Journal of Survey Statistics and Methodology (2024) ***12***, 458–492

BAYESIAN INTEGRATION OF PROBABILITY AND NONPROBABILITY SAMPLES FOR LOGISTIC REGRESSION

CAMILLA SALVATORE *
SILVIA BIFFIGNANDI
JOSEPH W. SAKSHAUG
ARKADIUSZ WIŚNIOWSKI
BELLA STRUMINSKAYA

Probability sample (PS) surveys are considered the gold standard for population-based inference but face many challenges due to decreasing response rates, relatively small sample sizes, and increasing costs. In contrast, the use of nonprobability sample (NPS) surveys has increased significantly due to their convenience, large sample sizes, and relatively low costs, but they are susceptible to large selection biases and unknown selection mechanisms. Integrating both sample types in a way that exploits their strengths and overcomes their weaknesses is an ongoing area of methodological research. We build on previous work by proposing a method of supplementing PSs with NPSs to improve analytic inference for logistic regression coefficients and potentially reduce survey costs. Specifically, we use a Bayesian framework for inference. Inference relies on a

Camilla Salvatore is a Postdoctoral Fellow in Statistics at the Department of Methodology & Statistics, Utrecht University, Sjoerd Groenmangebouw, Padualaan 14, 3584 CH Utrecht, The Netherlands. Silvia Biffignandi is a Full Professor of Economic Statistics (Retired) at the Department of Management, Economics and Quantitative Methods, University of Bergamo, Bergamo, Italy. Joseph W. Sakshaug is a Professor of Statistics at the Department of Statistical Methods Research, Institute for Employment Research, Nuremberg, Germany. Arkadiusz Wiśniowski is a Professor of Social Statistics & Demography at the Social Statistics Department, University of Manchester, Manchester, UK. Bella Struminskaya is an Associate Professor of Methodology and Statistics at the Department of Methodology & Statistics, Utrecht University, Utrecht, The Netherlands.
Camilla Salvatore conducted the majority of her work as a PhD candidate at the Department of Economics, Management and Statistics (DEMS), University of Milano-Bicocca. We greatly acknowledge the DEMS Data Science Lab for supporting this work by providing computational resources. The study design and analysis were not preregistered.
*Address correspondence to Camilla Salvatore, Department of Methodology & Statistics, Utrecht University, Sjoerd Groenmangebouw, Padualaan 14, 3584 CH Utrecht, The Netherlands; E-mail c.salvatore@uu.nl.

https://doi.org/10.1093/jssam/smad041 Advance access publication 17 November 2023

probability survey with a small sample size, and through the prior structure we incorporate supplementary auxiliary information from a less-expensive (but potentially biased) NPS survey fielded in parallel. The performance of several strongly informative priors constructed from the NPS information is evaluated through a simulation study and real-data application. Overall, the proposed priors reduce the mean-squared error (MSE) of regression coefficients or, in the worst case, perform similarly to a weakly informative (baseline) prior that does not utilize any nonprobability information. Potential cost savings (of up to 68 percent) are evident compared to a probability-only sampling design with the same MSE for different informative priors under different sample sizes and cost scenarios. The algorithm, detailed results, and interactive cost analysis are provided through a Shiny web app as guidance for survey practitioners.

KEY WORDS: Bayesian inference; Data integration; Online access panel; Selection bias; Web survey.

Statement of Significance

Probability sample surveys are the gold standard for population-based inference, but they can be prohibitively expensive, particularly those with large sample sizes. Thus, survey organizations must weigh the option of fielding smaller probability samples or embracing less expensive, but potentially biased, nonprobability sample surveys. We propose a Bayesian data integration approach where inference is based on a probability-based survey with a small sample size and supplementary auxiliary information from a nonprobability survey fielded in parallel is incorporated through the prior structure. We show that the approach can improve inference about parameters in a logistic regression model and potentially reduce survey costs. A Shiny web app with an interactive cost analysis feature is provided for survey practitioners.

1. INTRODUCTION

Probability sampling has long been considered the gold standard method for designing large-scale, population-based surveys. It provides a scientifically sound framework for making population-based inference as a function of the sample design with measurable bounds of uncertainty (Neyman 1934). All population units have a known (or knowable) nonzero chance of selection and through inverse probability weighting it is possible to construct design-unbiased estimators (Kish 1965). However, unbiasedness is threatened by the practical realities of surveys, including noncoverage, nonresponse, and other error sources defined in the total survey error (TSE) framework (Biemer 2010). Nonresponse in particular has significantly increased over the years, mainly driven by increasing noncontacts and refusals. Strategies to cope with nonresponse, such as through higher incentives or more intensive fieldwork efforts have failed to stymie this trend, and have increased survey costs to a point where fielding large probability sample (PS) surveys has become cost-prohibitive for many researchers and survey organizations (Luiten et al. 2020).

Given these budget pressures, many researchers and survey organizations have embraced the more affordable and timely alternative of nonprobability sample (NPS) surveys, specifically those conducted via online access (or volunteer opt-in) panels. Due to their relatively low cost, the use of online NPS surveys has increased in recent years (Biffignandi and Bethlehem 2021). However, because they rely on individuals self-enrolling themselves into the panel and agreeing to take part in periodic surveys, they are strongly susceptible to selection bias, often more so than carefully designed PS surveys with low response rates (Dutwin and Buskirk 2017; Cornesse et al. 2020). Consequently, a current strand of research treats NPS surveys as a supplement rather than as a replacement for PS surveys and develops approaches for integrating both sample types. The present study builds on this ongoing research.

Specifically, we propose a Bayesian approach to supplement a PS survey with small sample size with information from an NPS survey fielded in parallel to improve inference about logistic regression coefficients for a binary outcome. Several strongly informative priors constructed from the NPS information are evaluated in terms of the mean-squared error (MSE) of the posterior estimates through a simulation study and real-data application. We show that supplementing PS surveys with these priors produces coefficient estimates with lower MSEs and at lower costs compared to not using any NPS information and relying solely on PS data for inference. An R Shiny web app (https://bayesdataintegration.shinyapps.io/shiny_bayes_data_integration/) is provided that displays the algorithm and full results and includes an interactive cost analysis tool to estimate potential cost savings of the method for different cost scenarios.

The remainder of the article is organized as follows. Section 2 provides background on the topic and reviews the relevant literature. Section 3 outlines the research aims. Section 4 introduces the methodological framework. Section 5 presents the simulation results and in section 6 the results of the real-data application are presented. The article concludes in section 7 with a general discussion of the findings, recommendations, and potential research extensions.

2. BACKGROUND

Probability sampling is considered the fundamental method for making population-based inferences. Given its well-established statistical framework for unbiased estimation with measurable bounds of uncertainty (Neyman 1934), and it is a well-known methodological framework for studying nonsampling (e.g., coverage, measurement, nonresponse) errors (Biemer 2010), probability sampling continues to be the preferred approach for official statistics and academic research (Bethlehem et al. 2009; Brick 2011). However, achieving desired levels of precision can be challenging. Large sample sizes are often necessary, which can be costly and time-consuming to obtain, especially for researchers and survey organizations working with modest budgets and tight time schedules. For these reasons, some researchers are turning to less expensive, but potentially more biased, NPS surveys (Cornesse et al. 2020). Below we discuss the strengths and weaknesses of NPS surveys and consider their use as a supplement to PS surveys.

Participants in NPS surveys are typically recruited from online access panels or directly from visited web sites, including search engines and social media sites (Baker et al. 2010). Thousands of internet users opt-in to online access panels and crowdsourcing platforms, often in exchange for incentives or rewards to complete periodic web surveys. Thus, NPS web surveys can be completed by a large number of respondents rapidly and at low cost, enabling timely statistics (Kreuter et al. 2020; Astley et al. 2021) and reaching rare or hard-to-interview populations (Berzofsky et al. 2018). However, the drawbacks of online access panels are that there is no explicit sampling frame of the general internet population, the data-generating process is typically outside the researcher's control, and the lack of a known random selection mechanism renders the classical design-based approach to inference inappropriate. Moreover, parts of the general population are not covered, such as people without internet access and those who were not exposed to, or targeted by, the access panel's advertising efforts (Bethlehem 2010). Another concern arises from the presence of bots and professional respondents, who participate in many online panels solely to make money and may engage in undesirable response behaviors, leading to less reliable data (Hillygus et al. 2014; Johnson et al. 2021). Moreover, the empirical evidence suggests that the accuracy of NPS survey estimates is usually lower than those obtained from PS surveys (Yeager et al. 2011; Dutwin and Buskirk 2017; Cornesse et al. 2020).

For these reasons, serious concerns remain about the generalizability of NPS surveys and their suitability as a replacement for PS surveys. If accurate population-based inference is the goal, then PS surveys appear to be generally superior to NPS surveys, where the latter may be viewed more as an auxiliary data source or supplement to PS surveys—a viewpoint that forms the basis of the current study. Thus, the field of survey research is in a situation where PS surveys are known to have higher data quality but large samples are expensive while NPS surveys are convenient and more affordable but can suffer from large selection biases. Consequently, a natural avenue of research is the integration of both PS and NPS surveys to exploit their respective advantages in a way that overcomes their respective disadvantages and minimizes overall survey costs (Couper 2013; Miller 2017; Beaumont 2020; Rao 2021).

Classic data integration approaches include the construction of pseudo-weights and/or calibrated weights based on auxiliary variables or population totals (Elliot 2009; DiSogra et al. 2011; Raghunathan et al. 2020; Robbins et al. 2021). An alternative approach is a doubly robust inference where the quasi-randomization approach for the construction of pseudo-weights and the modeling of the outcome variable of interest are combined (Elliott and Valliant 2017; Yang et al. 2020). A key aspect of this approach is that the estimator is approximately unbiased if either one of the models is correctly specified. Another approach is mass imputation which comes from the missing data literature (Kim et al. 2021). Small area estimation methods are also applied for integrating multiple data sources (Ganesh et al. 2017; Beaumont and Rao 2021). Moreover, the availability of new data sources and unstructured big data offers new methodological possibilities (Stier et al. 2020). Kim and Tam (2021) address the problem of finite population inference when integrating big data sources and a PS survey accounting for both selection bias and measurement error without making missing at random (MAR) assumptions. Another option is to integrate both data sources under a Bayesian framework using latent class or hierarchical models (Sakshaug et al. 2019; Hsiao et al. 2020; Wiśniowski et al. 2020; Alexander et al. 2022).

Many studies on combining PS and NPS focus on finite population inference. Nevertheless, other types of inference can also be of interest, such as the study of associations and model parameters. While descriptive estimates tend to have larger discrepancies between the two sample types compared to correlations and regression coefficients (Pasek 2016), the literature is mixed with some studies reporting strong correspondence between PS and NPS surveys for regression coefficients and other studies reporting larger discrepancies (Malhotra and Krosnick 2007; Callegaro et al. 2014; Thompson and Pickett 2020). The presence of selection bias in NPS surveys for regression coefficients has recently been studied by West et al. (2021) who propose indices of nonignorable selection bias in linear and probit regression models based on a pattern-mixture model and on the availability of aggregate auxiliary data.

