Bayesian Predictive Inference with Survey Weights for Binary Response: A Simulation Study and a Numerical Example

Lingli Yang* Balgobin Nandram†

September 27, 2023

Abstract

We consider the problem of Bayesian predictive inference for binary response with covariates and survey weights. Our method makes use of the combination of probability survey samples that have been enhanced by auxiliary data. The incorporation of survey weights into a logistic regression model, which creates a thorough and logical analytical paradigm, is at the core of our methodology. Our investigation covers six different models that were carefully created to include both normalized and unnormalized weighted likelihoods. Three iterations of adjusted survey weights—original, trimmed, and calibrated—are taken into account within this spectrum. The Metropolis-Hastings sampler is the implementation algorithm for our analysis. Building on this foundation, we use the stratification and surrogate sampling technique to expand our inferences to finite population parameters. We conduct a thorough evaluation that includes a simulation study and a real-world dataset focused on body mass index (BMI) in order to assess the performance and efficacy of our models. Our findings show how powerful models with normalized density functions and adjusted trimmed weights are. These models exhibit a unique capability for higher estimation accuracy while maintaining fidelity to the fundamental principles of Bayesian inference. The results of our study have broad implications for the field of research as a whole, highlighting the significance of the framework we proposed and the exceptional value of trimmed weights that have been adjusted for the purpose of driving effective predictive inference in survey-oriented research studies.

Keywords: Survey Weights, Bayesian Predictive Inference, Body Mass Index

*lyang8@wpi.edu, Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609

†balnan@wpi.edu, Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609

1 Introduction

Survey sampling is a fundamental statistical methodology employed to collect reliable and representative data from a subset of a larger population. It serves as a means to reduce costs and capture essential information about key variables of interest. By strategically selecting a representative sample, researchers can make informed predictions and inferences based on the gathered data. This method is crucial for accurate estimation and forecasting at the population level. Survey weights play a pivotal role in the domain of survey sampling, addressing potential biases present within the sample data. Their purpose is to rectify any discrepancies between the sample and the target population, ensuring that inferences and estimates derived from the sample data are both precise and unbiased (Lohr, 2021). However, improper utilization of survey weights can lead to skewed estimates and misguided inferences. This has been acknowledged in previous research, where complexities in employing survey weights were highlighted (Gelman, 2007). Nevertheless, Lohr (2007) countered that while there are limitations to the utility of survey weights, diligent efforts should be made to maximize their potential.

Generating survey weights involves assigning weights to each unit within the sample, often through the inverse of the probability of selection. This correction compensates for potential bias introduced by the sampling design. For instance, groups with lower probabilities of selection receive higher weights to ensure the sample reflects the broader population. In survey datasets, weights can also encompass adjustments for non-response and post-stratification to align auxiliary variable distributions with known population distributions (Chen et al., 2017). Notably, survey weights can even be generated from a single nonprobability sample, as demonstrated by Elliott and Valliant (2017) and Chen et al. (2020). Robbins et al. (2021) concentrated on design-based methods such as calibration and propensity score weighting (without parametric models). Nandram and Rao (2021), Nandram et al. (2021) showed a full Bayesian approach to incorporate a nonprobability sample and a probability sample. Yang et al. (2023) concluded the performance of nine methods but they did not take auxiliary information into consideration. Despite this, our focus remains on evaluating the performance of various adjusted weights within a parametric model context, rather than on the estimation of survey weights themselves.

When dealing with binary study variables, Bayesian logistic regression emerges as a prominent probabilistic model. This framework elucidates the relationship between covariates and the binary response variable through Bayesian inference, offering posterior distributions of model parameters based on the observed data. Unlike traditional logistic regression, Bayesian logistic regression incorporates parameter uncertainty, yielding distributions over possible parameter values. This enables probabilistic predictions and quantification of prediction uncertainty. While weighting units can mitigate the impact of unequal inclusion in the sample, the estimates often remain inefficient. Covariates serve not only as auxiliary information but also as crucial components for modeling, specifically in regression and the corresponding coefficient estimation. Numerous studies such as Roberts et al. (1987), Archer and Lemeshow (2006), Rader et al. (2017), and Chen and Nandram (2023) have explored different aspects of incorporating survey designs into logistic regression analysis. The same idea is shown in the work of Barasa and Muchwanju (2015), but they only analyzed the regression coefficients, rather than making predictions and inferences about a finite population quantity.

This paper builds upon and refines the findings presented in Yang and Nandram (2023), providing an updated and enhanced analysis of the simulation study and the numerical example. In this study, we aim to evaluate the effectiveness of six distinct models for Bayesian predictive inference on finite population parameters. Section 2 delves into the nuances of adjusted survey weights, encompassing adjusted original, adjusted trimmed, and adjusted calibrated survey weights. These weights are then integrated into both unnormalized and normalized models to facilitate Bayesian predictive inference. Section 3 offers insight from a simulation study, leveraging metrics such as absolute relative bias (ARB), posterior standard deviation (PSD), posterior root mean squared error (PRMSE), coverage probability (CP), and width of highest posterior density interval (Wid) to evaluate the six models. In Section 4, we apply the models to real-world data on body mass index (BMI) from the Third National Health and Nutrition Examination Survey (NHANES III), investigating posterior means and posterior standard deviations. Lastly, Section 5 provides recommendations for further research while assessing the study's strengths and limitations comprehensively. Technical methods are presented in the appendix.

2 Bayesian Predictive Inference

2.1 Data Source and Objectives

In this section, we describe the objectives of our study and introduce the body mass index (BMI) dataset that serves as the focal point of our analysis. Our study revolves around the comprehensive Bayesian predictive inference framework for binary responses. To achieve this, we utilize probability survey samples that are enriched with auxiliary data. Our primary objective is to develop and evaluate a set of models for conducting Bayesian predictive inference on finite population parameters, taking into account the complexities introduced by survey weights and covariates.

To illustrate the efficacy of the proposed framework, we employ the Third National Health and Nutrition Examination Survey (NHANES III) dataset. NHANES III is a nationally representative survey conducted by the National Center for Health Statistics (NCHS) to assess the health and nutritional status of the U.S. population. This dataset is particularly suitable for our study due to its extensive collection of demographic, health, and lifestyle information. The NHANES III dataset contains measurements of various health-related variables, including binary responses, covariates, and survey weights. Of particular interest to our study is the body mass index (BMI) data, which serves as a binary response variable indicating the presence or absence of being obese. The dataset also includes covariates that provide valuable auxiliary information for our modeling, enabling us to account for potential confounding factors. By utilizing the NHANES III dataset and focusing on BMI as the binary response variable, we aim to showcase the practical applicability and effectiveness of our proposed framework in addressing real-world survey-based studies involving binary responses, survey weights, and covariates. In the subsequent sections of this paper, we delve into the details of how the Bayesian predictive inference framework is applied to the BMI dataset, the model comparisons, and the implications of our findings.

2.2 Adjusted Survey Weights

In this section, we establish a comprehensive framework for incorporating survey weights into the statistical inference of binary variables from probability survey samples, all while harnessing the power of pertinent auxiliary data. Our approach encompasses three variations of adjusted survey weights: adjusted original weights, adjusted trimmed weights, and adjusted calibrated weights. We begin by outlining the foundational concepts and methodologies underlying each variation. We initiate our exploration by considering the fundamental concept of adjusted original weights. Within this paradigm, we focus on the scenario where the selection probability exclusively depends on the observed covariates (i.e., ignorable selection). Imagine drawing a sample of size n from a finite population with size N. For each unit i, let y_i represent the response, $\underset{\sim}{x_i}$ symbolize the vector of observed covariates, and W_i denote the corresponding (original) survey weight. To estimate the unknown finite population size (N) along with the population total (T), the classical Horvitz-Thompson estimators are defined as $\hat{T} = \sum_{i=1}^{n} W_i y_i$, $\hat{N} = \sum_{i=1}^{n} W_i$.

To further enhance the robustness of our methodology, we delve into the concept of trimmed weights. Inspired by the Winsorization technique, which mitigates the impact of outliers, we seek to alleviate the influence of extreme values in our survey weights. By rescaling the weights and adopting a threshold approach, we curtail the undue impact of outliers, leading to more resilient and reliable predictive inferences; see Rao (1966) and Basu (1971). In the context of our work, we opt for a threshold defined as $W_0 = Q_3 + 1.5(Q_3 - Q_1)$, where Q_1 signifies the first quartile and Q_3 signifies the third quartile of the data distribution. This calibrated threshold ensures that we encapsulate the essence of the data's dispersion while judiciously mitigating the undue influence of extreme values. With this threshold in place, we proceed to transform our original survey weights into their trimmed counterparts, denoted as $\underset{\sim}{W}^*$. This transformation follows a distinct prescription, as detailed by the following formulation:

$$W_i^* = \begin{cases} W_0, & W_i \geq W_0 \\ aW_i, & W_i < W_0 \end{cases}, \tag{1}$$

where a represents a carefully chosen rescaling parameter, imbued with the purpose of preserving the total weights in the adjusted scheme ($\sum_{i=1}^{n} W_i^* = \sum_{i=1}^{n} W_i = \hat{N}$). This judicious rescaling ensures that the recalibrated weights seamlessly integrate into the broader statistical framework, poised to furnish predictive inferences fortified against the perturbing effects of extreme values.

