Bayesian Modeling for Mental Health Surveys

Sharifa Zakiya Williams

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Public Health
under the Executive Committee of the
Department of Biostatistics
Mailman School of Public Health

COLUMBIA UNIVERSITY

2018

Abstract

Bayesian Modeling for Mental Health Surveys

Sharifa Zakiya Williams

Sample surveys are often used to collect data for obtaining estimates of finite population quantities, such as disease prevalence. However, non-response and sampling frame under-coverage can cause the survey sample to differ from the target population in important ways. To reduce bias in the survey estimates that can arise from these differences, auxiliary information about the target population from sources including administrative files or census data can be used. Survey weighting is one approach commonly used to reduce bias. Although weighted estimates are relatively easy to obtain, they can be inefficient in the presence of highly dispersed weights. Model-based estimation in survey research offers advantages of improved efficiency in the presence of sparse data and highly variable weights. However, these models can be subject to model misspecification. In this dissertation, we propose Bayesian penalized spline regression models for survey inference about proportions in the entire population as well as in subpopulations. The proposed methods incorporate survey weights as covariates using a penalized spline to protect against model misspecification. We show by simulations that the proposed methods perform well, yielding efficient estimates of population proportion for binary survey data in the presence of highly dispersed weights and robust to model misspecification for survey outcomes. We illustrate the use of the proposed methods to estimate the prevalence of lifetime temper dysregulation disorder among National Guard service members overall and in sub-populations defined by gender and race using the Ohio Army National Guard Mental Health Initiative 2008-2009 survey data. We further extend the proposed framework to the setting where individual auxiliary data for the population are not available and utilize a Bayesian bootstrap approach to complete model-based estimation of current and undiagnosed depression in Hispanics/Latinos of different national backgrounds from the 2015 Washington Heights Community Survey.

Contents

List of Figures

List of Tables

Acknowledgements

It is impossible to overstate my gratitude to my advisor, Dr. Qixuan Chen. With great diligence, dedication, and care, she provided good teaching, sound advice, encouragement, excellent ideas, and a positive mentorship environment. Without Qixuan, my journey from student to researcher would have been much more difficult. I would especially like to thank Qixuan for always being accessible and for the late hours she spent working with me in the weeks leading to my defense.

Special thanks to my dissertation committee chair: Dr. Yuanjia Wang and members: Drs. Sandro Galea, Jeff Goldsmith, and Peter Muennig for their insightful comments and encouragement as well as the challenging questions which only served to improve the quality of this work.

Further thanks to my classmates and friends as well as the administrative team and faculty in the Biostatistics Department for providing a fun environment in which to learn and grow. Thanks for never being tired of me stopping by for paper, to book conference rooms, ask questions, or to get a snack; I will miss our hallway conversations.

To the excellent teachers and mentors I have had along this journey beginning with my high school teacher Mr. Russell Bell and including my college mentor Ms. Deedee Davis: you will never know how much you have affected my academic and professional trajectory. Thanks for sharing your passion. To Drs. Ana Abraido-Lanza and Crystal Fuller-Lewis for mentorship and financial support of my studies through the Initiative for Maximizing Student Development training program.

Importantly, I would like to thank my mother, Mary, for loving me into confidence and strength, and setting the ultimate example of hard work and excellence. I hope any success I have or approximate, shines on you. To my father, Thomas, for leading the way. Having your encouragement and support, knowing you completed this journey, was invaluable. To my grandparents, siblings, step-father, aunts, uncles, cousins, and friends who form the village that shared in many of my disappointments and successes during life and this writing process, I am eternally grateful.

Lovingly dedicated to my late grandmother
Mable Elizabeth Smith née Thomas

Chapter 1

Introduction

1.1 Overview

This dissertation develops Bayesian multilevel regression model-based approaches that utilize auxiliary information to improve the robustness of inference in survey data challenged with non-response and sampling frame under-coverage. In this chapter we briefly review important statistical concepts in survey sampling and introduce the data motivating our work.

1.2 Introduction to Survey Sampling

Survey research methods can be seen as being composed of two main stages: design and inference. Appropriate application of these methods in public health facilitate inference about large populations using moderately sized samples that are relatively low cost and more generalizable to a target population than other study designs.

1.2.1 Survey design

Survey design focuses on the statistical aspects of taking a sample. It provides guidance on defining a target population, the validity of a sample, and measurement error [1]. A critical aspect of survey design is how data are collected. Typically, a sampling technique is used to draw sample units from an enumerated list of target population units also known as a sampling frame. An ideal sample would be a smaller-sized direct translation of the target population, perfectly reflecting each characteristic of the whole population [1]. This is nearly impossible to obtain as populations, particularly human populations, are complex with units entering, leaving, and re-entering the population over time. However, steps can be taken to ensure that our sample will reflect the characteristics of interest in the population as closely as possible; that is, it will be representative.