Considering budget constraints, researchers may not always have the resources to afford large PS size. Rather than relying solely on potentially biased NPS data, an alternative strategy involves collecting a small PS alongside a larger NPS. Designing the survey in this way and exploiting data integration can result in cost savings and improved inference. A recent application of this approach was described in Johnson et al. (2021), who investigated the association between gun ownership and perceptions about COVID-19 by integrating a small PS ($n = 77$) survey conducted in parallel with a larger ($n = 1{,}120$) NPS survey using a similar Bayesian framework to the one we propose, but for continuous outcome variables. Our proposed Bayesian framework contributes additional machinery to the analysis of binary outcomes that can be applied under similar study designs. In section 3, we provide a more detailed explanation of our research objectives.

3. RESEARCH AIMS

The present study focuses on integrating PS and NPS surveys for improving analytic inference about coefficients for logistic regression models and potentially reducing survey costs, which is still an emerging topic in the literature. Our contribution focuses on supplementing a PS survey with small sample size with information from a parallel NPS survey with overlapping analysis variables. This approach is likely to be attractive to survey researchers interested in making population-based inferences from PS surveys but cannot afford to field large PSs. These researchers are typically faced with the decision of reducing the size of their PSs and tolerating large variances, or opting for a larger and cheaper, but potentially more biased NPS survey. By fielding a small PS in parallel with a relatively inexpensive supplementary NPS survey, the larger variances of the PS survey can potentially be reduced, though possibly at the expense of introducing bias depending on how much weight is given to the NPS data in the estimation. Thus, our framework aims to provide a practical data integration solution that overcomes the limitations of each standalone sample type while prioritizing the PS data and its inferential properties and possibly reducing costs (relative to fielding a large PS-only survey).

To combine the information coming from the two samples, we consider a Bayesian framework where inference is based on the PS survey and available information from the potentially biased NPS survey is supplied through a strongly informative prior. We build on the previous work of Sakshaug et al. (2019), Wiśniowski et al. (2020), and Nandram and Rao (2021), who proposed a similar framework for the analysis of a continuous outcome variable using linear regression. However, categorical data analysis, and particularly the modeling of binary outcomes is of key interest in the health and social sciences, where the objective is to study the classification of behaviors, attitudes, and characteristics (e.g., healthcare coverage, unemployment, voting, and illness, among others). Thus, we extend the previous work by developing an approach for modeling binary outcomes with covariates using logistic regression and leave the analysis of other categorical data types to future work.

To evaluate the proposed method, we conduct a simulation study to compare the performance of several strongly informative priors in terms of the MSE for the posterior estimates according to different selection mechanisms, selection probabilities, and PS and NPS sample sizes. The comparisons are made in reference to a weakly informative ("baseline") prior in which no NPS information is supplied and inference is based on PS survey data alone. In contrast to previous studies, which generally assume a MAR selection mechanism for the NPS data, we do not make such an assumption and also evaluate the framework in the missing not at random (MNAR) context, where the NPS selection mechanism depends on the outcome variable of interest. Under this framework, incorporating biased NPS data through a strongly informative prior is likely to result in posterior estimates that have more bias, but possibly less variance compared to using a naïve baseline prior that does not utilize any NPS information. Thus, our main interest lies in investigating under which conditions this reduction in variance offsets increases in bias, thus leading to lower posterior MSEs relative to a probability-only sample. We expect that any MSE reductions will be most evident when considering small PS sizes, where the strongly informative priors will have the most influence on reducing the posterior variance. In general, the results will likely vary according to the underlying selection mechanism of NPS data and their level of bias.

In addition to the simulation study, we evaluate the strongly informative priors through a real-data application involving a nationally representative, probability-based web survey and several overlapping nonprobability web surveys conducted in parallel with potentially different selection mechanisms. A cost analysis is performed to study the extent to which the strongly informative priors produce posterior estimates at a lower cost for the same MSE as would be obtained from a (potentially more expensive) probability-only sample. Here, we expect that the largest potential cost savings will occur when PS sizes are small and the MSE reduction is large.

4. METHODOLOGY

4.1 The Bayesian Inferential Framework

The Bayesian framework offers a unified approach for integrating multiple data sources of different sizes and quality in a natural way, that is, through the prior structure. In the proposed methodology, inference is based on the PS survey and additional information from the NPS survey is incorporated through a strongly informative prior. This is the setup used in the previously cited work (Sakshaug et al. 2019; Wiśniowski et al. 2020; Nandram and Rao 2021). Nandram and Rao (2021) also considered the reverse setup, where inference is based on the NPS survey with supplementary PS survey data incorporated through a strongly informative prior, but concluded that this was inferior to the approach we consider. In particular, the authors state that "it is erroneous to use the PS as the prior or to use the PS to supplement the NPS, rather one should use the NPS to supplement the PS, provided the study variable is available in the PS" (Nandram and Rao 2021, p. 1593). Thus, we do not consider this setup in the current study.

Our aim is to improve inference about logistic regression model parameters by fielding a small PS survey and supplementing it with overlapping NPS survey information collected in parallel. We assume that the PS survey is of high quality (i.e., unbiased) despite its small sample size and that the NPS survey might be subject to large selection biases and thus has lower quality.

We consider logistic regression to model a binary outcome with covariates. Let us denote with Y_{PS} the binary response vector of size $n_{\text{PS}} \times 1$ and X_{PS} the $n_{\text{PS}} \times k$ design matrix from a PS survey. The PS data are denoted by $D_{\text{PS}} = (n_{\text{PS}}, Y_{\text{PS}}, X_{\text{PS}})$. Similarly, the data from a parallel NPS survey are denoted by $D_{\text{NPS}} = (n_{\text{NPS}}, Y_{\text{NPS}}, X_{\text{NPS}})$. The likelihoods of the NPS and PS data are denoted by $L(\boldsymbol{\beta}|D_{\text{NPS}})$ and $L(\boldsymbol{\beta}|D_{\text{PS}})$, respectively.

The logistic model is presented in equation (1), where $\theta_i = \frac{\exp(X_i'\boldsymbol{\beta})}{1+\exp(X_i'\boldsymbol{\beta})}$ are the success probabilities:

$$Y_{\text{PS}i} \sim \text{Ber}(\theta_i),$$
$$\text{logit}(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \beta_{0\text{PS}} + \sum_{j=1}^{k} \beta_{j\text{PS}} X_{\text{PS}ij} \text{ for } i = 1, \ldots, n_{\text{PS}}. \tag{1}$$

Inference is expressed through the posterior distribution, $\pi(\boldsymbol{\beta}|D_{\text{PS}}, D_{\text{NPS}})$, which is based on the likelihood function for the PS data $L(\boldsymbol{\beta}|D_{\text{PS}})$ and the prior distribution $\pi(\boldsymbol{\beta})$ through Bayes theorem.

4.2 Construction of the Prior Distributions

Eliciting the prior is a key step in Bayesian analysis. A strongly informative prior distribution incorporates previous information or beliefs about the parameters before the data are observed. Such information may come from the literature, historical data, or expert opinions. If there is no previous information or beliefs to incorporate, then vague or weakly informative priors are used to reflect this lack of knowledge.

4.2.1 Strongly informative priors

We derive several strongly informative prior specifications that incorporate information from an NPS survey about a model parameter. We first consider multiple variations of a normally distributed prior, which is a common choice for constructing priors. The general idea behind this class of proposed priors is to set the location parameter equal to the maximum-likelihood estimate (MLE) of the target parameter from the substantive model of interest obtained from the NPS data and scale this information through the scale parameter. We propose different formulations for the scale parameter taking into account the distance between the MLEs from the separate PS and NPS data sources and the NPS size. We refer to this class of priors as distance priors.

The first prior from this class of distance priors is simply referred to as the *Distance prior*, which was originally proposed by Sakshaug et al. (2019) for continuous outcomes:

$$\beta_j \sim \mathcal{N}(\hat{\beta}_{j\mathrm{NPS}}, |\hat{\beta}_{j\mathrm{PS}} - \hat{\beta}_{j\mathrm{NPS}}|). \tag{2}$$

The scale parameter is set equal to the absolute difference between the two MLEs from the PS and NPS data, denoted by $\hat{\beta}_{\mathrm{PS}}$ and $\hat{\beta}_{\mathrm{NPS}}$, respectively. These MLEs can be estimated using standard logistic regression functions in statistical software packages. The larger the absolute difference (an indication of larger selection bias in the NPS data), the smaller the influence of the prior on the posterior.

A shortcoming of such a formulation is that the scale parameter can be very small or equal to zero in extremely unlikely cases (Sakshaug et al. 2019). Thus, as an alternative we propose to take the maximum value of the squared difference between the ML estimates and the variance of the MLE based on the NPS data. Then, we use the NPS size to shrink the prior around $\hat{\beta}_{j\mathrm{NPS}}$. To do that, we use an inverse logarithmic scaling factor $\frac{1}{\log(n_{\mathrm{NPS}})}$, where n_{NPS} is the length of the vector of the NPS data (Wiśniowski et al. 2020). This may lead to potentially more bias but lower posterior variance. We refer to this prior as the *Distance-log prior*:

$$\beta_j \sim \mathcal{N}\left(\hat{\beta}_{j\mathrm{NPS}}, \sqrt{\frac{1}{\log(n_{\mathrm{NPS}})} \cdot \max\left((\hat{\beta}_{j\mathrm{PS}} - \hat{\beta}_{j\mathrm{NPS}})^2, \hat{\sigma}^2_{\hat{\beta}_j\mathrm{NPS}}\right)}\right). \tag{3}$$

While these two prior formulations are not new, the literature suggests that they have never been applied in a logistic regression setting.

We propose a slightly modified prior specification using the common logarithm instead of the natural logarithm, which will result in a slightly wider distribution. This prior is referred to as the *Distance-log10 prior*:

$$\beta_j \sim \mathcal{N}\left(\hat{\beta}_{j\mathrm{NPS}}, \sqrt{\frac{1}{\log_{10}(n_{\mathrm{NPS}})} \cdot \max\left((\hat{\beta}_{j\mathrm{PS}} - \hat{\beta}_{j\mathrm{NPS}})^2, \hat{\sigma}^2_{\hat{\beta}_j\mathrm{NPS}}\right)}\right). \tag{4}$$

If participation in the NPS survey depends directly on the outcome variable of interest, then also the intercept will be biased. Thus, we consider a mixed formulation for the distance prior specifications in equations (2)–(4), where the prior for the intercept is replaced by a weakly informative Student-t prior distribution with three degrees of freedom, t_3. We refer to this set of priors as the mixed-distance priors (Mixed-Distance, Mixed-Distance-log, and Mixed-Distance-log10, respectively). In all of these prior formulations, the issue of using the PS data twice arises. Indeed, inference is based on the PS data which are also used as a reference to construct the scale (variance) parameter of the priors. The use of PS data on the second-order prior component essentially serves as protection against a prior informed from a severely biased NPS data source from dominating the posterior inference.

Lastly, we propose the *Power prior* to integrate the two sample types. Chen et al. (1999; 2000) and Ibrahim and Chen (2000) introduced this new class of strongly informative prior distributions based on the availability of historical data. The prior's properties have been discussed in different contexts, for example, in regression models (Ibrahim and Chen 2000), variable selection, logistic regression (Chen et al. 1999), and generalized linear models (Chen et al. 2000). The term *historical data* refers to both data from previous studies or similar parallel studies that are used to inform the model parameters. The Power prior is mainly used in clinical trials and health applications (De Santis 2006; Ibrahim et al. 2012; Haddad et al. 2017; Thompson et al. 2021; De Santis and Gubbiotti 2023; Wang et al. 2023). The Power prior has been previously used to integrate PS and NPS surveys by Nandram and Rao (2021). They consider the case of a random power parameter with values sampled using a grid method. We propose a different approach that dynamically borrows information from the NPS data based on its similarity with the parallel PS data.