In contrast, the calibrated weights are fashioned through a more intricate optimization process. The process strives to achieve harmony between the aggregate totals and the weighted sum of auxiliary variables. The use of a distance function $G(u)$ and the concept of Lagrange multipliers guides this optimization. The calibrated weights, ultimately determined by the Lagrange multipliers and the distance function, reflect a balanced relationship between auxiliary variables and aggregate totals. Incorporating calibrated survey weights serves as a potent strategy to rectify the under-representation of specific population segments within the sample (Haziza and Beaumont, 2017). This approach proves indispensable when striving to enhance the precision and credibility of prevalence estimates for obesity within the finite population. The calibrated survey weights stand as a hallmark technique within survey research, adeptly addressing the challenges posed by non-representative samples or the under-representation of distinct subgroups.

To embark on this calibration journey, we commence by assuming the availability of aggregate totals, denoted as $\underset{\sim}{t}$. These totals might emanate from sources such as censuses, governmental documents, or meticulously curated online data. The calibration process hinges on the notion of calibrating the survey weights, denoted as $\underset{\sim}{\tilde{W}}$, to align them with the aggregate totals. This alignment is guided by a distance function $G(u)$, wielded to minimize the divergence between the calibrated and original weights. The function $G(u)$ adheres to essential properties: it is non-negative ($G(u) \geq 0$), vanishes at unity ($G(1) = 0$), is differentiable ($G(u)$ has a derivative, denoted as $g(u)$), and is strictly convex. The calibration leverages the relative importance of each unit, as encapsulated by the factor q_j.

To realize this calibration process, we venture into the realm of optimization. Our objective is to minimize a functional form involving both the weighted distance function and a constraint imposed by the calibration equations. The Lagrangian multiplier method emerges as a versatile tool for handling such optimization tasks, with the Lagrange multipliers $\underset{\sim}{\lambda} = (\lambda_1, \ldots, \lambda_p)'$ playing a pivotal role. The formulation takes shape as follows:

$$\underset{\tilde{W}j}{\operatorname{argmin}} \sum_{j=1}^{n} \frac{\tilde{W}_j}{q_j} G\left(\frac{\tilde{W}j}{W_j}\right) \quad \text{s.t.} \quad \sum_{j=1}^{n} \tilde{W}_j x_j = \underset{\sim}{t}. \tag{2}$$

Upon successfully obtaining the Lagrange multipliers $\hat{\underset{\sim}{\lambda}}$, the calibrated weights readily unfold through the relationship:

$$\tilde{W}_j = W_j g^{-1}\left(q_j \hat{\underset{\sim}{\lambda}}' \underset{\sim}{x}_j\right), \quad j = 1, \ldots, n \tag{3}$$

The calibration process embraces a specific choice of the distance function—specifically, the simple Euclidean distance represented by $G(u) = \frac{1}{2}(u-1)^2$. This choice ushers in closed-form solutions, facilitating the practical implementation of the procedure. The calibration process entails pivotal algebraic steps, culminating in the calibration weights:

$$\tilde{W}_j = W_j \left(1 + q_j \hat{\underset{\sim}{\lambda}}' \underset{\sim}{x}_j\right), \quad j = 1, \ldots, n. \tag{4}$$

It is noteworthy that under specific conditions ($\hat{\underset{\sim}{\lambda}} = \underset{\sim}{0}$), calibrated weights coincide with original survey weights. The intricacies of the calibration process can, however, yield negative calibrated weights, particularly when the sample fails to accurately represent the population, owing to non-response bias, sampling error, or inappropriate weighting adjustments. To navigate this challenge, we adopt a practical approach of rescaling negative weights to be positive. Specifically, weights less than 1 are set to 1 to uphold a representative minimum size. This rescaling is followed by normalization to preserve the integrity of the sample size. By this means, we ensure that the calibrated weights are primed to fulfill their role as stalwart pillars of accurate predictive inferences.

The effective sample size serves as a pivotal metric, quantifying the degree of independent information within a dataset. In the realm of statistics, it acts as a corrective measure, addressing the potential discrepancies between observed sample size and the actual independence within the data. This adjustment accounts for various complexities, including interdependence between observations, clustering effects, and non-random sampling techniques (Potthoff et al., 1992). A larger effective sample size signifies a wealth of independent information, leading to more precise estimations and narrower confidence intervals. Conversely, a smaller effective sample size implies a dearth of independent information, potentially resulting in less precise estimates and wider confidence intervals. Denoted by n_e, the effective sample size anchors our adjustment process. We embark on constructing three types of adjusted weights for each unit, indexed by $i = 1, \ldots, n$:

$$n_e = \frac{\left(\sum_{j=1}^{n} W_j\right)^2}{\sum_{j=1}^{n} W_j^2}, \quad w_i = \frac{n_e W_i}{\sum_{j=1}^{n} W_j}, \tag{5}$$

$$n_e^* = \frac{\left(\sum_{j=1}^n W_j^*\right)^2}{\sum_{j=1}^n W_j^{*2}}, \quad w_i^* = \frac{n_e^* W_i^*}{\sum_{j=1}^n W_j^*}, \tag{6}$$

$$\tilde{n}_e = \frac{\left(\sum_{j=1}^n \tilde{W}_j\right)^2}{\sum_{j=1}^n \tilde{W}_j^2}, \quad \tilde{w}_i = \frac{\tilde{n}_e \tilde{W}_i}{\sum_{j=1}^n \tilde{W}_j}. \tag{7}$$

By introducing these adjustments, we elevate the accuracy and reliability of our methodology. The refined weights - adjusted original weights $\underset{\sim}{w}$, adjusted trimmed weights $\underset{\sim}{w}^*$, and adjusted calibrated weights $\underset{\sim}{\tilde{w}}$ - surpass their unadjusted counterparts in terms of precision. Notably, the use of adjusted trimmed weights is instrumental in curbing the undue influence of outlier observations, rendering the methodology more robust against data that may deviate from conventional assumptions. These assumptions include the normality of the data distribution or the homogeneity of variance. On the other hand, calibrated weights are designed to rectify discrepancies in variables of interest, such as age, gender, or education level, thereby enhancing the representativeness of the survey data. This dual approach of trimmed and calibrated weights forms a cohesive strategy that fortifies the validity and trustworthiness of our predictive inferences.

2.3 Weighted Density

In this study, we focus on the binary response, and the logistic regression model with p covariates is used to model the binary response. To be more specific, the following logistic regression model for the population is considered,

$$y_i \mid \underset{\sim}{\beta} \overset{ind}{\sim} \text{Bernoulli}\left\{\frac{e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}{1+e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}\right\}, i = 1, \ldots, N, \tag{8}$$

where N, the population size, and the nonsampled $\underset{\sim}{x}_i$ may be unknown, and these can come from an external source or a calibrated bootstrap of the sampled covariates. Note that there are no survey weights in the population model. Suppose that a probability sample of size n is taken from a finite population. With given survey weights of each unit in the sample, $\underset{\sim}{W}$, the Horvitz-Thompson estimator of the loglikelihood for the entire population, $L(\underset{\sim}{\theta})$, is

$$\widehat{L(\underset{\sim}{\theta})} = \sum_{i=1}^n W_i \log\left\{f(y_i|\underset{\sim}{\theta})\right\},$$

where we use W_i to illustrate our models but replace the weights with adjusted survey weights w_i in the application. This is also called pseudo-likelihood function, which means $g(y_i|\theta) \propto f(y_i \mid \underset{\sim}{\theta})^{W_i}, i = 1, \ldots, n$, where $g(y_i \mid \underset{\sim}{\theta})$ is the density function of y_i. Note that $f(y_i \mid \underset{\sim}{\theta})^{W_i}$ is an unnormalized 'density' function. To obtain the normalized one, we need to insert the normalization constant, $\int f(y_i \mid \underset{\sim}{\theta})^{W_i} dy_i$, such that $g(y_i \mid \underset{\sim}{\theta})$ integrates to 1. Therefore, we have,

$$g(y_i|\underset{\sim}{\theta}) = \frac{(f(y_i|\underset{\sim}{\theta}))^{W_i}}{\int (f(y_i|\underset{\sim}{\theta}))^{W_i} dy_i}.$$