Most sampling techniques can be described as being probability or non-probability sampling procedures. Probability sampling is a technique in which the chance of being selected into the sample is known for all the units in the population; thus the probability of the resulting sample can be calculated [1]. Probability sampling techniques result in samples that are expected to support efficient statistical analysis and be representative of the population from which they are drawn. Common types of probability sampling techniques include simple random sample, stratified random sampling, probability-proportional-to-size (PPS) sampling, and cluster sampling [2]. More complex probability sampling procedures may combine two or more probability sampling techniques in what is called multistage sampling [1, 2]. However implemented, a probability sample of a few thousand units can provide accurate information on a target population of millions [1, 2]. Non-probability sampling on the other hand, does not involve random selection and as such, it is more difficult to determine whether the samples obtained are

representative of the population. It is often necessary to employ non-probability sampling techniques in situations where probabilistic sampling is not feasible or practical [2]. While non-probability sampling has its place, the primary goal of survey research is to obtain efficient, approximately unbiased estimators and to make appropriate inference about a population using the estimators obtained from a representative sample. As such, employing a probability sampling technique is an important part of ensuring approximately unbiased estimate of population parameters [2].

When a probability sampling strategy is applied, the survey sample obtained will differ from the target population in expected ways i.e. by design. These are called sampling errors. Weighting adjustments are usually utilized to correct the differences between the samples and the population that arise due to sampling errors [3]. This adjustment compensates for differential sampling rates and produces approximately unbiased estimates of parameters in the target population [3]. Often samples will also differ from the target population due to non-sampling errors [2–7]. Non-sampling errors include any errors not attributable to sampling variation such as non-response, where some sampled individuals provide no information because of non-contact or refusal to respond, or sampling frame under-coverage, where not all the units in the target population are included in the enumerated list [3]. Ignoring non-sampling errors in survey research can lead to biased results.

1.2.2 Survey inference

This stage involves selecting and using estimators for parameters of interest as well as variance estimation. Of interest in this thesis are methods for survey inference that correct for differences between the sample and target population that arise due to non-response and under-coverage using auxiliary data from administrative files or census

data. There are two primary inferential paradigms that exist for survey data inference: design-based and model-based. The primary distinction between design- and model-based inference lies in the source of randomness. In the design-based framework, the population is considered fixed and the sample is seen as the realization of a stochastic process. Inference in this case is based on the distribution of estimates generated by the sampling design and is free of any assumptions about the distribution of the population values [8]. In model-based inference, the population is seen as the realization of a stochastic process and as such, population values in this case are considered random and certain assumptions regarding their distribution in the population must be made [8].

1.2.2.1 Design-based inference

As previously mentioned, in the design-based framework the population is fixed and randomness is introduced via sample selection. We cannot predict exactly how precise an estimate is as this requires information about the target population that we do not have access to; however, we judge the precision by examining the randomization distribution i.e. the frequency distribution generated for the estimate as a result of repeated random sampling of all possible samples permissible under the sampling design from the population [8, 9]. This distribution is sometimes referred to as the reference distribution.

The following notation will be utilized for this section. For a population with N units, let $\mathbf{Y}= (y_1, \ldots, y_N)$, where y_i is the survey variable for unit i, and let $\mathbf{I}= (I_1, \ldots, I_N)$ denote the inclusion indicator variable, where $I_i = 1$ if unit i is included in the sample and $I_i = 0$ otherwise. Design-based inference in based on the distribution of $\mathbf{I}$, with the survey variables, $\mathbf{Y}$ treated as fixed quantities [10]. For a randomly

selected sample, s of size n, we attempt to measure and record survey variables for all observation units in the sample $\mathbf{y}= (y_1, \ldots, y_n)$. Please note that sample is obtained via probability sampling technique and will not necessarily contain the first n units of the population unless these units happen to be drawn by the sampling technique; this is rare [9]. Recall that each possible random sample s has a known probability of selection π. An estimate, say a sample mean, $\bar{y}$ can be computed for the sample and is the average of the values of survey variable $\mathbf{y}$ on the individual observation units in the sample [9]. The frequency distribution of the estimates can be calculated for samples obtained from the population via probability sampling selection methods. Therefore it is possible to calculate how frequently any sample will be selected and the estimate for each sample. Samples obtained via selection techniques that are non-probability, are not amenable to design-based inference [9].