In our context, the NPS survey can be viewed as the *historical data*. The degree of influence of the NPS data on the posterior inference is determined by the power parameter $0 \leq a \leq 1$. The case of $a = 0$ corresponds to no borrowing of information while $a = 1$ reflects the case of full borrowing. Although it is possible to specify a prior for the power parameter a, we derive a fixed value for this parameter. The initial prior for $\boldsymbol{\beta}$ is denoted by $\pi_0(\boldsymbol{\beta})$. Equation (5) shows the prior specification for fixed a:

$$\pi(\boldsymbol{\beta}, a|D_{\mathrm{NPS}}) \propto L(\boldsymbol{\beta}|D_{\mathrm{NPS}})^a \pi_0(\boldsymbol{\beta}). \tag{5}$$

Thus, the resulting posterior in equation (6) is also proportional to the NPS data:

$$\pi(\boldsymbol{\beta}|D_{\mathrm{PS}}, D_{\mathrm{NPS}}, a) \propto L(\boldsymbol{\beta}|D_{\mathrm{PS}})L(\boldsymbol{\beta}|D_{\mathrm{NPS}})^a \pi_0(\boldsymbol{\beta}). \tag{6}$$

We set the prior $\pi_0(\boldsymbol{\beta})$ to be weakly informative as in equation (7). We choose not to set a to an arbitrary fixed value because it may lead to high MSE values in the presence of high selection bias. Instead, we derive a in an automated fashion according to the similarity between the MLEs obtained from the PS and NPS data. Specifically, we set a equal to the p-value resulting from the Hotelling's T^2 test for the difference between the two vectors, $\hat{\boldsymbol{\beta}}_{\mathrm{PS}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{NPS}}$. p-Values close to 1 indicate strong evidence for the null hypothesis and, in such cases, it is suggested to *borrow* more heavily from the NPS data. For p-values close to 0 (indicating significant differences between the two vectors), the amount of information borrowed is smaller or, in the worst-case, is nil. This method allows for automatic rescaling of $L(\boldsymbol{\beta}|D_{\mathrm{NPS}})$ based on the difference between the MLEs.

4.2.2 Weakly informative (baseline) prior

To form the basis for evaluating the informative priors, we evaluated several vague priors to serve as baseline priors, including uniform and some normally distributed priors centered around zero with large-scale parameters. However, the results were not satisfying, especially for small sample sizes. Thus, we discarded these priors and focused our attention on weakly informative priors.

We refer to Gelman et al. (2008) for a discussion of suitable weakly informative priors for logistic regression. The authors do not recommend the use of vague normal priors and give preference to the location-scale family of the Student's t-distribution. In particular, they suggest a *t*-density function with 7 degrees of freedom and scale equal to 2.5, which is close to the likelihood of a single binomial trial. A more conservative choice is also proposed, which is a Cauchy prior with scale parameter equal to 2.5. However, Ghosh et al. (2018) show that in such cases sampling from the posterior is challenging in the presence of separation and, thus, it is recommended to use a t-distribution with degrees of freedom ν between 3 and 7. We use $\nu = 3$. The resulting prior is formalized in equation (7) and referred to as the *Baseline prior*:

$$\beta_j \sim \text{Student}\ (\nu = 3, \mu = 0, \sigma = 2.5). \tag{7}$$

4.3 Posterior Estimation

For the simulation and real-data application, the posterior distributions based on both strongly informative and weakly informative priors are numerically approximated. We use the No-U-Turn sampler, implemented in R (R Core Team 2020) and Stan (Stan Development Team 2019), which is a variant of the Hamiltonian Monte Carlo algorithm. Specifically, we used the R packages rstan (Stan Development Team 2021) and rstanarm (Goodrich et al. 2020). The posterior distributions were obtained using four MCMC chains with samples of 7,000 each and 3,500 burn-in samples which ensured convergence of all chains.

5. SIMULATION STUDY

5.1 The Simulation Framework

The simulation study is designed to evaluate the proposed strongly informative priors under a variety of real-world settings. All settings involve the analysis of binary outcome variables, which is a common application in the health and social sciences. For example, a researcher might be interested in analyzing the propensity to commit a crime, become divorced, drop out of school, or become inflicted with an illness. In some cases, the binary outcome is unbalanced, that is, there is a greater proportion of zeros than ones (or vice versa). For this reason, we consider both balanced and unbalanced outcomes in the simulation study (and also in the application in section 6). To reflect further practical scenarios, we also consider different PS and NPS sizes when setting up the simulation. Because PS surveys are the "gold standard" for inference, but large sample sizes can be prohibitively expensive, we focus our attention on a range of PS sizes, from very small (50–100 cases) to modestly large (up to 1,000 cases). On the contrary, we consider larger NPS sizes which we reasonably assume to be more affordable than similarly sized PS surveys.

Based on these practical considerations, we assume that the outcome variable is generated from the logistic model in equation (1) with two binary predictors: $X_{i1} \sim \text{Ber}(0.5)$ and $X_{i2} \sim \text{Ber}(0.5)$. To test the stability of the results, we consider three specifications for the population regression coefficients, $\beta = (\beta_0, \beta_1, \beta_2)$, namely:

$$\begin{aligned} \beta_{\text{NEG}} &\in (0.5, -1.3, -0.9) \\ \beta_{\text{MIX}} &\in (0.5, -1.3, 0.9) \quad . \\ \beta_{\text{POS}} &\in (0.5, 1.3, 0.9) \end{aligned}$$

These specifications consider both balanced and unbalanced scenarios for the outcome and different cell proportions when combining the three variables. The proportions of Y are 0.37, 0.57, and 0.81, respectively. Table 1 shows the cross-tabulation of the outcome with the covariates under these three scenarios. Thus, we take into consideration population structures in which certain subgroups (derived from cross-tabulating all variables) are smaller in comparison to others.

Under this model, we simulate a population of size $N = 1{,}000{,}000$. The PS is then drawn from this population using simple random sampling without replacement (*srswor*). The use of more complex sampling designs (e.g.,

Table 1. Cross-Tabulation of Variables Used in the Three Simulated Populations

Y	X_1	X_2	NEG	MIX	POS
0	0	0	0.09	0.09	0.09
1	0	0	0.16	0.16	0.16
0	1	0	0.17	0.17	0.04
1	1	0	0.08	0.08	0.21
0	0	1	0.15	0.05	0.05
1	0	1	0.10	0.20	0.20
0	1	1	0.21	0.12	0.02
1	1	1	0.04	0.13	0.23

stratified, cluster) is a topic we leave to future work. We consider different PS sizes, including very small and larger sizes, $n_{PS} \in \{50, 100, 150, 200, 300, 500, 750, 1000\}$. This range of sizes is chosen to reflect different budgetary restrictions that might constrain the sample size that a researcher can afford, but also to align with previous research that has proposed a similar approach to facilitate comparisons (see Sakshaug et al. 2019; Wiśniowski et al. 2020). As previously mentioned, we expect the most significant reductions in MSE (if any) to occur for the very small PS sizes compared to the larger PS sizes (see section 3).

For generating the NPS we first simulate a self-selected panel of individuals who declared their willingness to complete online surveys. From this panel we extract two simple random samples without replacement of different sizes, $n_{NPS} \in \{1000, 5000\}$. We assume that the panel population reflects an online access panel and is thus affected by self-selection. Indeed, the real-life process for joining the panel and eventually completing the survey includes different stages of selection (Valliant and Dever 2011). First, individuals must have an internet connection and visit the recruitment website. Then, they decide to join the panel completing all the required steps. For a specific survey, a sample of individuals is selected from the panel and they can decide whether to participate or not. For simplicity, we assume that all selected units will participate and fully complete the questionnaire (i.e., no unit nonresponse, item nonresponse, or break-offs) without measurement error, just as we assume for the PS survey. Thus, we assign to each population unit a positive probability of participation denoted by p. In general, we set p to be low in order to account for the selection process described above. However, due to the rise of big data sources and the potential of conducting surveys through social media, it may be possible to reach a very large, but very specific part of the population. Thus, we also consider higher values of p.

When the probability of participation depends directly on Y, we are in the case of nonignorable selection bias (or MNAR). Otherwise, if p depends only on observed covariates then we have a MAR selection mechanism; that is, after controlling for X_1 and X_2 in the model, the coefficients will be unbiased. In the latter case, we account for all variables that explain the selection mechanism. However, the MAR assumption is strong and may not hold in practice. Thus, it is important to consider different selection mechanisms when evaluating the proposed data integration method.

We consider five selection mechanisms: (i) p depends on Y only (MNAR); (ii) p depends on Y and X_1 (MNAR); (iii) p depends on Y and X_2 (MNAR); (iv) p depends on X_1 and X_2 (MAR); and (v) p depends on Y, X_1, and X_2 (MNAR). If the probabilities of participation are equal for all units, then there is no selection bias. To introduce bias, we consider four scenarios of varying probabilities of participation p for specific subgroups defined by the value of the selection variables:

$$p = \begin{cases} \{.10, .20, .50, .90\} & \text{if the value of the selection variable(s) is 1} \\ .10 & \text{otherwise} \end{cases},$$

where $p = .1$ reflects the case of no selection bias and $p = .9$ high selection bias.

Then, the probability of participation p is used to generate the participation indicator $P_i \sim \text{Ber}(p_i)$ for $i \in \{1, \ldots, N\}$ for each individual in the population. It follows that the size of the panel N_{Panel} is random. Both PS and NPS are constructed cumulatively and thus, all cases in the smaller samples are always included in the larger ones. We consider standardized covariates for comparability and also because this can reduce auto-correlation in MCMC chains.

The simulation is repeated 100 times (the results were consistent with more repetitions). To compare the performance of the strongly informative priors against the weakly informative baseline prior, we consider the MSE of the posterior estimates. Given the true value of the generic coefficient β, namely β^*, the MSE is defined as follows:

$$\begin{aligned} \text{MSE}\Big(\pi(\boldsymbol{\beta}|D_{\text{PS}}, D_{\text{NPS}})\Big) &= \text{Bias}^2\Big(\pi(\boldsymbol{\beta}|D_{\text{PS}}, D_{\text{NPS}})\Big) + \text{Var}\Big(\pi(\boldsymbol{\beta}|D_{\text{PS}}, D_{\text{NPS}})\Big), \\ &= [\bar{\pi}(\boldsymbol{\beta}|D_{\text{PS}}, D_{\text{NPS}}) - \beta^*]^2 + \text{Var}\Big(\pi(\boldsymbol{\beta}|D_{\text{PS}}, D_{\text{NPS}})\Big), \end{aligned} \tag{8}$$

where in equation (8) $\bar{\pi}(\boldsymbol{\beta}|D_{\text{PS}}, D_{\text{NPS}})$ is the mean of the posterior distribution for a given coefficient and $\text{Var}(\pi(\boldsymbol{\beta}|D_{\text{PS}}, D_{\text{NPS}}))$ is the posterior variance. When describing the results, we always refer to the median value of MSE estimates obtained from the 100 iterations.

When working with small sample sizes and categorical variables, the problem of quasi-complete separation and the Hauck–Donner effect may arise (Yee 2021). In such cases, even if the algorithm converges without any evidence of predicted probabilities numerically equal to 0 or 1, coefficients and standard errors can assume very large and implausible values. We observed the presence of such issues for a few small samples (mainly of size 50). For this reason, we use the median instead of the mean across all simulations to limit the influence of a small number of outliers. The mean values are also available in the Shiny web app.