For the sample, we have $(W_i, \underset{\sim}{x}_i, y_i), i = 1, \ldots, n$. Without the normalization,

$$f_1\left(y_i \mid \underset{\sim}{\beta}\right)=\left(\frac{e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}{1+e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}\right)^{W_i y_i}\left(\frac{1}{1+e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}\right)^{W_i(1-y_i)}, i=1, \ldots, n, \tag{9}$$

independently, but note that this is not a probability mass function in y_i, even though it is typically employed in model-based analysis of data derived from survey sampling. With normalization,

$$\begin{aligned}
f_2(y_i|\underset{\sim}{\theta}) = & \frac{(f_1\left(y_i|\underset{\sim}{\theta}\right))^{W_i}}{f_1\left(y_i=0|\underset{\sim}{\theta}\right))^{W_i}+f_1\left(y_i=1|\underset{\sim}{\theta}\right))^{W_i}} && (10)\\
= & \frac{\left(\frac{e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}{1+e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}\right)^{y_i W_i}\left(\frac{1}{1+e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}\right)^{(1-y_i) W_i}}{\left(\frac{e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}{1+e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}\right)^{W_i}+\left(\frac{1}{1+e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}}}\right)^{W_i}} && (11)\\
= & \left(\frac{e^{W_i \underset{\sim}{x}_i' \underset{\sim}{\beta}}}{1+e^{W_i \underset{\sim}{x}_i' \underset{\sim}{\beta}}}\right)^{y_i}\left(\frac{1}{1+e^{W_i \underset{\sim}{x}_i' \underset{\sim}{\beta}}}\right)^{(1-y_i)}. && (12)
\end{aligned}$$

Therefore,

$$y_i \mid \underset{\sim}{\beta} \overset{ind}{\sim} \text{Bernoulli}\left\{\frac{e^{W_i \underset{\sim}{x}_i' \underset{\sim}{\beta}}}{1+e^{W_i \underset{\sim}{x}_i' \underset{\sim}{\beta}}}\right\}, i=1, \ldots, n. \tag{13}$$

Once more, the Bayesian paradigm favors the normalized form.

Then, using a flat prior on $\underset{\sim}{\beta}$, $\pi(\underset{\sim}{\beta})=1$, the joint pseudo-posterior density and the joint posterior density are,

$$\pi_1(\underset{\sim}{\beta} \mid \underset{\sim}{y}) \propto\left\{\frac{e^{\sum_{i=1}^n y_i W_i \underset{\sim}{x}_i' \underset{\sim}{\beta}}}{\prod_{i=1}^n(1+e^{\underset{\sim}{x}_i' \underset{\sim}{\beta}})^{W_i}}\right\}, \ \underset{\sim}{\beta} \in R^p, \tag{14}$$

$$\pi_2(\underset{\sim}{\beta} \mid \underset{\sim}{y}) \propto\left\{\frac{e^{\sum_{i=1}^n y_i W_i \underset{\sim}{x}_i' \underset{\sim}{\beta}}}{\prod_{i=1}^n(1+e^{W_i \underset{\sim}{x}_i' \underset{\sim}{\beta}})}\right\}, \ \underset{\sim}{\beta} \in R^p. \tag{15}$$

For both posterior distributions, the prior is assumed to be a uniform distribution on $\underset{\sim}{\beta}$; see Chen et al. (2008) for more details about Jeffreys' prior.

In this case, we can use the Metropolis-Hastings sampler (or the Gibbs sampler) to obtain samples of $\underset{\sim}{\beta}$; see Appendix A for more details. For both posteriors, large values of W_i will give unacceptably small variance, so it is necessary to replace the original weights with adjusted weights. After we replace the original survey weights with three adjusted weights in these two posteriors, we have six models, and they are illustrated in the following flowchart,

2.4 Bayesian Predictive Inference for Logistic Regression

In the realm of Bayesian predictive inference, our initial step involves the specification of a prior distribution for the model parameters $\beta_1, \beta_2, \ldots, \beta_{p-1}$. This prior can be informed by our pre-existing beliefs or domain knowledge concerning the interplay between the predictor variables and the outcome. Subsequently, we employ Bayes' theorem to refine

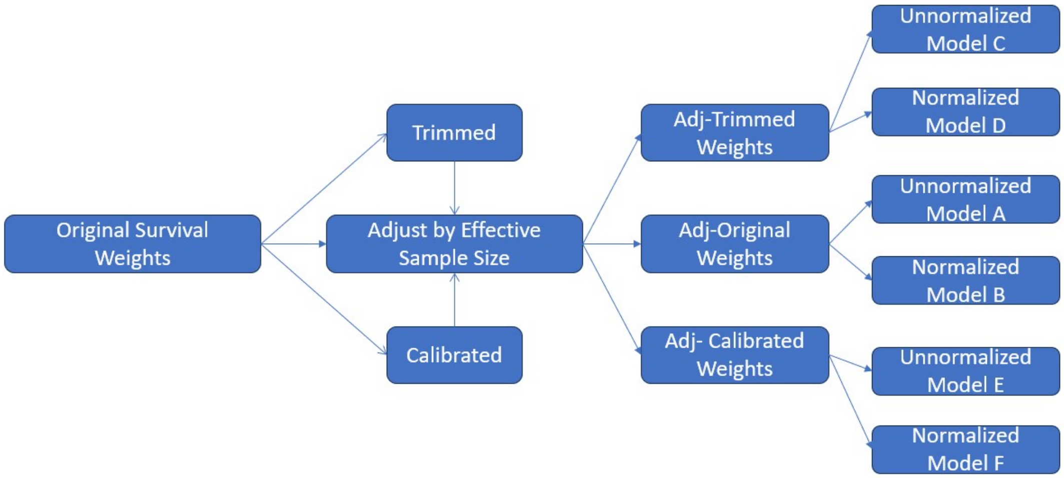

Figure 1: A flowchart of the six models

this prior, yielding the posterior distribution of the parameters in light of the observed data:

$$\pi(\underset{\sim}{\beta}|y, \underset{\sim}{x}) \propto \pi(y|\underset{\sim}{x}, \underset{\sim}{\beta})\pi(\underset{\sim}{\beta}),$$

where, $\pi(y|\underset{\sim}{x}, \underset{\sim}{\beta})$ represents the likelihood function, delineating the likelihood of the observed data (y) given both the predictor variables $(\underset{\sim}{x})$ and the model parameters $(\underset{\sim}{\beta})$. Simultaneously, $\pi(\underset{\sim}{\beta})$ embodies the prior distribution encompassing the parameters.

To model the binary response, we consider using a logistic regression model. Bayesian predictive inference in logistic regression is a powerful technique for estimating the probability of a binary outcome, such as success or failure, based on a set of predictor variables. This method begins by establishing a prior distribution over the model parameters, which encapsulates our initial beliefs or knowledge about the relationship between the predictors and the outcome. Subsequently, this prior is updated using Bayes' theorem in light of the observed data. The resulting posterior distribution over the parameters provides a refined estimate. In our logistic regression model, we express the log odds of the binary outcome in terms of the predictor variables:

$$\log \frac{p(y=1|\underset{\sim}{x})}{1-p(y=1|\underset{\sim}{x})} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_{p-1} x_{p-1},$$

where $p(y=1|\underset{\sim}{x})$ represents the probability of the binary outcome equals 1 $(y=1)$ given the predictor variables $\underset{\sim}{x}$, β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_{p-1}$ are the coefficients for the predictor variables $x_1, x_2, \ldots, x_{p-1}$. The logistic function ensures that the output lies within the range of 0 to 1. Following the derivation of the posterior distribution, it becomes possible to predict the probability of the binary outcome for new observations, thereby enabling insightful inferences.

In this investigation, a central hurdle lies in estimating unobserved variables. These play a crucial role in predicting the proportion of the finite population, yet they are typically beyond our knowledge. To tackle this, we employ a stratification method that enables predictive inferences about the population's interests. In this approach, the population is partitioned into k distinct strata or cells. Assuming that the sample covariates adequately represent the population, especially with a large sample size, unobserved variables can be logically excluded. This is because the sampled covariates are likely to encompass a broad range of population characteristics. Within the k-th stratum, the corresponding variables $\underset{\sim}{x}_k$ are well-known, and every unit shares the same set of covariates. By successfully stratifying the population, the need to estimate non-sampled variables—typically a challenging and resource-intensive endeavor—is efficiently obviated. Consequently, we can employ the Horvitz-Thompson estimator for each stratum, aggregating the original survey weights of each unit in the stratum. This is represented by $\hat{N}_k = \sum_{i \in \kappa} W_i$, where κ denotes the set of indices in stratum k.

