} = u_i + s_{ij}$, and a weakly-informative prior may be adopted for the division means, say $u_i \sim N(0, 10^4)$. We implemented this alternative model and compared the results with those from the proposed model above. As expected, in the absence of a hyperparameter with shrinkage prior for the standard deviation of the division means, the model predictions in the 50-State Survey currency stay closer to the 50-State Survey survey estimates (they are shrunk less towards the overall model trend) and present more overall variability across domains (states or census divisions). However, both
of these differences are small, because the proposed model does not imply much shrinkage in the model predictions to begin with. The results are omitted.

Removing the smoothing level from the model would provide yet another alternative model specification. This level is included to help improve the precision of the final model predictions, and is not necessary for the bridging problem (reconciliation of the two survey estimates). We implemented this alternative model with bridging only and compared its results to those from the proposed model with both bridging and smoothing. As expected, the model predictions in the 50-State Survey currency stay closer to the 50-State Survey survey estimates (no smoothing effects), there is no reduction in the 50-State Survey standard errors, and the model predictions in the National Survey currency are less precise and more variable across all the divisions. The results are omitted.

7. Discussion. We developed statistical methodology to bridge the 2016 National Survey and 50-State Survey so that the adjusted census division estimates from the 50-State Survey are comparable to estimates from the National Survey, and state-level estimates based on the 50-State Survey are now available in the National FHWAR currency. The hierarchical model presented in this paper accounts for data available from multiple sources and at different levels of geography, and provides a solution to overcome the challenging bridging task. The sources of the data and the code are made available as a supplement to this paper, and may serve as starting points for any additional analyses.

The methods presented can be extended to more than two surveys and to reconciliation for other purposes, such as improving estimation or forecasting. For example, an additional sampling level may be included in the hierarchical model specification for an additional survey $A$ with information available at a level of aggregation $g$, $\mathbf{Y}_g^A \sim N(\mathbf{\mu}_g^A, \text{Var}(\mathbf{Y}_g^A))$, with corresponding bridging and smoothing levels, as needed or applicable. Also, multiple quantities of interest may be bridged jointly, to further improve the precision of their estimates and to maintain their observed sampling relationship. For example, the sampling level may be extended to $\mathbf{Y}_g^A \sim N(\mathbf{\mu}_g^A, \text{Var-Cov}(\mathbf{Y}_g^A))$, where $\mathbf{Y}_g^A$ is a multivariate vector of estimates at level $g$, constructed using data from survey $A$. To further improve estimation, auxiliary data with good predictive power may be used. For example, the smoothing model may be extended to use state-specific information $x_{ij}$ as covariates, $\mu_{ij}^{RI} = x_{ij}^T \beta + u_i + s_{ij}$.

In our application, we constructed and modeled survey estimates at different levels of aggregation—state and census division-level—because there was no state indicator available in the public use files containing the National FHWAR Survey data for 2016. One could model survey direct estimates available at the same level of aggregation. For this, the sampling models would both be specified at the same level of aggregation, and the identity between low and high levels of aggregation would no longer be applicable. The smoothing effects in the smoothing model could still be specified at multiple levels of aggregation, in order to borrow strength across each level as well as from the nested structure of the levels. For our application, the level of aggregation was defined in terms of geography, but other domains could be considered, for example socio-economic domains.

REFERENCES