5.2 Simulation Results

For ease of visualization, figure 1 shows the median MSEs of the regression coefficients for three selected informative priors: Power, Distance, and Mixed-Distance. The results are shown for the case of no selection bias ($p = .1$) and high selection bias ($p = .9$) with sample sizes restricted to n_{PS}

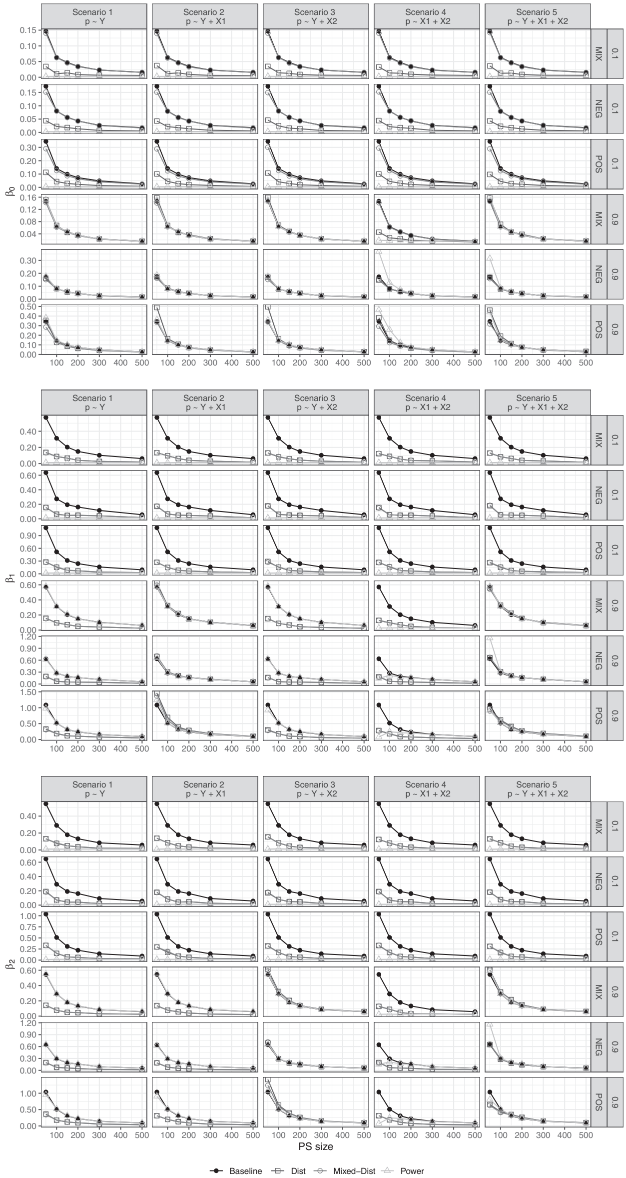

Figure 1. Median MSEs for Regression Coefficients Averaged Over 100 Simulations and for $n_{\text{NPS}} = 5{,}000$ by Probability Sample (PS) Size. Note: Four priors are considered: Distance (Dist), Mixed-Distance (Mixed-Dist), Power, and Baseline. Each panel shows the combination of the five selection scenarios, the three population models: MIX—$\beta_{\text{MIX}} \in (0.5, -1.3, 0.9)$, NEG—$\beta_{\text{NEG}} \in (0.5, -1.3, -0.9)$, and POS—$\beta_{\text{POS}} \in (0.5, 1.3, 0.9)$, and the case with no selection bias ($p = .1$) and high selection bias ($p = .9$).

$\leq$ 500 and for $n_{\text{NPS}} = 5{,}000$. For $n_{\text{PS}} \geq 500$, the MSEs of the informative priors are indistinguishable from those of the baseline prior, and between $n_{\text{NPS}} =$ 1,000 and $n_{\text{NPS}} = 5{,}000$ the differences in MSEs are minuscule. Detailed results for the variance and bias components are available in appendix A in the supplementary data online.

The first three rows of each panel in figure 1 show that for the MAR and MNAR selection scenarios with no bias ($p = .1$), all strongly informative priors produce remarkably lower MSEs than the baseline (weakly informative) prior, especially when $n_{\text{PS}} < 200$. The only exception is for the intercept, where the Mixed-Distance prior, by design, yields MSEs close to the baseline prior. In the presence of large selection bias ($p = .9$; the bottom three rows of each panel in figure 1), the strongly informative priors also reduce the MSEs relative to the baseline prior, but to a lesser extent than the no-bias case. The MSE reductions are hardly affected by whether the outcome is balanced (MIX) or unbalanced (POS and NEG), with the exception that the Power prior tends to perform slightly worse than the baseline prior for very small unbalanced samples. This exception notwithstanding, we may conclude from the figure that the strongly informative priors produce MSEs that are generally smaller or, in the case of high selection bias, similar to those of the baseline prior for all coefficients and selection scenarios.

The full results (including the bias and variance components of the MSE) for all priors and sample sizes are available in the Shiny web app under the tab *Simulation/Results* and in appendix A in the supplementary data online for the scenarios considered in this section. The results clearly show a bias-variance trade-off. In general, the strongly informative priors lead to the posterior estimates being more biased than the estimates based on the baseline prior (which uses PS data only). However, the results also demonstrate that a careful scaling of the NPS information in the prior reduces the variance and, thus, improves the MSE relative to the baseline prior.

The bias-variance trade-off is especially critical for the scenario with the highest level of selection bias ($p = .9$), where the Distance prior yields the smallest MSE relative to the other prior formulations. In the worst case, the Distance prior performs similarly to the baseline prior, while for the Distance-log10 and Distance-log priors, the MSEs can become slightly higher than those of the baseline prior. The Power prior performs especially well for small PS sizes (50–150 observations) and in the MAR selection scenario.

To summarize the full results, we look at the overall performance of each prior across all simulation settings and coefficients, keeping the NPS size equal to 5,000. The first column of table 2 shows the percentage of instances where the MSEs obtained using a strongly informative prior is lower than the MSE obtained using the baseline prior. With this measure, the Mixed-Distance and Mixed-Distance-log10 priors perform best, each yielding 82 percent of the MSEs smaller than the baseline prior. Further, to better understand how large the differences are when the MSEs of the strongly informative priors are *worse*

Table 2. The Percentage of Instances Where the MSE Obtained using the Strongly Informative Prior (MSE_{INF}) Is Less Than the MSE Obtained using the Corresponding Baseline Prior (MSE_{BASE}), and the Percentage of Instances Where the Relative Difference (RD) is Lower Than a Prespecified Threshold of <5%, <10%, <20%, and <30% for the Instances Where MSE_{INF} Is Greater Than MSE_{BASE}

Strongly inf. priors	$MSE_{INF} \leq MSE_{BASE}$	$MSE_{INF} > MSE_{BASE}$			
		$\leq$ 5% RD	$\leq$ 10% RD	$\leq$ 20% RD	$\leq$ 30% RD
Dist	78	45	73	91	96
Mixed-Dist	82	62	82	98	100
Dist-log	65	5	11	21	29
Mixed-Dist-log	78	17	27	38	49
Dist-log10	72	17	34	60	74
Mixed-Dist-log10	82	33	45	72	85
Power	64	74	80	86	89

NOTE.—The priors are: Distance (Dist), Mixed-Distance (Mixed-Dist), Distance-log (Dist-log), Mixed-Distance-log (Mixed-Dist-log), Distance-log10 (Dist-log10), Mixed-Distance-log10 (Mixed-Dist-log10), and Power. A detailed breakdown of results by sample sizes, selection and bias scenarios, and balanced/unbalanced outcomes is available in the Shiny web app under the menu *Simulation/Summary*.

(i.e., larger) than those of the corresponding baseline prior, we calculate the relative difference (RD) between those MSEs defined as:

$$RD = \frac{MSE_{INF} - MSE_{BASE}}{MSE_{BASE}} \quad \text{if } MSE_{INF} > MSE_{BASE}. \tag{9}$$

The other columns of table 2 show the percentage of instances where the RD is lower than 5 percent, 10, 20, and 30 percent, respectively, when $MSE_{INF} > MSE_{BASE}$. In general, an RD up to 5 percent may be considered small, moderate up to 10–20 percent, and large otherwise, though we acknowledge such judgments are subjective.

We observe that the Distance and Mixed-Distance priors perform best, that is, almost always with RDs smaller than 30 percent (96 and 100 percent, respectively), followed by the Power, Mixed-Distance-log10, Distance-log10, Mixed-Distance-log10, and Distance-log priors.

The simulation study demonstrated that the use of strongly informative priors is beneficial to improve the MSEs of coefficient estimates from logistic regression models, especially when the PS size is smaller than 200 observations. However, the amount of such improvements depends on the level of selection bias in the NPS data and the selection mechanism. In the worst-case selection bias scenario ($p = .9$), there is evidence that the MSEs from the strongly informative priors are similar to those of the baseline prior. In the rare instances where $MSE_{INF} > MSE_{BASE}$, the differences are usually relatively small. The MSE reductions are mainly driven by a reduction in the posterior variance which offsets the increase in bias. Overall, the Mixed-Distance prior performs best, with 82 percent of the MSEs being lower than those of the baseline prior and the remaining MSEs never exceeding a 30 percent RD (table 2).

6. APPLICATION: AMERICAN TRENDS PANEL

6.1 The Data

To evaluate the method in a practical setting, a real-data application is presented with an actual PS survey, the American Trends Panel (ATP; Keeter 2019), and nine parallel NPS web surveys carried out by different vendors, which reflect real-world selection scenarios. The ATP is the Pew Research Center's probability-based online panel used for conducting public opinion research. It is representative of the general population aged 18 years and older in the United States and covers both the online and the offline population—before 2016 offline individuals were provided with paper questionnaires or were interviewed by telephone, while in the subsequent years panelists were supplied with the necessary technological tools.

Panel members were originally recruited in 2014 from the Political Polarization and Typology survey (Dimock et al. 2014), a national RDD survey. Additional panelists have been recruited via random-digit-dial telephone surveys in 2015, 2017, and 2018. Panelists are invited to complete at least one survey in each monthly wave. Survey duration is 15 min and a system of financial incentives is implemented. The data we analyze were collected in waves 5 (Pew Research Center 2014a), 7 (Pew Research Center 2014b), and 10 (Pew Research Center 2015a) in 2014 and 2015. We note that wave 5 was part of a mode experiment in which panel members who use the internet were randomly assigned to either web or telephone mode. We analyze the full sample independently of the mode, though a sensitivity check yielded the same conclusions when excluding the telephone cases.

During the same period, Pew sponsored the parallel collection of nine NPS web surveys from different vendors (Pew Research Center 2015b; Kennedy et al. 2016). The same questionnaire was administrated to all respondents with the questions overlapping with those in the ATP, but in different waves. The required sample size was about 1,000 respondents. All vendors implemented quota sampling based on different variables, including age, gender, education, and also other nondemographic variables. More survey details, including target population, response rate, and sample sizes are available in Kennedy et al. (2016) and in the Shiny web app under the menu *Real Data Analysis/Data*.