According to the population distribution (equation 8), we can express the total count, t_k, for the k-th stratum as follows:

$$t_k \mid \underset{\sim}{\beta} \overset{ind}{\sim} \text{Binomial}\left\{ \hat{N}_k, \frac{e^{\underset{\sim}{x}'_k \underset{\sim}{\beta}}}{1 + e^{\underset{\sim}{x}'_k \underset{\sim}{\beta}}} \right\}, \quad k = 1, \ldots, K, \tag{16}$$

where, t_k represents the total count of the binary outcome ($y = 1$) in the k-th stratum. The estimate of the finite population proportion is then given by:

$$\hat{T} = \frac{\sum_{k=1}^{K} t_k}{\sum_{k=1}^{K} \hat{N}_k}.$$

Additionally, this method, a form of stratified analysis, is particularly effective when all covariates are discrete with only a few levels. It also works well when a specific continuous variable can be discretized into a small number of levels. The discretization of continuous variables simplifies their handling, allowing us to apply methods designed for discrete data. This approach not only reduces complexity but also lessens the computational load of the methods. Moreover, in cases where auxiliary information is accessible in the form of known marginal counts, the raking approach provides an alternative and more robust procedure within strata. However, it is important to note that such auxiliary information is often not readily available in most scenarios. As a result, we commonly resort to employing survey weights, as we have done here. This makes prediction using the already fitted logistic regression the preferred method. To illustrate, consider a single stratum with s successes and f failures, a sample size of $n = s + f$, and a stratum size of N. Raking allocates N units proportionally to obtain $(s/n)N$ and $(f/n)N$. It is crucial to acknowledge that this proportionality assumption may not always hold; for a comprehensive treatment of general multiway tables using iterative proportional fitting (IPF), refer to Deville et al. (1993).

In summary, by segmenting sampled variables into distinct strata, we can efficiently make predictive inferences for finite populations. This is accomplished by leveraging the covariates and cumulative weights of each stratum. Such a stratification strategy empowers us to effectively utilize sampled variables for estimation, eliminating the need to estimate non-sampled variables. This streamlines the process and reinforces the reliability of our findings.

3 Simulation

In this section, we employ a design-based approach to generate both the finite population and the sample, as outlined in Nandram and Rao (2021) and Chen et al. (2020). We only need to generate the population and the probability sample. We also vary the strength of the association between the response and covariates. Age, race, and sex, which are the three covariates, determine the structure of the simulation data as expressed below:

$$z_i = 24.2449 + 0.0559x_{1i} + 1.2656x_{2i} + 1.2525x_{3i} + \epsilon_i,$$

$$y_i = \begin{cases} 1, & \text{if } z_i \geq 30, \\ 0, & \text{if } z_i < 30, \end{cases} \quad \text{for } i = 1, \ldots, N,$$

where,

$$x_{1i} \overset{\text{ind}}{\sim} \text{Discrete Uniform}(20, 90), x_{3i} \overset{\text{ind}}{\sim} \text{Bernoulli}(0.5),$$
$$x_{2i} \mid x_{1i}, x_{3i} \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{e^{a_i}}{1 + e^{a_i}}\right), \epsilon_i \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2),$$

with $a_i = [24.2449 + 0.0559x_{1i} + 1.2525x_{3i}]^{\frac{1}{10}}$. For the simulation studies, we adjust the values of σ by trial and error to control the correlation coefficient ρ between z_i and the linear predictor $\underset{\sim}{x}_i' \underset{\sim}{\beta}$ at 0.2, 0.3, 0.5, and 0.8.

We then proceeded to create a probability-based sample with a target size of $n = 200$ using the randomized systematic probability-proportional-to-size (PPS) sampling technique. This involved calculating selection probabilities (π_i) according to the formula:

$$\pi_i = \frac{n_i b_i}{\sum_{i=1}^{N} b_i}, \quad i = 1 \ldots N,$$

where $b_i = \theta + 0.2x_{1i} + 5x_{2i} + 5x_{3i}$. The value of θ was determined through iterative testing to ensure that the variation in survey weights met our specified criteria of $\max\{b_i\}/\min\{b_i\} \approx 50$. Following the application of the PPS sampling method with $\underset{\sim}{\pi}$ to select our samples, we assigned survey weights ($W_i = 1/\pi_i$) to each unit in the sample ($i = 1, \ldots, n$). Additionally, we deliberately identified 5 samples with a response variable of 0 and increased their corresponding weights by a factor of 3, categorizing these five units as outliers. Subsequently, we computed these three adjusted weights based on $\underset{\sim}{W}$. In the case of adjusted calibrated weights, it was assumed that the overall population totals were accessible.

In this simulation, our parameter of interest is the finite population proportion. To evaluate the performance of a given estimator, we employ several metrics:

1. Absolute Relative Bias (ARB):

$$\text{ARB} = \frac{1}{H} \sum_{h=1}^{H} \left| \frac{PM^{(h)} - T}{T} \right|.$$

2. Posterior Standard Deviation (PSD):

$$\text{PSD} = \frac{1}{H}\sum_{h=1}^{H} PSD^{(h)}.$$

3. Posterior Root Mean Squared Error (PRMSE):

$$\text{PRMSE} = \frac{1}{H}\sum_{h=1}^{H} \sqrt{\left(PM^{(h)} - T\right)^2 + (PSD^{(h)})^2}.$$

4. Coverage Probability (CP):

$$\text{CP} = \frac{1}{H}\sum_{h=1}^{H} I\left(C_{025}^{(h)} \leq T \leq C_{975}^{(h)}\right).$$

5. Width of the Highest Posterior Density Interval (Wid):

$$\text{Wid} = C_{975}^{(h)} - C_{025}^{(h)}.$$

Here, T represents the true finite population proportion, $PM^{(h)}$ is the posterior mean, $PSD^{(h)}$ is the posterior standard deviation, $I(\cdot)$ is the indicator function, and $C_{025}^{(h)}$ and $C_{975}^{(h)}$ denote endpoints (not necessarily percentiles). These metrics are computed from the h^{th} simulated sample, with $H = 100$ representing the total number of simulation runs.

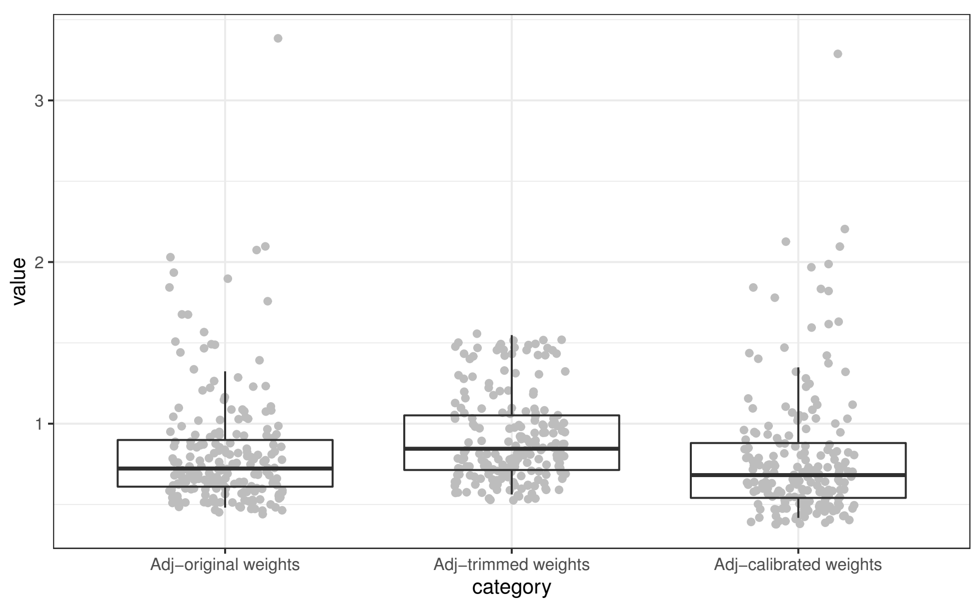

Figure 2: Comparison boxplots with jittered points of three different weights using one simulation dataset with $n = 200$.

Figure 2 displays the boxplot from a single simulation iteration. It illustrates that the adjusted calibrated weights exhibit median, quartiles, and range values that closely align with those of the adjusted original weights. This suggests that the calibration process did not introduce significant bias into the dataset. By trimming some weight outliers and redistributing weights to ensure $\sum_{i=1}^{n} W_i^* = \sum_{i=1}^{n} W_i = \hat{N}$, the boxplot of the adjusted trimmed weights appears slightly higher and more consistent compared to the others, without any outliers. The corresponding effective sample sizes for this run are $n_e = 162$, $n_e^* = 184$, and $\tilde{n}_e = 147$.

Table 1: Comparison of six models using simulation data with $\rho = 0.2, 0.3, 0.5, 0.8$

	ρ	A	B	C	D	E	F
ARB	0.2	0.106	0.118	0.075	0.069	0.104	0.116
	0.3	0.093	0.109	0.076	0.077	0.093	0.105
	0.5	0.105	0.130	0.084	0.090	0.105	0.128
	0.8	0.108	0.117	0.098	0.101	0.109	0.122
PSD	0.2	0.029	0.033	0.025	0.026	0.028	0.033
	0.3	0.028	0.033	0.025	0.026	0.028	0.033
	0.5	0.026	0.031	0.022	0.024	0.026	0.031
	0.8	0.020	0.019	0.017	0.018	0.020	0.019
PRMSE	0.2	0.004	0.006	0.002	0.002	0.004	0.005
	0.3	0.003	0.004	0.002	0.002	0.003	0.004
	0.5	0.003	0.004	0.002	0.002	0.003	0.004
	0.8	0.002	0.002	0.001	0.002	0.002	0.002
CP	0.2	0.69	0.71	0.76	0.80	0.68	0.69
	0.3	0.68	0.77	0.80	0.84	0.68	0.79
	0.5	0.72	0.74	0.77	0.79	0.67	0.72
	0.8	0.71	0.61	0.65	0.68	0.70	0.61
Wid	0.2	0.112	0.131	0.099	0.102	0.110	0.128
	0.3	0.110	0.130	0.097	0.100	0.108	0.128
	0.5	0.103	0.121	0.088	0.096	0.102	0.120
	0.8	0.080	0.075	0.068	0.070	0.078	0.075

In Table 1, we provide a comprehensive comparison of the simulation results. It is evident that model D consistently outperforms models A, B, C, E, and F in terms of lower bias, enhanced precision, and overall superior performance in estimating the parameter of interest. Models A and E, which utilize unnormalized densities with adjusted original and calibrated weights respectively, exhibit similar results across all ρ levels. This suggests that the difference between adjusted original weights $\underset{\sim}{w}$ and adjusted calibrated weights $\underset{\sim}{\tilde{w}}$ is negligible. The same observation holds for models B and F, which utilize normalized densities. However, models C and D, which employ adjusted trimmed weights, demonstrate more substantial variations in these metrics, particularly at lower ρ levels (0.2 and 0.3). They display lower Absolute Relative Bias (ARB) and Posterior Root Mean Squared Error (PRMSE), and higher Coverage Probability (CP). This indicates an advantage of using adjusted trimmed weights in this simulation, especially in scenarios involving weight outliers. It is worth noting that at lower ρ levels (0.2 and 0.3), the normalized model with adjusted trimmed weights (model D) outperforms the unnormalized model with adjusted trimmed weights (model C). Model D tends to exhibit relatively lower or similar ARB values, along with higher CP values. Weight outliers in survey data occur when certain survey units are assigned disproportionately large or small weights compared to others. In binary samples, these outlier weights can significantly impact the predictive inference of the population. The subsequent simulation involves generating synthetic data with varying levels of outlier weights (by manipulating the fraction of outliers from 0 to 0.5) and assessing the performance of the six models in terms of ARB. Specifically, we randomly select some units in the sample based on the outlier fraction and multiply their original weights by five. We then proceed to compute the posterior means of all six models and plot their respective ARBs.

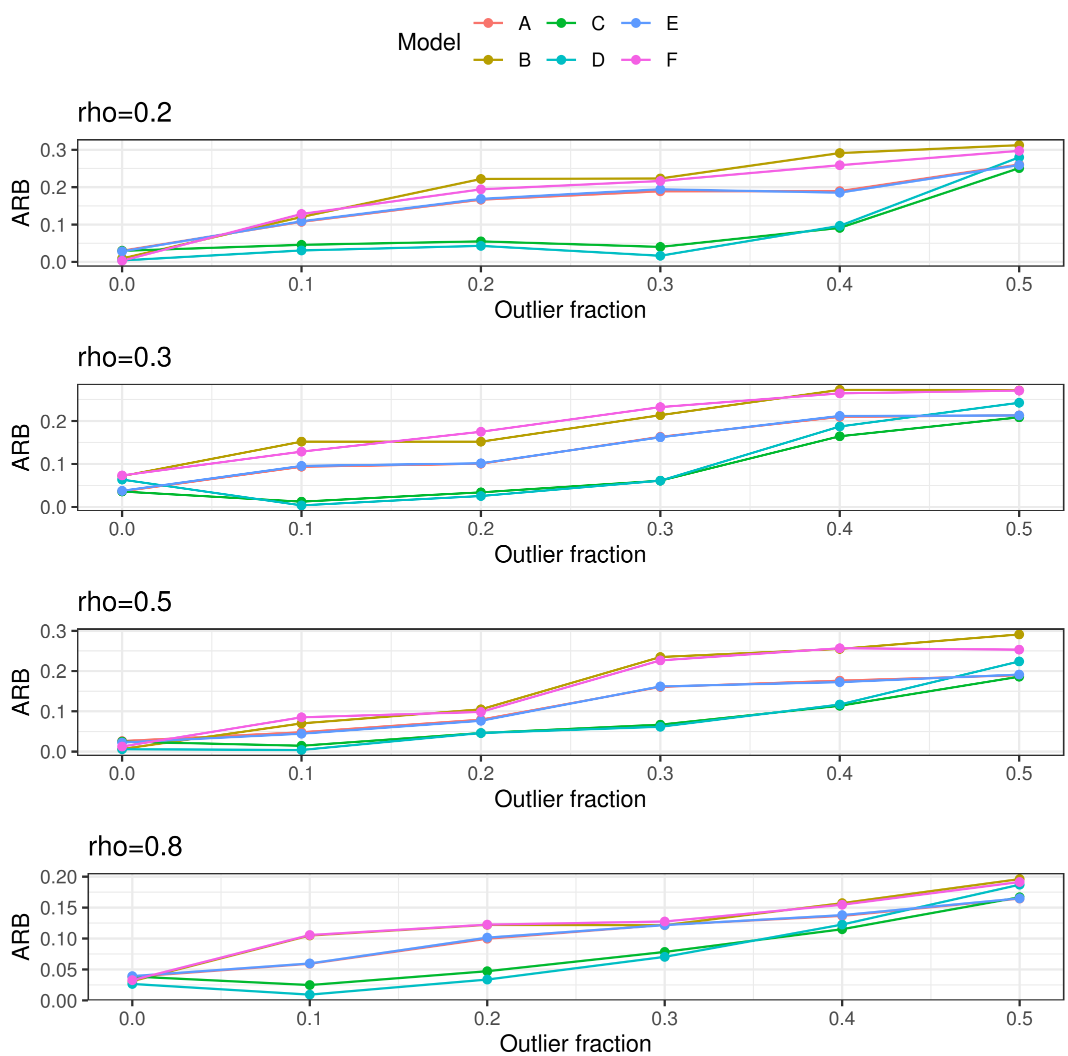

Figure 3: ARB Comparison of six models using simulation data with $\rho = 0.2, 0.3, 0.5, 0.8$ and different outlier fractions.

The performance of the six models at varying levels of outlier fractions is depicted in Figure 3. When the outlier fraction is set to 0, the results indicate negligible differences among the six models. As the outlier fraction increases, models employing adjusted trimmed weights (models C and D) consistently outperform the others. This outcome aligns with expectations, as weight trimming or Winsorization is effective in mitigating the impact of outlier weights. It is noteworthy that at lower outlier fractions (0.1, 0.2, and 0.3), the model employing normalized density (model D) exhibits lower Absolute Relative

Bias (ARB) values compared to the model with unnormalized density (model C). However, across all outlier fractions, the normalized density models (models B and F) display higher ARB values compared to the unnormalized density models (models A and E). Similarly, there is little distinction between the unnormalized models using adjusted original weights (model A) and adjusted calibrated weights (model E), and the same applies to the normalized models (models B and F).

4 Application on Body Mass Index

Examining the proportion of the population classified as obese is a crucial aspect of assessing the health of a finite population. In this section, we employ our six models using a sample dataset of Body Mass Index (BMI) measurements from individuals across eight counties in California, sourced from the Third National Health and Nutrition Examination Survey (NHANES III). This dataset comprises 1,867 observations, providing a robust basis for constructing a contingency table to represent the finite population covariates. Among the covariates commonly associated with BMI data, age, race, and sex are discrete. Age, spanning from 20 to 89, encompasses approximately 70 levels, while race and sex each possess two levels. Consequently, there are 280 unique covariate vectors $\underset{\sim}{x}$.