Six categorical outcome variables are considered. In the case of nonbinary classification, variables were re-coded in a binary fashion. All question wordings are available in the Shiny web app under the menu *Real Data Analysis/Data/Variable Coding*. The questions relate to smoking at least 100 cigarettes in one's entire life (SMOKING; 1 = yes, 0 = no), volunteering in the last 12 months (VOLUNTEERING; 1 = yes, 0 = no), health insurance coverage (HEALTHCARE COVERAGE; 1 = yes, 0 = no), frequency of voting in local elections (ALWAYS VOTE; 1 = always, 0 = otherwise), how many people they trust in their neighborhood (NEIGHBORHOOD TRUST; 1 = all people, 0 = otherwise), and how safe they feel when walking in their neighborhood at night (NEIGHBORHOOD SAFETY; 1 = very safe, 0 = otherwise).

Covariates include binary age (AGE; 1 = 50+, 0 = otherwise), gender (GENDER; 1 = male, 0 = female), education (EDU; 1 = college graduate or higher, 0 = otherwise), and the continuous survey weight variable (SVY WEIGHT; log-transformed). Only some vendors provided weights but the Pew team constructed ATP-style weights for all NPS surveys with the aim to reduce selection bias through a raking adjustment to population benchmarks. We use these weights in the analysis by including them as covariates in the regression models. Before analyzing the data, we drop all observations with missing values and standardize all covariates. More details on the percentage of missing data and the final sample size for each variable are available in the Shiny web app (*Real Data Analysis/Data/Description*).

Figure 2 shows the estimated proportions with 95 percent confidence intervals for a selection of outcome variables (smoking, always vote, and neighborhood trust) across all samples. While there is no evidence of significant differences between the ATP and NPS estimates for the smoking and neighborhood trust variables, differences are evident for the always vote variable in NPS surveys C, D, F, H, and I. The proportions of the other outcome variables are presented in the Shiny web app (*Real Data Analysis/Data/Additional Plots*).

Logistic regression coefficient estimates, based on maximum-likelihood (ML) estimation, are also provided separately for the PS and NPS survey data. As an example, figure 3 shows the estimates for the smoking variable. The figures for always vote and neighborhood trust are in appendix B in the supplementary data online and those for the remaining variables are available in the Shiny web app (*Real Data Analysis/Data/Additional Plots*). Detailed results about the parameter estimates, standard errors, and goodness-of-fit statistics are available in the supplementary data online. Figure 3 shows that the NPS regression coefficients for smoking differ only slightly from the ATP estimates, with the exception of the education variable in sample NP-D and the gender variable in sample NP-I. For the neighborhood trust variable

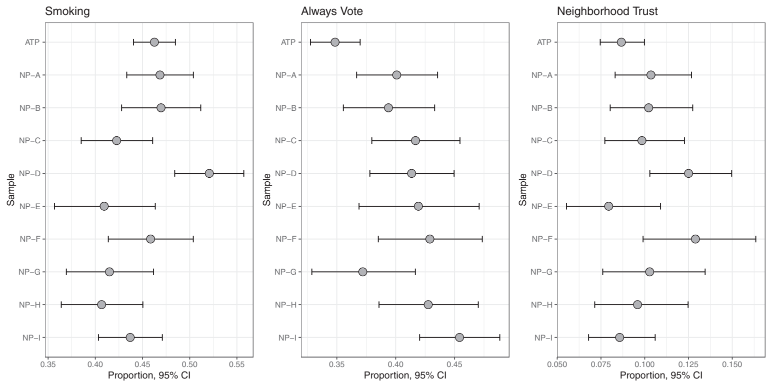

Figure 2. Estimated Sample Proportions (Weighted) with 95 Percent Confidence Intervals for a Selection of Outcome Variables for Each Survey.

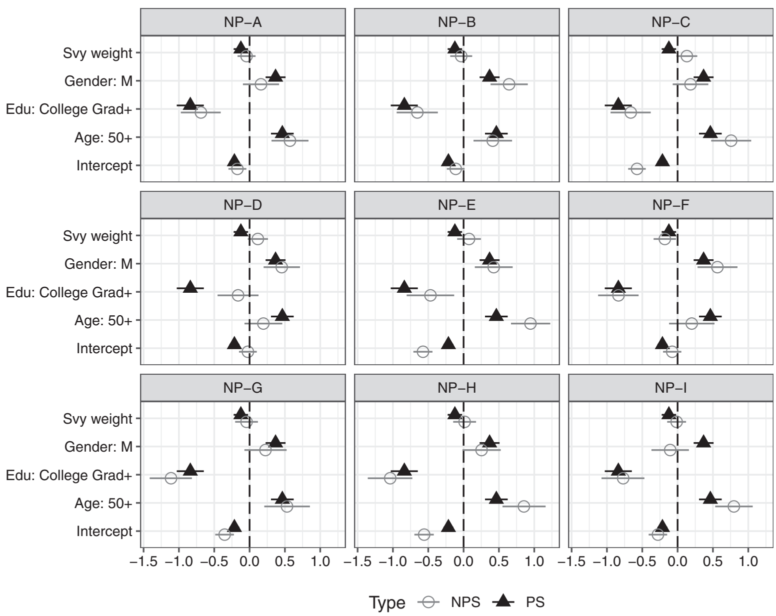

Figure 3. Maximum Likelihood Estimates of Logistic Regression Coefficients and 95 Percent Confidence Intervals for the Smoking Outcome Variable in the ATP (Triangles) and Nine Non-Probability Surveys (Circles).

(figure B.2 in the supplementary data online), the NPS and ATP coefficients are very similar, except for age in samples NP-A, NP-B, and NP-D. For always vote (figure B.1 in the supplementary data online), there are notable differences for the education variable where coefficients have opposite signs for all NPS surveys. The same is true for gender in NP-I.

From these comparative analyses, we conclude that the descriptive estimates are more dissimilar between the ATP and NPS surveys compared to the regression estimates, which is consistent with the literature (Pasek 2016). Contrary to the simulation study, which included scenarios with high selection bias, it appears that in the considered real data, the regression coefficients are not heavily affected by a high level of bias.

To apply the proposed methodology, we simulate a situation in which only a small PS size survey is conducted along with a parallel NPS survey. We consider PS sizes $n_{PS} \in \{50, 100, 150, 200, 300, 500\}$. The PS data are drawn with *srswor* from the full ATP data and are assumed to be unbiased. The samples are constructed cumulatively, such that respondents selected for the smaller samples are included in the larger samples. For the NPS surveys, the original sample sizes (approximately 1,000 each) are used.

To compute the posterior bias, the true values are defined by the vector of ML estimates obtained using the full ATP sample (where the sample size ranges between 3,106 and 3,331 respondents depending on the outcome variable of interest), namely $\boldsymbol{\beta}^*$, which provides an unbiased result by assumption. The entire procedure is repeated 100 times and, as in the simulation study, the median MSE values are reported across all repetitions. Only for the healthcare coverage variable, which is highly unbalanced, the model was not estimable for some iterations which is likely due to the lack of variation for the smallest PS sizes. As an ad hoc remedy, additional iterations were performed for this outcome variable until 100 estimable results were obtained.

6.2 Results

For brevity, we discuss the results for only a selection of outcome variables: smoking, always vote, and neighborhood trust, and only one NPS data source (NP-A). The results for all other outcome variables and NPS data sources are shown in the Shiny web app under the menu *Real Data Analysis/Results and Summary*. Figure 4 shows the median posterior bias, variance, and MSE for the smoking outcome across the 100 repetitions and for the selected priors as in the simulation study. The figures for always vote and neighborhood trust are shown in figures B.3 and B.4 in the supplementary data online.

For the smoking and neighborhood trust outcomes, which are affected by a low level of selection bias (see figure 3 and figure B.2 in the supplementary data online), all strongly informative priors produce lower MSEs than the baseline prior for all coefficients particularly when $n_{PS} < 200$. The reduction in posterior MSEs is mainly driven by a reduction in the posterior variance. The Power prior leads to the largest reductions in MSE relative to the baseline prior, especially for the smallest PS sizes ($n_{PS} \in \{50, 100\}$). However, the

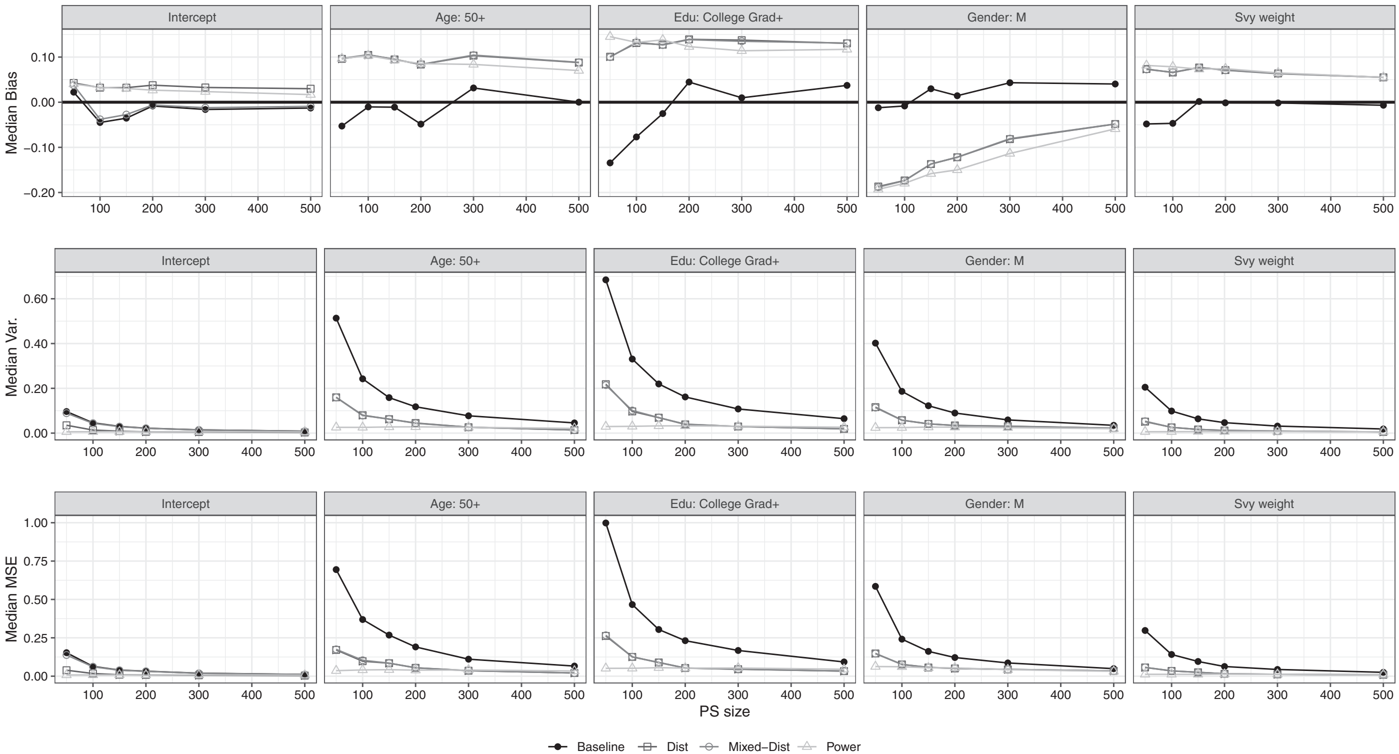

Figure 4. Median Bias, Variance, and MSE of Coefficient Estimates for the Smoking Outcome Using the NP-A Survey to Supply the Prior Information. Note: Four priors are considered: Distance (Dist), Mixed-Distance (Mixed-Dist), Power, and Baseline.

more dissimilar are the ML coefficient estimates between the PS and NPS surveys, the more similar the MSEs of the estimates based on strongly informative priors become to those of the baseline prior. Indeed, for the neighborhood trust variable in sample NP-A, the MLE for the age coefficient is significantly different from the ATP estimate (see figure B.2 in the supplementary data online), which yields MSE curves for the strongly informative priors that are similar to those of the baseline prior (figure B.4 in the supplementary data online).