For this example concerning eight counties in California, we utilized web scraping along with the 1990 census report of California to obtain the following totals: population size $N = 4,035,862$, age $= 36.7 \times N$, race $= .719 \times N$, and sex $= .497 \times N$. This provides a basis for making meaningful comparisons between the outcomes of the various models.

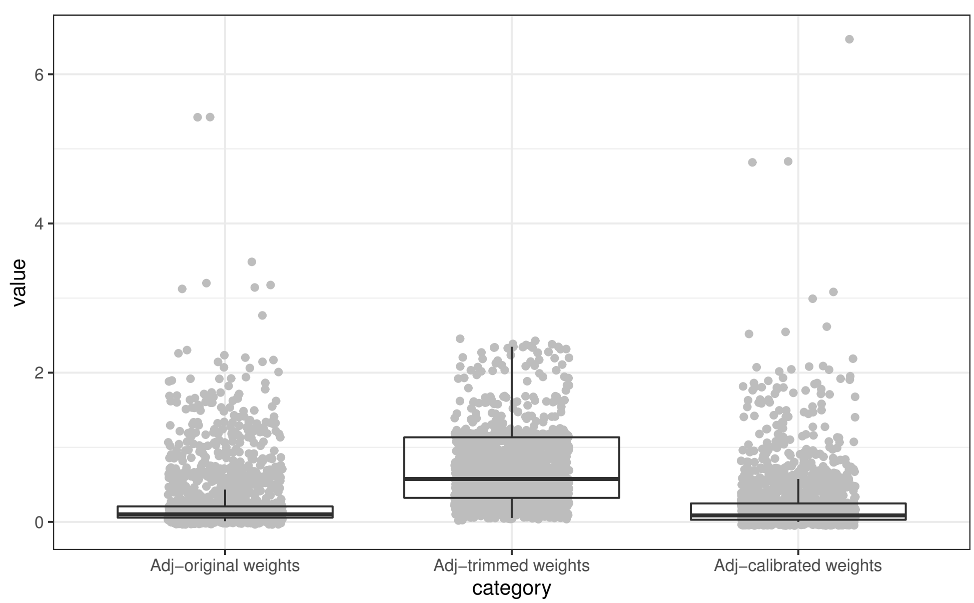

Figure 4: Comparison boxplot with jittered points of three different weights using BMI data

It is important to note that the inclusion of jittered points in the boxplot might visually suggest that some data points fall below the zero line. However, it should be emphasized that all points are positive. In Figure 4, we observe that the boxplots of adjusted original weights and adjusted calibrated weights exhibit a heavy tail. This indicates the presence of some outliers with weights surpassing the maximum boundary, which can have a significant impact on the estimates. The corresponding effective sample sizes are $n_e = 498, n_e^* = 1301, \tilde{n}_e = 498$, compared to the actual sample size $n = 1,867$.

As is well-known, outliers can exert a substantial influence on the effective sample size. Consider a dataset with an abundance of outliers used to estimate the mean of a population. These outliers can tug the mean away from its true value, leading to a larger variance and subsequently a diminished effective sample size. In such a scenario, the effective sample size would be lower than the actual sample size, signifying that there is less reliable information in the data for making inferences about the population. In the case of the BMI dataset, the substantial presence of outliers leads to a reduction in the two effective sample sizes computed using adjusted original weights and adjusted calibrated weights.

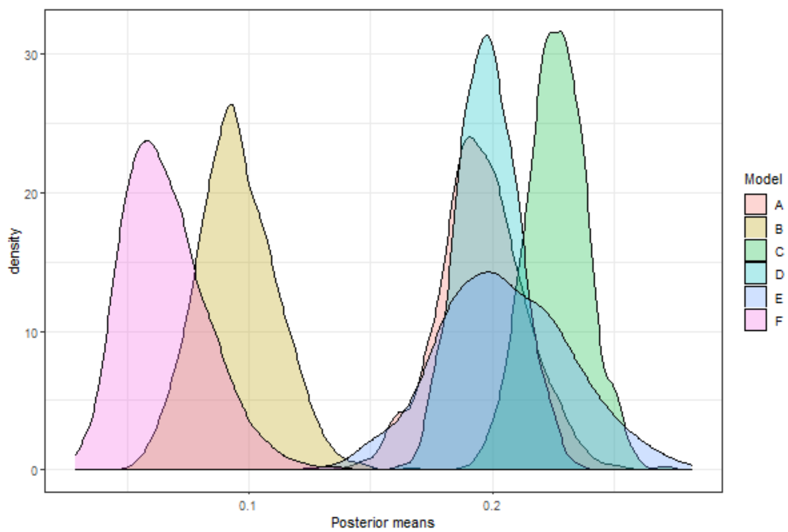

Figure 5: Comparison of densities of posterior means (PM) using BMI data

The distributions of posterior means (PMs) depicted in Figure 5 indicate that these normalized models with adjusted original weights (model B) and adjusted calibrated weights (model F) tend to skew towards higher values compared to other models. This raises potential concerns regarding the accuracy of their estimates. A more comprehensive examination of the results can be found in the accompanying Appendix A, which includes convergence tests.

Table 2: Comparison of six models using BMI data

Model	PM	PSD	PCV	95% CI
A	0.196	0.017	0.089	(0.160, 0.232)
B	0.095	0.016	0.172	(0.065, 0.127)
C	0.226	0.012	0.055	(0.202, 0.251)
D	0.198	0.012	0.059	(0.175, 0.222)
E	0.195	0.018	0.092	(0.162, 0.233)
F	0.094	0.017	0.178	(0.063, 0.128)

In Table 2, we provide posterior summaries of the six models for estimating the finite population proportion of obesity. These models are evaluated using four key metrics: posterior mean (PM), posterior standard deviation (PSD), posterior coefficient of variation (PCV), and the 95% credible interval. From the table, models A, C, D, and E may be reasonable but others are unreasonable. It is evident that for adjusted original weights and adjusted calibrated weights, models employing unnormalized densities (Models A and E) demonstrate greater resilience against weights with outliers, yielding more stable estimates of the population proportion in the presence of extreme weights compared to models employing normalized densities (Models B and F). Upon closer inspection, it is notable that responses associated with extreme weights predominantly result in 0 values. Consequently, this drives the PMs of Models B and F towards 0. In contrast, when weight trimming is implemented as a strategy to mitigate the influence of these outlier weights, Models C and D exhibit higher PMs. Furthermore, Models C and D boast the smallest PSD values (0.012), indicating the least amount of uncertainty in their estimates. This is a crucial characteristic in statistical analysis, as reduced uncertainty typically leads to more reliable results. The low PCVs for Models C and D (0.055 and 0.059, respectively) further substantiate their robustness, suggesting the lowest relative variability in their estimates. Conversely, when utilizing adjusted original weights and adjusted calibrated weights (as in Models A, B, E, and F), the resulting interval is wider compared to Models C and D. This broadened interval could potentially introduce a higher degree of uncertainty into the analysis.

Figure 6 provides an overview of the posterior distributions of $\underset{\sim}{\beta}$ in the context of BMI data. It is evident that in cases with small sample sizes, the posterior densities of $\underset{\sim}{\beta}$ exhibit a considerable variance. However, when employing the normalized model in scenarios with small sample sizes, the distributions tend to contract slightly. Conversely, for cases with a large effective sample size and models utilizing adjusted trimmed survey weights, there is no significant discrepancy between the unnormalized and normalized models.

5 Conclusion

By integrating pertinent auxiliary data, our study provides a comprehensive framework for conducting Bayesian predictive inference on binary responses utilizing probability survey samples. We introduce three distinct sets of adjusted survey weights: the adjusted original survey weights, the adjusted trimmed survey weights, and the adjusted calibrated survey weights. These weights are incorporated into both unnormalized and normalized densities within the Bayesian framework. Subsequently, we implement the logistic regression model using the Metropolis-Hastings sampler, enabling predictive inference on finite population interests through surrogate sampling and the stratification approach.

The simulations and analysis of the BMI dataset highlight the advantages of models incorporating adjusted trimmed weights in conjunction with normalized density functions, especially in scenarios with small correlation coefficients. These models not only excel in estimation accuracy but also align with the fundamental tenets of the Bayesian paradigm, ensuring the validity of posterior distributions. In situations involving exceedingly large population sizes, the normalized posterior employing adjusted trimmed weights demonstrates enhanced resilience against outliers in weights, ultimately leading to more precise estimations of the parameters of interest. Consequently, such models merit careful consideration for their potential advantages when addressing scenarios with small correlation coefficients and large population sizes.