For the always vote variable, the strongly informative priors reduce the MSEs with respect to the baseline prior for almost all coefficients, or in the worst case, the MSE curves are similar to the baseline. In particular, as expected from the analysis of the ML regression coefficients (see figure B.3 in the supplementary data online), there is no significant reduction in the MSEs using the strongly informative priors for the education coefficient. Indeed, the Distance and the Mixed-Distance priors produce MSEs that are equal to or slightly lower than the baseline prior. In contrast, the Power prior produces larger MSEs than the baseline prior for PS sizes between 100 and 200. Thus, while the distance priors protect against excessive selection bias in the education coefficient, this is not true of the Power prior, which decreases the posterior variance but not enough to offset the large education bias. In general, the largest reduction in MSEs relative to the baseline prior is achieved through the Distance-log prior, which is driven by reductions in the posterior variance.

As for the simulation study (section 5), we summarize the results of the application in table 3 by showing the percentage of instances where the MSE of a coefficient using the strongly informative priors is lower than the MSE obtained using the baseline prior, and, for the instances where the performance of the strongly informative priors is worse (i.e., $MSE_{INF} > MSE_{BASE}$), the percentage of instances where the RD (9) is lower than 5 percent and 30 percent, representing the two extremes, across all PS sizes and NPS surveys used to construct the strongly informative priors. In the Shiny web app under the menu *Real Data Analysis/Summary*, such a table is available for each combination of n_{PS} and NPS survey. As inferred from section 6.1, most regression coefficients from the NPS data are not affected by a high level of selection bias, thus, the strongly informative priors lead to lower MSEs in most cases, as expected.

The neighborhood trust and healthcare coverage variables are two particularly notable cases in terms of MSE reduction using the strongly informative priors. In more than 99 percent of instances, the distance priors yield lower MSEs than the baseline prior, indicating improvements in the MSEs regardless of which NPS survey is used to supply the prior information. The Power prior performs similarly and such percentages are about 98 and 97 percent for these two outcomes, respectively. Similar results are achieved when considering the smoking and neighborhood safety outcomes, where the best priors are the Mixed-Distance-log10 (99 percent) and the Distance (97.4 percent) for the two outcomes, respectively. For always vote, the best prior is the Distance-log10 (93.3 percent) and, for volunteering, it is the Mixed-Distance-log10 together

Table 3. The Percentage of Instances Where the MSE Obtained using the Strongly Informative Prior (MSE_{INF}) Is Less Than or Equal to the MSE Obtained using the Corresponding Baseline Prior (MSE_{BASE}), and the Percentage of Instances Where the Relative Difference (RD) Is Lower Than a Prespecified Threshold of <5% and <30% for the Instances Where MSE_{INF} Is Greater Than MSE_{BASE}

Strongly inf. priors	Smoking			Always vote			Volunteering		
	$MSE_{INF} \leq MSE_{BASE}$	$MSE_{INF} > MSE_{BASE}$		$MSE_{INF} \leq MSE_{BASE}$	$MSE_{INF} > MSE_{BASE}$		$MSE_{INF} \leq MSE_{BASE}$	$MSE_{INF} > MSE_{BASE}$	
		$\leq$ 5% RD	$\leq$ 30% RD		$\leq$ 5% RD	$\leq$ 30% RD		$\leq$ 5% RD	$\leq$ 30% RD
Dist	94.8	14.3	100	92.6	65	100	91.5	65.2	100
Mixed-Dist	97.0	100	100	90.4	69.2	100	97.8	83.3	100
Dist-log	92.2	14.3	95.2	79.6	40	94.6	87.4	35.3	97.1
Mixed-Dist-log	96.3	60	100	82.2	35.4	95.8	96.7	44.4	100
Dist-log10	94.8	14.3	100	93.3	55.6	100	91.1	66.7	100
Mixed-Dist-log10	99.6	100	100	89.3	65.5	100	97.8	83.3	100
Power	90.7	56	76	67.8	39.1	80.5	73.3	65.3	76.4

(continued)

Table 3. Continued

Strongly inf. priors	Neighborhood trust			Neighborhood safety			Healthcare coverage		
	$MSE_{INF} \leq MSE_{BASE}$	$MSE_{INF} > MSE_{BASE}$		$MSE_{INF} \leq MSE_{BASE}$	$MSE_{INF} > MSE_{BASE}$		$MSE_{INF} \leq MSE_{BASE}$	$MSE_{INF} > MSE_{BASE}$	
		$\leq$ 5% RD	$\leq$ 30% RD		$\leq$ 5% RD	$\leq$ 30% RD		$\leq$ 5% RD	$\leq$ 30% RD
Dist	99.6	0	100	97.4	85.7	100	97.4	50	100
Mixed-Dist	99.3	50	100	95.6	91.7	100	95.6	50	100
Dist-log	99.6	0	100	93.3	27.8	100	93.3	75	100
Mixed-Dist-log	99.6	0	100	94.1	31.2	100	94.1	50	100
Dist-log10	99.6	100	100	97.4	57.1	100	97.4	0	100
Mixed-Dist-log10	99.3	50	100	96.7	77.8	100	96.7	50	100
Power	98.2	60	100	66.3	69.2	90.1	66.3	62.5	100

NOTE.—The priors are: Distance (Dist), Mixed-Distance (Mixed-Dist), Distance-log (Dist-log), Mixed-Distance-log (Mixed-Dist-log), Distance-log10 (Dist-log10), Mixed-Distance-log10 (Mixed-Dist-log10), and Power. A detailed breakdown of results by PS sizes and NPS surveys are available in the Shiny web app under the menu *Real Data Analysis/Summary*.

with the Distance prior (97.8 percent). The performance of the Power prior is generally worse compared to the distance priors. The worst result is for the always vote and neighborhood safety outcomes where in only 67 percent of cases does the Power prior produce lower MSEs than the baseline prior. Nevertheless, for all strongly informative priors the RD usually does not exceed 30 percent when $MSE_{INF} > MSE_{BASE}$.

In summary, there is evidence that the strongly informative priors reduce MSEs for logistic regression coefficients relative to a weakly informative baseline prior in real-world settings. The smaller MSEs are driven by a reduction in variability and the largest reductions occur for small PS sizes (50–200 observations). These results are also consistent with the simulation study. When selection bias is low (as for neighborhood trust, healthcare coverage), all priors perform similarly well, although the largest reductions in MSEs are achieved with the Power prior for very small PS sizes (50–100 observations) and with the Distance-log prior (and its mixed version) for sample sizes up to 200 observations. However, as the selection bias increases, the Distance or Distance-log10 priors (and their mixed versions) tend to be superior. The mixed priors usually perform better than their nonmixed counterparts and the results are generally consistent across balanced and unbalanced outcome variables. A final aspect to consider regarding the power prior is the values assigned to the power parameter. Additional details about the median values of the power parameter across the 100 repetitions and for all variables considered in the case study are available in figures B.5 and B.6 in the supplementary data online. There is clear evidence of dynamic borrowing, with borrowing decreasing as the size of the PS increases. Furthermore, the degree of borrowing from the same NPS varies across the analyzed outcomes variables, indicating that there is no universal power parameter value that is suitable for all situations. In figure S.1 in the supplementary data online, we also compare the dynamic power prior with a power prior with fixed a parameter under three scenarios: high borrowing (0.9), half borrowing (0.5), and low borrowing (0.1). The results show that the dynamic power parameter prior outperforms all other fixed power parameter priors when the selection bias is very high, thus, offering protection against wrong inferences.

6.3 Cost Analysis

The simulation and real-data application showed that supplementing a small PS survey (100–200 cases) with prior information from a parallel NPS survey results in lower MSEs for logistic regression coefficients compared to not supplementing. To the extent that NPS surveys are less expensive than PS surveys, these results suggest that the same MSE values might be achieved at a lower cost with an integrated sample compared to a larger and likely more expensive standalone PS survey. Cost savings are an important justification for supplementing PS data with NPS data and should be considered when applying the proposed method.

To demonstrate the extent to which the strongly informative priors lead to cost savings (or losses) with respect to a PS-only survey with the same MSEs, we implement a cost analysis which takes into account hypothetical, yet realistic, costs for the ATP and NPS data sources. The cost analysis procedure is described step-by-step in appendix C in the supplementary data online, together with a sensitivity analysis for different assumed costs of the PS and NPS surveys. In short, assuming per-respondent PS and NPS costs, we first estimate the expected cost of fielding a PS-only survey with baseline prior that would achieve the same MSE as fielding parallel PS and NPS surveys and analyzing them using the strongly informative priors. Then, we compare this cost estimate to the cost of fielding the parallel PS and NPS surveys to estimate the percentage of savings (or losses). The cost analysis can be performed interactively within the Shiny web app under the menu *Real Data Analysis/Cost Analysis*, where users can specify different per-respondent costs for the PS and NPS surveys. For illustration, here we assume the cost per respondent to be \$5 in the NPS survey and \$30 in the PS survey. The cost of the PS survey equates to roughly \$2 per interview minute for a 15-minute interview, which is consistent with the cost of similar PS surveys (https://openpanelalliance.org/pricing.php).

The results reveal a mixed picture as the presence and amount of cost savings depend on the prior structure and the NPS survey used. Figure 5 shows the distribution of percent savings(+)/losses(−) for the smoking outcome variable across the nine NPS surveys for different PS sizes (50, 100, 150) and for the Distance, Distance-log, Distance log-10, and Power priors. Figures C.1 and C.2 in the supplementary data online show the same plots for the always vote and neighborhood trust outcomes, respectively. An interactive visualization is available for all outcome variables, PS sizes, and priors in the Shiny web app under (*Real Data Analysis/Cost Analysis/Savings Distribution*). Looking at the three figures, it is evident that for a PS size of 50, the Power prior nearly always leads to higher cost savings compared to the other strongly informative priors. The Distance-log prior also leads to cost savings in most cases. The Distance and the Distance-log10 priors always lead to losses. However, as the PS size increases ($n_{PS} \geq 100$), the pattern reverses and the Power prior starts to generate losses rather than savings and the performance of the distance priors improves, especially the Distance-log prior which generates the largest cost savings. In general, the median level of cost savings for the mixed-distance priors is lower than for the nonmixed formulations, except for the healthcare and neighborhood trust outcomes, where the median savings are similar for both prior types or slightly higher for the mixed priors.

Table 4 summarizes, for each PS size, the prior formulation (and corresponding NPS survey) that leads to the largest cost savings (in percent) for the selected outcomes: smoking, neighborhood trust, and always vote. Results for

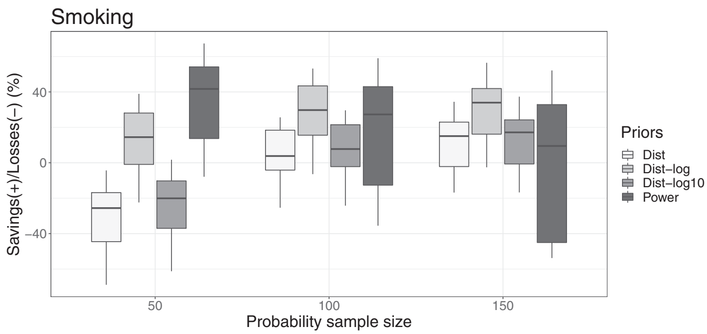

Figure 5. Percentage Cost Savings(+)/Losses(−) for the Smoking Outcome.

the additional outcomes are available in the Shiny web app under (*Real Data Analysis/Cost Analysis/Max. Savings*). For small PS sizes (50–100), the Power prior yields the largest percent cost savings across the three outcomes (range: 37–68 percent). In the case of larger PS sizes (150–500), the distance priors, specifically the Distance-log (range: 42–58 percent) and Mixed-Distance-log (56–61 percent) priors produce the largest cost savings for the three outcomes. Thus, there are indications of a potential cost savings (up to 68 percent) by integrating the PS and NPS surveys using informative priors under assumed per-respondent costs. However, whether savings occur and the amount of those savings varies depending on which NPS survey is used to construct the prior.

Table C.1 in the supplementary data online displays the sensitivity analysis for the largest cost savings according to different cost structures and for all outcome variables. Specifically, we consider situations where the cost per PS unit is between two and six times larger than the NPS unit cost. The results indicate that when the PS costs are at least three times larger than the NPS costs, then the majority of the best-performing strongly informative priors yield significant cost savings. However, when the PS costs are only twice as large as the NPS costs, then estimated cost savings tend to become marginal or negative.

7. DISCUSSION

NPS surveys have received a lot of attention recently, mainly due to the convenience, timeliness, and cost-effectiveness of online access panels and the availability of *big data* sources. However, their well-known risks of selection bias and their generally inferior accuracy compared to PS surveys is why researchers often view them as a supplement, rather than as a replacement, for PS surveys (Cornesse et al. 2020). However, conducting large-scale PS surveys can be prohibitively expensive for researchers working with modest budgets. For researchers who wish to carry out a PS survey but cannot afford a

Table 4. Best-Performing Strongly Informative Priors in Terms of Percent Cost Savings for a Selection of Outcome Variables and Different PS Sizes

PS size	Results	Smoking	Neighborhood trust	Always vote
50	Strongly inf. prior	Power	Power	Power
	NPS survey	NP-A	NP-I	NP-H
	Exp. cost (base. prior)	$20,250	$20,433	$10,359
	Blended cost (inf. prior)	$6,610	$6,500	$6,535
	Savings %	67.36	68.19	36.91
100	Strongly inf. prior	Power	Power	Dist-log
	NPS survey	NP-A	NP-I	NP-H
	Exp. cost (base. prior)	$19,807	$19,746	$11,516
	Exp. cost (best prior)	$8,110	$8,000	$8,035
	Savings %	59.05	59.48	30.23
150	Strongly inf. prior	Dist-log	Mixed-Dist-log	Dist-log
	NPS survey	NP-A	NP-H	NP-H
	Exp. cost (base. prior)	$22,075	$23,998	$18,234
	Exp. cost (best prior)	$9,610	$9,535	$9,535
	Savings %	56.47	60.27	47.71
200	Strongly inf. prior	Dist-log	Mixed-Dist-log	Dist-log
	NPS survey	NP-A	NP-H	NP-H
	Exp. cost (base. prior)	$26,710	$28,809	$23,173
	Exp. cost (best prior)	$11,110	$11,035	$11,035
	Savings %	58.4	61.27	52.38
300	Strongly inf. prior	Dist-log	Mixed-Dist-log	Dist-log
	NPS survey	NP-A	NP-H	NP-H
	Exp. cost (base. prior)	$30,677	$34,932	$27,357
	Exp. cost (best prior)	$14,110	$14,035	$14,035
	Savings %	54.00	59.82	48.70
500	Strongly inf. prior	Dist-log	Mixed-Dist-log	Dist-log
	NPS survey	NP-A	NP-I	NP-H
	Exp. cost (base. prior)	$36,352	$45,870	$34,732
	Exp. cost (best prior)	$20,110	$20,000	$20,035
	Savings %	44.68	56.4	42.32

large sample size, a potential compromise is to carry out a PS survey with small sample size and a parallel supplementary NPS survey that collects the same analysis variables and use Bayesian data analysis methods to integrate both data sources. This approach has been proposed by previous authors for the case of modeling continuous outcome variables using linear regression (Sakshaug et al. 2019; Wiśniowski et al. 2020; Nandram and Rao 2021), with some of these studies demonstrating that this approach produces coefficient estimates with smaller MSEs compared to PS-only surveys and at lower costs.

The current study extends this previous work and contributes to this emerging research area by proposing a Bayesian methodology for analyzing binary outcome variables collected from parallel PS and NPS surveys using logistic regression, which is of great importance in the social sciences for studying attitudes, behaviors, and characteristics of populations. Specifically, we proposed several novel strongly informative priors that exploit auxiliary information from a parallel NPS survey data to improve coefficient estimates from small-sized PS surveys.

The strongly informative priors were evaluated through a simulation study which showed that they achieve smaller MSEs for coefficient estimates compared to those achieved exclusively using PS survey data with a weakly informative baseline prior. This was particularly true for the case of no (or low) selection bias in the NPS data and for PS sizes less than 200, where the Distance-log and the Power priors produced the smallest MSEs, effectively driving down the posterior variance. In the case of large selection bias, the strongly informative priors performed mostly similarly to the baseline prior with the Distance prior being superior to the other strongly informative prior formulations for small PS sizes. We then evaluated the approach in a real-data application by modeling six binary outcomes from an actual PS survey and several parallel NPS surveys reflecting different selection scenarios one might face in practice. The strongly informative priors again yielded significant reductions in MSEs, compared to the baseline prior, especially for the smaller PS sample sizes. The Power prior was superior to the other strongly informative priors in terms of MSE reduction for very small sample sizes (50–100), whereas the Distance-log prior and its mixed version performed better for slightly larger sample sizes (150–200). For PS sizes larger than 200, all of the strongly informative priors performed similarly to the baseline prior with respect to MSEs.

An important novelty of the method lies in its ability to achieve the same MSE values as would a larger (and likely more expensive) PS-only survey at a potentially lower cost. Using assumed but realistic cost data for the parallel PS and NPS surveys, we showed indications of potential cost savings for the informative priors for different PS sizes. In general, for a PS survey of 50–100 respondents, the Power prior showed high potential cost savings (up to 68 percent), while the Distance-log and its mixed version were among the best performers for larger PS sizes (achieving potential cost savings of up to about 60 percent). Thus, researchers with low-to-moderate budgets may benefit from using the proposed data integration strategy to minimize both costs and errors.

As a general recommendation for practitioners, the Power prior appears to be the most appropriate choice for small PS sizes (up to 100 observations) and the Distance-log and Mixed-Distance-log priors for larger PS sizes (up to 500 respondents). These recommendations hold regardless of the outcome variable and the NPS survey considered. Nevertheless, we recommend performing a sensitivity analysis and comparing estimates obtained using different priors.

The present study entails some limitations and areas for future work. First, the method rests on the assumption that the PS survey is unbiased or less biased than the parallel NPS survey. This assumption may not always hold in practice. In addition, the method does not account for measurement errors, which may differ between PS and NPS survey data (Einarsson et al. 2022). We leave these topics for future work. Moreover, it would be worthwhile to extend the current framework to other types of categorical variables (e.g., multinomial, ordinal) and account for complex sample design features (e.g., stratification). The approach may also be extended in a multivariate setting by taking into account the distributions of several outcomes simultaneously. Exploring alternative methods for selecting the Power parameter is another topic for future development.

In conclusion, it's becoming increasingly common for survey researchers who wish to carry out a PS survey to face budgetary constraints or timeliness considerations that preclude fielding large sample sizes, while fielding small samples will yield large variances for survey estimates. To tackle this issue and improve inference while maintaining cost-effectiveness, researchers can design studies that field a PS survey with small-sample size and supplement it with a parallel NPS survey under the proposed framework, which can result in improved inference (reduced variances and MSEs) and potentially lower survey costs.

The proposed data integration method can be easily implemented in any statistical software, which supports Bayesian computation. To assist researchers, an R Shiny web app has been developed which provides the replication code for applying the methodology and allows readers to browse the full results of the simulation and application in more detail (see appendix D in the supplementary data online). In addition, a key feature of the Shiny app is the possibility to dynamically implement the cost analysis with user-entered PS and NPS cost data. This may be useful for practitioners interested in designing and integrating parallel PS and NPS surveys and wish to compare the results under different cost scenarios.

Supplementary Materials

Supplementary materials are available online at academic.oup.com/jssam.

REFERENCES

Alexander, M., Polimis, K., and Zagheni, E. (2022), "Combining Social Media and Survey Data to Nowcast Migrant Stocks in the United States," *Population Research and Policy Review*, 41, 1–28.

Astley, C. M., Tuli, G., Mc Cord, K. A., Cohn, E. L., Rader, B., Varrelman, T. J., Chiu, S. L., Deng, X., Stewart, K., Farag, T. H., Barkume, K. M., LaRocca, S., Morris, K. A., Kreuter, F., and Brownstein, J. S. (2021), "Global Monitoring of the Impact of the Covid-19 Pandemic through Online Surveys Sampled from the Facebook User Base," *Proceedings of the National Academy of Sciences*, 118, 1–10.

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Dillman, D., Frankel, M. R., Garland, P., Groves, R. M., Kennedy, C., Krosnick, J., Lavrakas, P. J., Lee, S., Link, M., Piekarski, L., Rao, K., Thomas, R. K., and Zahs, D; Prepared for the AAPOR Executive Council by a Task Force operating under the auspices of the AAPOR Standards Committee, with members including: (2010), "Research Synthesis: Aapor Report on Online Panels," *Public Opinion Quarterly*, 74, 711–781.

Beaumont, J.-F. (2020), "Are Probability Surveys Bound to Disappear for the Production of Official Statistics?," *Survey Methodology*, 46, 1–29.

Beaumont, J.-F., and Rao, J. (2021), "Pitfalls of Making Inferences from Non-Probability Samples: Can Data Integration through Probability Samples Provide Remedies?" *The Survey Statistician*, 83, 11–22.

Berzofsky, M. E., McKay, T., Hsieh, Y. P., and Smith, A. (2018), "Probability-Based Samples on Twitter: Methodology and Application," *Survey Practice*, 11, 4936.

Bethlehem, J. (2010), "Selection Bias in Web Surveys," *International Statistical Review*, 78, 161–188.

———. (2009), "The Rise of Survey Sampling," Discussion Paper 09015, Statistics Netherlands, The Hague/Heerlen.

Biemer, P. P. (2010), "Total Survey Error: Design, Implementation, and Evaluation," *Public Opinion Quarterly*, 74, 817–848.

Biffignandi, S., and Bethlehem, J. (2021), *Handbook of Web Surveys*, Hoboken, NJ: John Wiley & Sons, Inc.

Brick, J. M. (2011), "The Future of Survey Sampling," *Public Opinion Quarterly*, 75, 872–888.

Callegaro, M., Villar, A., Yeager, D. S., and Krosnick, J. A. (2014), "A Critical Review of Studies Investigating the Quality of Data Obtained with Online Panels Based on Probability and Nonprobability Samples," in *Online Panel Research: A Data Quality Perspective*, eds. M. Callegaro, R. Baker, J. Bethlehem, A.S. Göritz, J.A. Krosnick, and P.J. Lavrakas, Hoboken, NJ: John Wiley & Sons, Inc., pp. 23–53.