Future research on Bayesian predictive inference, incorporating covariates and survey

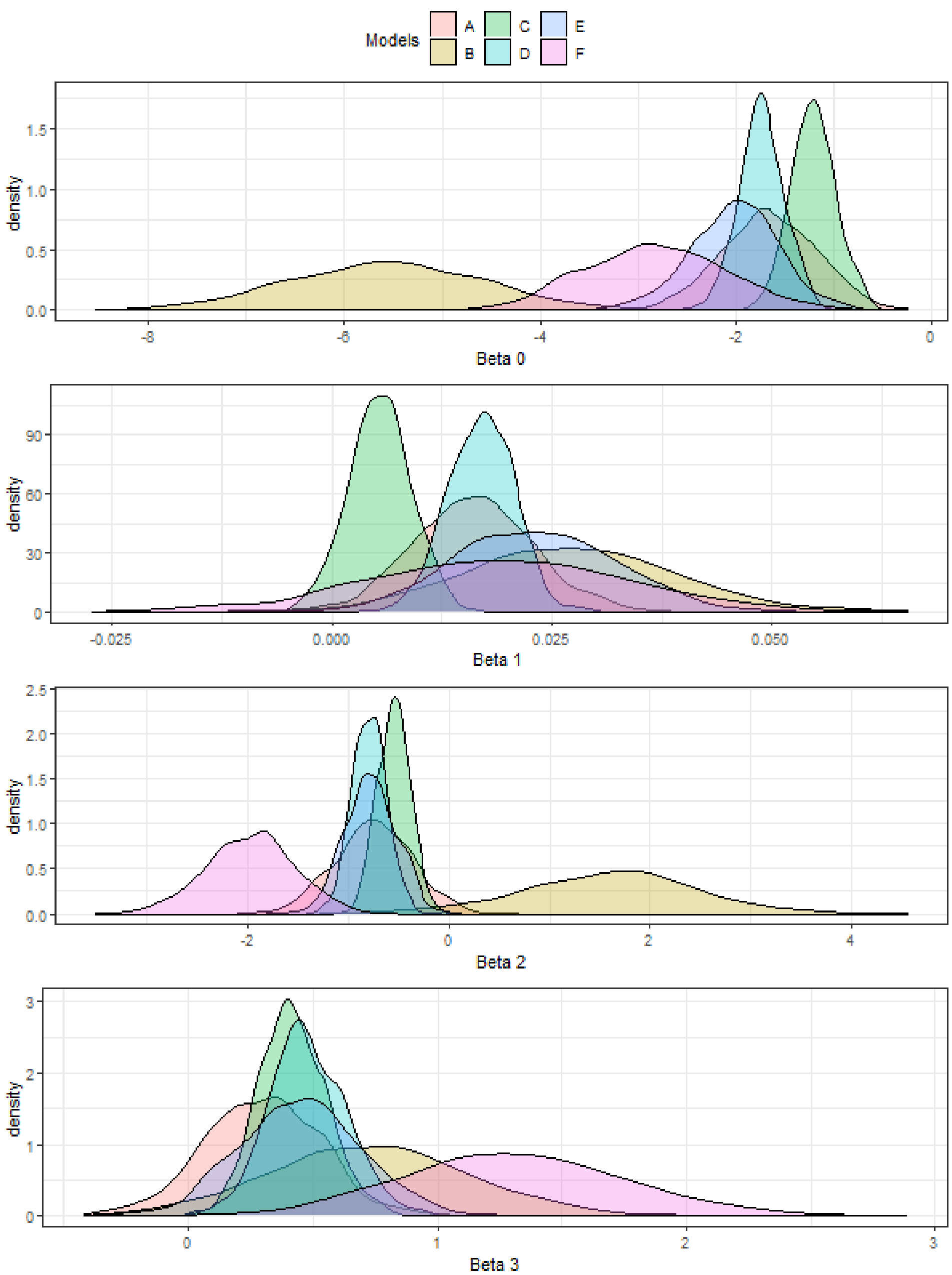

Figure 6: Comparison of posterior densities of $\underset{\sim}{\beta}$ using BMI data

weights, is anticipated to be extensive. The proposed framework lends itself to further extension and application in scenarios involving different types of response variables, such as categorical or continuous responses. Moreover, it can be adapted to accommodate more intricate survey data, including issues related to nonresponse and nonprobability sampling.

Appendix

A. Metropolis-Hastings Algorithm for $\underset{\sim}{\beta}$

The Metropolis-Hastings algorithm is a Markov chain Monte Carlo (MCMC) technique utilized for generating samples from a target probability distribution that is challenging to sample from directly. Widely applied in Bayesian statistics and other contexts where sampling from a complex distribution is essential, this algorithm is invaluable. Its fundamental concept is to simulate a Markov chain whose stationary distribution corresponds to the target distribution of interest. The algorithm progresses through a sequence of iterations. At each iteration, a candidate state is proposed, and acceptance or rejection is determined based on an acceptance probability. This probability is designed to ensure that the chain converges to the target distribution, even if the initial distribution is distant from it.

In our study, we employ the Metropolis-Hastings algorithm to generate a sample of $\underset{\sim}{\beta}$, using $\pi(\underset{\sim}{\beta} \mid \underset{\sim}{y})$ from (14) as the target density. To do this, we must specify a candidate generating (proposal) density, denoted as $q(\underset{\sim}{\beta})$. The choice of this proposal distribution significantly impacts the algorithm's performance. Ideally, the proposal distribution should exhibit high probability density in regions where the target distribution has a high probability density, thereby minimizing the number of rejected proposals. Additionally, it should possess some probability density in regions where the target distribution has low probability density, allowing the chain to explore the entire space.

For unimodal target densities, it is reasonable to consider a normal approximation distribution. Specifically, by using the mode of our target density as the mean, denoted as $\hat{\underset{\sim}{\beta}}$, and the negative of the inverse-Hessian matrix at the mode as the covariance matrix, denoted as $\hat{\Sigma}$, we obtain a normal distribution. This leads to:

$$\underset{\sim}{\beta} \mid \underset{\sim}{y} \overset{app}{\sim} \text{Normal}\left\{\hat{\underset{\sim}{\beta}}, \sigma^2\hat{\Sigma}\right\},$$

$$\frac{\gamma}{\sigma^2} \sim \chi^2_{\gamma},$$

where γ represents the degrees of freedom of the chi-squared distribution. The degrees of freedom are also adjusted to ensure the jumping rate falls within the range of $25\% - 75\%$.

The algorithm works as follows,

1. Start with an initial value, $\underset{\sim}{\beta}_0$.
2. At each iteration t:
   (a) Generate a new state, $\underset{\sim}{\beta}_t$, from $q(\underset{\sim}{\beta})$.
   (b) Calculate the acceptance probability, $\alpha(\underset{\sim}{\beta}_t, \underset{\sim}{\beta}_{t-1}) = \min\{1, \frac{\pi(\underset{\sim}{\beta}_{t-1})q(\underset{\sim}{\beta}_t)}{\pi(\underset{\sim}{\beta}_t)q(\underset{\sim}{\beta}_{t-1})}\}$, where $\underset{\sim}{\beta}_{t-1}$ is the current state.
   (c) Generate a uniform random variable u from [0,1].

(d) If $u \leq \alpha(\underset{\sim}{\beta}_t, \underset{\sim}{\beta}_{t-1})$, accept the proposed state and set $\underset{\sim}{\beta}_{t+1} = \underset{\sim}{\beta}_t$. Otherwise, reject the proposed state and set $\underset{\sim}{\beta}_{t+1} = \underset{\sim}{\beta}_{t-1}$.

3. Repeat step 2 until convergence.

For both the simulation and BMI example, we set the degrees of freedom equal to 8, resulting in jumping rates of $(0.468, 0.460, 0.461, 0.462, 0.469, 0.468)$ for all models in BMI. Additionally, we configure the iteration count to 15,000, perform a burn-in of 5,000, and retain every tenth value to obtain the posterior samples of parameters. Subsequently, we utilize trace plots and auto-correlations, conduct the Geweke test for stationarity, and calculate the effective sample sizes to verify their convergence.

Table 3: P-values and Effective Sample Size for Models A-F in BMI

	Model A		Model B	
	P-value	ESS	P-value	ESS
Beta 1	0.623	1000	0.620	1000
Beta 2	0.383	1000	0.492	1000
Beta 3	0.435	1000	0.908	1000
Beta 4	0.592	1000	0.926	1000
	Model C		**Model D**	
	P-value	ESS	P-value	ESS
Beta 1	0.296	1079	0.124	1000
Beta 2	0.078	902	0.421	1164
Beta 3	0.543	762	0.662	1000
Beta 4	0.302	1099	0.606	1000
	Model E		**Model F**	
	P-value	ESS	P-value	ESS
Beta 1	0.454	1096	0.447	1000
Beta 2	0.706	1000	0.537	1000
Beta 3	0.848	1000	0.069	1000
Beta 4	0.028	1000	0.907	1000

Table 3 presents the results of a Geweke test and the Effective Sample Size (ESS) for different parameters ($\underset{\sim}{\beta}$) under six different models in the BMI application. In the Geweke test, a higher p-value suggests that the chain has converged well. In our table, all p-values are greater than 0.05, apart from β_4 of Model E (unnormalized density with adjusted calibrated weights). Effective Sample Size (ESS) is another diagnostic measure used in MCMC analysis. In our case, the ESS for all parameters of each model is very close to or exceeds 1000 (the length of $\beta_i, i = 1, \ldots, 4$), except β_3 of Model C (unnormalized density with adjusted trimmed weights). According to all these measurements, we can conclude that all models across all parameters have good convergence of the MCMC chains and the results obtained from these models can be considered reliable.

B. Tuning Parameter in Normalized Cases

In a sense, the following likelihood represents the "correct" likelihood of the population:

$$\widehat{L(\underset{\sim}{\theta})} = \sum_{i=1}^{n} W_i \log\left\{f(y_i|\underset{\sim}{\theta})\right\}. \tag{B.1}$$

However, as mentioned before, it is desirable to conduct everything within the Bayesian paradigm. After incorporating weights into the likelihood function, if we normalize the pseudo-likelihood function to be a proper density such that $g(y_i \mid \underset{\sim}{\theta})$ integrates to 1, we have:

$$g(y_i|\underset{\sim}{\theta}) = \frac{(f(y_i|\underset{\sim}{\theta}))^{W_i}}{\int (f(y_i|\underset{\sim}{\theta}))^{W_i} dy_i}. \tag{B.2}$$

However, in the binary case, after introducing the normalization constant $\int f(y_i \mid \underset{\sim}{\theta})^{W_i} dy_i$, the likelihood function is no longer the likelihood function for the population because the normalization constant depends on the parameter θ. Thus, there is a trade-off between likelihood and proper density.

One idea is to introduce a tuning parameter γ to control this trade-off. With the tuning parameter γ, the likelihood function and density become:

$$\widehat{L(\underset{\sim}{\theta})} = \sum_{i=1}^{n} \gamma W_i \log\left\{f(y_i|\underset{\sim}{\theta})\right\}, \tag{B.3}$$

and

$$g(y_i|\underset{\sim}{\theta}) = \frac{(f(y_i|\underset{\sim}{\theta}))^{\gamma W_i}}{\int (f(y_i|\underset{\sim}{\theta}))^{\gamma W_i} dy_i}. \tag{B.4}$$

It is important to note that the solutions of equations (B.1) and (B.4) are the same when the denominator is a function of $\underset{\sim}{\theta}$, so (B.4) is unbiased in that case. Then, we can assume a prior distribution for γ or perform cross-validation based on the conditional predictive ordinate (CPO) or classification accuracy to select the appropriate tuning parameter.

C. Illustration of impact of normalization constant

Consider the following models for y:

$$f(y \mid \theta) = \theta e^{-\theta y}, \quad \theta, y > 0.$$

Let $W > 0$ be a fixed real number. Then,

$$(f(y \mid \theta))^W = \theta^W e^{-\theta W y}, \quad y > 0,$$

which is not a density function of y. In this way, the density function of y becomes:

$$\frac{(f(y \mid \theta))^W}{\int (f(y \mid \theta))^W dy} = \frac{\theta^W e^{-\theta W y}}{\int \theta^W e^{-\theta W y} dy}, \quad y > 0.$$

Assuming the prior of θ, $\pi(\theta) = 1, \theta > 0$, we have two forms of the posterior distribution of θ:

$$\pi(\theta \mid y) \propto \theta^W e^{-\theta W y}, \quad \text{i.e. } \theta \mid y \sim \text{Gamma}(W + 1, Wy),$$

and

$$\pi(\theta \mid y) \propto \theta W e^{-\theta W y}, \quad \text{i.e. } \theta \mid y \sim \text{Gamma}(2, Wy).$$

Obviously,

$$E_1(\theta \mid y) = \frac{W+1}{Wy} = (W+1)a, \quad E_2(\theta \mid y) = \frac{2}{Wy} = 2a,$$

$$Var_1(\theta \mid y) = \frac{W+1}{(Wy)^2} = (W+1)a^2, \quad Var_2(\theta \mid y) = \frac{2}{(Wy)^2} = 2a^2,$$

where $a = 1/(W+1)$ is an irrelevant value. If $W = 1$, there is no difference between the posterior distribution of unnormalized and normalized density. If $W < 1$, $E_1(\theta|y) < E_2(\theta|y)$ and $Var_1(\theta|y) < Var_2(\theta|y)$. If $W > 1$, the opposite situation holds. It is clear that when the normalization constant is related to the parameter, it will affect our estimation. However, if $f(y \mid \theta)$ is normal, there will be no difference between unnormalization and normalization, and there will be no benefit in normalization.

References

Archer, K. J. and Lemeshow, S. (2006). Goodness-of-fit test for a logistic regression model fitted using survey sample data. *The Stata Journal*, 6(1):97–105.

Barasa, K. S. and Muchwanju, C. (2015). Incorporating survey weights into binary and multinomial logistic regression models. *Sci J Appl Math Stat*, 3(6):243–9.

Basu, D. (1971). An essay on the logical foundations of survey sampling, part i. foundations of statistical inferences, vp godambe and da sprott.

Chen, L. and Nandram, B. (2023). Bayesian logistic regression model for sub-areas. *Stats*, 6(1):209–231.

Chen, M.-H., Ibrahim, J. G., and Kim, S. (2008). Properties and implementation of Jeffreys's prior in binomial regression models. *Journal of the American Statistical Association*, 103(484):1659–1664.

Chen, Q., Elliott, M. R., Haziza, D., Yang, Y., Ghosh, M., Little, R. J., Sedransk, J., and Thompson, M. (2017). Approaches to improving survey-weighted estimates. *Statistical Science*, 32(2):227–248.

Chen, Y., Li, P., and Wu, C. (2020). Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, 115(532):2011–2021.

Deville, J.-C., Särndal, C.-E., and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American statistical Association*, 88(423):1013–1020.

Elliott, M. R. and Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, pages 249–264.

Gelman, A. (2007). Struggles with survey weighting and regression modeling. *Statistical Science*, 22(2):153–164.

Haziza, D. and Beaumont, J.-F. (2017). Construction of weights in surveys: A review. *Statistical Science*, 32(2):206–226.

Lohr, S. (2007). Comment: Struggles with survey weighting and regression modeling. *Statistical Science*, 22(2):175–178.

Lohr, S. L. (2021). *Sampling: design and analysis*. CRC press.

Nandram, B., Choi, J. W., and Liu, Y. (2021). Integration of nonprobability and probability samples via survey weights. *International Journal of Statistics and Probability*, 10(6):1–5.

Nandram, B. and Rao, J. (2021). A bayesian approach for integrating a small probability sample with a non-probability sample. *In JSM Proceedings, Survey Research Methods Section, Alexandria, VA: American Statistical Association*, pages 1568–1603.

Potthoff, R. F., Woodbury, M. A., and Manton, K. G. (1992). "Equivalent sample size" and "equivalent degrees of freedom" refinements for inference using survey weights under superpopulation models. *Journal of the American Statistical Association*, 87(418):383–396.

Rader, K. A., Lipsitz, S. R., Fitzmaurice, G. M., Harrington, D. P., Parzen, M., and Sinha, D. (2017). Bias-corrected estimates for logistic regression models for complex surveys with application to the united states' nationwide inpatient sample. *Statistical methods in medical research*, 26(5):2257–2269.

Rao, J. (1966). Alternative estimators in pps sampling for multiple characteristics. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 47–60.

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(5):1114–1145.

Roberts, G., Rao, N., and Kumar, S. (1987). Logistic regression analysis of sample survey data. *Biometrika*, 74(1):1–12.

Yang, L. and Nandram, B. (2023). Bayesian predictive inference model for binary response with covariates and survey weights. *Communications in Statistics - Theory and Methods (Under Review)*.

Yang, L., Nandram, B., and Choi, J. W. (2023). Bayesian predictive inference under nine methods for incorporating survey weights. *International Journal of Statistics and Probability*, 12(1):33–53.