Chen, M.-H., Ibrahim, J. G., and Shao, Q.-M. (2000), "Power Prior Distributions for Generalized Linear Models," *Journal of Statistical Planning and Inference*, 84, 121–137.

Chen, M.-H., Ibrahim, J. G., and Yiannoutsos, C. (1999), "Prior Elicitation, Variable Selection and Bayesian Computation for Logistic Regression Models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 223–242.

Cornesse, C., Blom, A. G., Dutwin, D., Krosnick, J. A., De Leeuw, E. D., Legleye, S., Pasek, J., Pennay, D., Phillips, B., Sakshaug, J. W., Struminskaya, B., and Wenz, A. (2020), "A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research," *Journal of Survey Statistics and Methodology*, 8, 4–36.

Couper, M. P. (2013), "Is the Sky Falling? new Technology, Changing Media, and the Future of Surveys," *Survey Research Methods*, 7, 145–156.

De Santis, F. (2006), "Power Priors and Their Use in Clinical Trials," *The American Statistician*, 60, 122–129.

De Santis, F., and Gubbiotti, S. (2023), "Borrowing Historical Information for Non-Inferiority Trials on Covid-19 Vaccines," *The International Journal of Biostatistics*, 19, 177–189.

Dimock, M., Doherty, C., Kiley, J., and Oates, R. (2014), "Political Polarization in the American Public." Technical Report, Washington, DC: Pew Research Center.

DiSogra, C., Cobb, C., Chan, E., and Dennis, J. M. (2011), "Calibrating Non-Probability Internet Samples with Probability Samples Using Early Adopter Characteristics," in *JSM Proceedings, Survery Research Methods Section*. Alexandria, VA: American Statistical Association, pp. 4501–4515.

Dutwin, D., and Buskirk, T. D. (2017), “Apples to Oranges or Gala versus Golden Delicious?: Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples,” *Public Opinion Quarterly*, 81, 213–239.

Einarsson, H., Sakshaug, J. W., Cernat, A., Cornesse, C., and Blom, A. G. (2022), “Measurement Equivalence in Probability and Nonprobability Online Panels,” *International Journal of Market Research*, 64, 484–505.

Elliot, M. R. (2009), “Combining Data from Probability and Non-Probability Samples Using Pseudo-Weights,” *Survey Practice*, 2, 1–7.

Elliott, M. R., and Valliant, R. (2017), “Inference for Nonprobability Samples,” *Statistical Science*, 32, 249–264.

Ganesh, N., Pineau, V., Chakraborty, A., and Dennis, J. M. (2017), “Combining Probability and Non-Probability Samples using Small Area Estimation,”in *JSM Proceedings*, *Survery Research Methods Section*. Alexandria, VA: American Statistical Association, pp. 1657–1667.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008), “A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models,” *The Annals of Applied Statistics*, 2, 1360–1383.

Ghosh, J., Li, Y., and Mitra, R. (2018), “On the Use of Cauchy Prior Distributions for Bayesian Logistic Regression,” *Bayesian Analysis*, 13, 359–383.

Goodrich, B., Gabry, J., Ali, I., and Brilleman, S. (2020), “Rstanarm: Bayesian Applied Regression Modeling via Stan,” R Package Version 2.21.1.

Haddad, T., Himes, A., Thompson, L., Irony, T., Nair, R., Modeling, M. C., and Participants, S. W. G.; MDIC Computer Modeling and Simulation Working Group Participants (2017), “Incorporation of Stochastic Engineering Models as Prior Information in Bayesian Medical Device Trials,” *Journal of Biopharmaceutical Statistics*, 27, 1089–1103.

Hillygus, D. S., Jackson, N., and Young, M. (2014), “Professional Respondents in Nonprobability Online Panels,” in *Online Panel Research: A Data Quality Perspective*, eds. M. Callegaro, R. Baker, J. Bethlehem, A.S. Göritz, J.A. Krosnick, and P.J. Lavrakas, Hoboken, NJ: John Wiley & Sons, Inc., pp. 219–237.

Hsiao, Y., Fiorio, L., Wakefield, J., Zagheni, E. (2020), “Modeling the Bias of Digital Data: An Approach to Combining Digital and Survey Data to Estimate and Predict Migration Trends.” Technical Report, Max Planck Institute for Demographic Research, Rostock, Germany.

Ibrahim, J. G., and Chen, M.-H. (2000), “Power Prior Distributions for Regression Models,” *Statistical Science*, 15, 46–60.

Ibrahim, J. G., Chen, M.-H., Xia, H. A., and Liu, T. (2012), “Bayesian Meta-Experimental Design: Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes,” *Biometrics*, 68, 578–586.

Johnson, R. M., Crifasi, C., Goodell, E. M. A., Wiśniowski, A., Sakshaug, J. W., Thrul, J., and Owens, M. (2021), “Differences in Beliefs about Covid-19 by Gun Ownership: A Cross-Sectional Survey of Texas Adults,” *BMJ Open*, 11, e048094.

Keeter, S. (2019), “Growing and Improving Pew Research Center’s American Trends Panel.” Technical Report, Washington, DC: Pew Research Center.

Kennedy, C., Mercer, A., Keeter, S., Hatley, N., McGeeney, K., and Gimenez, A. (2016), “Evaluating Online Nonprobability Surveys.” Technical Report, Washington, DC: Pew Research Center.

Kim, J. K., Park, S., Chen, Y., and Wu, C. (2021), “Combining Non-Probability and Probability Survey Samples through Mass Imputation,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184, 941–963.

Kim, J.-K., and Tam, S.-M. (2021), “Data Integration by Combining Big Data and Survey Sample Data for Finite Population Inference,” *International Statistical Review*, 89, 382–401.

Kish, L. (1965), *Survey Sampling*, New York: John Wiley and Sons, Inc.

Kreuter, F., Barkay, N., Bilinski, A., Bradford, A., Chiu, S., Eliat, R., Fan, J., Galili, T., Haimovich, D., Kim, B., LaRocca, S., Li, Y., Morris, K., Presser, S., Salomon, J. A., Sarig, T., Stewart, K., Stuart, E. A., and Tibshirani, R. (2020), “Partnering with Facebook on a University-Based Rapid Turn-around Global Survey,” *Survey Research Methods: SRM*, 14, 159–163.

Luiten, A., Hox, J., and de Leeuw, E. (2020), "Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys," *Journal of Official Statistics*, 36, 469–487.

Malhotra, N., and Krosnick, J. A. (2007), "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 Anes to Internet Surveys with Nonprobability Samples," *Political Analysis*, 15, 286–323.

Miller, P. V. (2017), "Is There a Future for Surveys?," *Public Opinion Quarterly*, 81, 205–212.

Nandram, B., and Rao, J. (2021), "A Bayesian Approach for Integrating a Small Probability Sample with a Non-Probability Sample," in *Proceedings of the American Statistical Association (JSM2021 - Virtual Conference)*, pp. 1568–1603.

Neyman, J. (1934), "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, 97, 558–625.

Pasek, J. (2016), "When Will Nonprobability Surveys Mirror Probability Surveys? considering Types of Inference and Weighting Strategies as Criteria for Correspondence," *International Journal of Public Opinion Research*, 28, 269–291.

Pew Research Center (2014a), "American Trends Panel Wave 5." Available at https://www.pewresearch.org/politics/dataset/american-trends-panel-wave-5/.

Pew Research Center (2014b), "American Trends Panel Wave 7." Available at https://www.pewresearch.org/politics/dataset/american-trends-panel-wave-7/.

Pew Research Center (2015a), "American Trends Panel Wave 10." Available at https://www.pewresearch.org/politics/dataset/american-trends-panel-wave-10/.

Pew Research Center (2015b), "Online nonprobability landscape study." Available at https://www.pewresearch.org/methods/dataset/online-nonprobability-landscape-study/.

R Core Team (2020), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing.

Raghunathan, T., Ghosh, K., Rosen, A., Imbriano, P., Stewart, S., Bondarenko, I., Messer, K., Berglund, P., Shaffer, J., and Cutler, D. (2020), "Combining Information from Multiple Data Sources to Assess Population Health," *Journal of Survey Statistics and Methodology*, 9, 598–625.

Rao, J. (2021), "On Making Valid Inferences by Integrating Data from Surveys and Other Sources," *Sankhya B*, 83, 242–272.

Robbins, M. W., Ghosh-Dastidar, B., and Ramchand, R. (2021), "Blending Probability and Nonprobability Samples with Applications to a Survey of Military Caregivers," *Journal of Survey Statistics and Methodology*, 9, 1114–1145.

RStudio, Inc (2022), "shiny: Easy Web Applications in r." Available at http://shiny.rstudio.com.

Sakshaug, J. W., Wiśniowski, A., Ruiz, D. A. P., and Blom, A. G. (2019), "Supplementing Small Probability Samples with Nonprobability Samples: A Bayesian Approach," *Journal of Official Statistics*, 35, 653–681.

Stan Development Team (2019), "Stan Modeling Language Users Guide and Reference Manual," Version, 2.29.

Stan Development Team (2021), "RStan: The R Interface to Stan," R Package Version 2.21.3.

Stier, S., Breuer, J., Siegers, P., and Thorson, K. (2020), "Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field," *Social Science Computer Review*, 38, 503–516.

Thompson, A. J., and Pickett, J. T. (2020), "Are Relational Inferences from Crowdsourced and Opt-in Samples Generalizable? comparing Criminal Justice Attitudes in the GSS and Five Online Samples," *Journal of Quantitative Criminology*, 36, 907–932.

Thompson, L., Chu, J., Xu, J., Li, X., Nair, R., and Tiwari, R. (2021), "Dynamic Borrowing from a Single Prior Data Source Using the Conditional Power Prior," *Journal of Biopharmaceutical Statistics*, 31, 403–424.

Valliant, R., and Dever, J. A. (2011), "Estimating Propensity Adjustments for Volunteer Web Surveys," *Sociological Methods & Research*, 40, 105–137.

Wang, J., Zhang, H., and Tiwari, R. (2023), "A Propensity-Score Integrated Approach to Bayesian Dynamic Power Prior Borrowing," *Statistics in Biopharmaceutical Research*, 1–10. DOI: 10.1080/19466315.2023.2223533.

West, B. T., Little, R. J., Andridge, R. R., Boonstra, P. S., Ware, E. B., Pandit, A., and Alvarado-Leiton, F. (2021), "Assessing Selection Bias in Regression Coefficients Estimated from Nonprobability Samples with Applications to Genetics and Demographic Surveys," *The Annals of Applied Statistics*, 15, 1556–1581.

Wiśniowski, A., Sakshaug, J. W., Perez Ruiz, D. A., and Blom, A. G. (2020), "Integrating Probability and Nonprobability Samples for Survey Inference," *Journal of Survey Statistics and Methodology*, 8, 120–147.

Yang, S., Kim, J. K., and Song, R. (2020), "Doubly Robust Inference When Combining Probability and Non-Probability Samples with High Dimensional Data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82, 445–465.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., and Wang, R. (2011), "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples," *Public Opinion Quarterly*, 75, 709–747.

Yee, T. W. (2021), "On the Hauck-Donner Effect in Wald Tests: Detection, Tipping Points, and Parameter Space Characterization," *Journal of the American Statistical Association*, 117, 1763–1774.