Using the design-based approach to inference can be advantageous because it takes into account survey design and provides reliable inference in large samples [10]. Although most design-based estimators obtained using a probability sampling technique are unbiased and consistent, they can be inefficient [10, 11]. Moreover, there are practical considerations with regard to the use of probability sampling in real world settings such as obtaining sampling frames, alternative modes of survey administration, and a growing interest in combining data sources to produce new information where standard design-based strategies may not be appropriate [11]. Finally, design-based methods become inapplicable in situations where the randomization distribution is corrupted by non-sampling errors such as non-response and measurement errors [12–14].

Additional weighting adjustments via post-stratification, raking, or response propensity models can be used to address non-response and under-coverage errors. Here, existing survey weights are adjusted according to the distribution of auxiliary data in the

target population and survey sample and the weighted sample observations are used for inference. These non-response weighting adjustments can introduce considerable variation in weights such that the increased variance overwhelms the reduction in bias and results in increased mean squared error [5, 6]. This is particularly the case where the number of auxiliary variables is large or when estimation in small or zero sample sub-populations is desired [5, 6, 15].

Model-assisted approach In the model-assisted approach to design-based survey inference, the aim is to use models to help address important considerations relating to non-response and under-coverage errors [10]. The role of the model in this case is to estimate the variation in the finite population without making the assumption that the model generated the population though it should look as though it could have been generated from the model [12]. This requires model assumptions about these non-sampling errors and their distributions in addition to the randomization distribution that is induced by sample design [12].

Lehtonen and Veijanen (1998) proposed a generalized regression model-assisted estimator that is design-consistent regardless of the validity of the working model under certain regularity conditions. Further, if the working model provides a good fit to the data then the residuals, should be less variable than the response values and the generalized regression estimator should be significantly more efficient than the basic design-weighted estimator.

1.2.2.2 Model-based inference

The basis of a model-based approach to sampling inference is that estimating finite population characteristics can be naturally expressed as a prediction problem [17]. Here,

both $\mathbf{I}=(I_1,\ldots,I_N)$ and $\mathbf{Y}=(y_1,\ldots,y_N)$ are considered random variables and a model is specified for the survey outcomes $\mathbf{Y}$, which then are used to predict the non-sampled values of the population [10, 11]. The view that finite population inference problems are actually prediction problems, leads naturally to a theory in which prediction models, not sample selection probabilities, are central [17]. There are two major variants of model-based inference: super-population modeling and Bayesian modeling.

Super-population modeling This is also also referred to as the frequentist model-based approach. The finite population values of $\mathbf{Y}$ are assumed to be a random sample from a super-population; a super-population is the hypothetical infinite population from which we sample a well-defined finite population. Here, $\mathbf{Y}$ are assigned a probability distribution indexed by fixed parameters θ. Inferences are based on the joint distribution of $\mathbf{Y}$ and $\mathbf{I}$ [13]. It is assumed that the population structure obeys this specified model and that the same model holds with respect to the sample [18]. To reduce or eliminate sampling bias, sample design is usually incorporated into the model.

Bayesian modeling In Bayesian modeling, parameters in the super-population model are assigned a prior distribution and inferences about finite population quantities or parameters are based on their posterior distributions. Bayesian inference requires specification of a distribution for the population values. Inferences for finite population quantities are then based on the posterior predictive distribution of the non-sampled values of $\mathbf{Y}$, $\mathbf{Y}_{\mathrm{i}\notin\mathrm{s}}$, given the sampled values $\mathbf{Y}_{\mathrm{i}\in\mathrm{s}}$. Specification of the distribution for the population values is often achieved via a parametric model indexed by parameters, combined with a prior distribution.

The related calibrated Basyeian approach represents a unified approach to survey inference that is model-based. Under this framework, inferences are Bayesian with models that yield inferences with good design-based properties. That is, Bayesian credible intervals when assessed as confidence intervals in repeated sampling should have close to nominal coverage [11]. Good calibration in survey research requires that Bayesian models incorporate sampling design features such as weighting, stratification, and clustering. Weighting and stratification are captured by including weights and stratification variables as covariates in the model while clustering is captured by Bayesian hierarchical modeling with clusters as random effects [11]. Prior distributions are generally weakly informative to allow the likelihood to dominate the posterior distribution [11]. This Bayesian approach is preferable to a frequentist model-based approach since using weakly informative priors over parameters tends to propagate uncertainty in estimating these parameters yielding better confidence coverage than procedures that fix parameters at their estimates [11].

In summary, model-dependent approaches have bridged the gap between survey inference and the rest of statistics [10, 17]. And often times, under the right conditions such as large samples and uninformative priors, model-based inference results parallel those from design-based inference; moreover, model-based inference not only matches but also outperforms design-based inference if the model is correctly specified [10]. The Bayesian approach, and hierarchical Bayes methods in particular, are attractive because of their ability to handle complex design features, modeling, and provide exact inferences on parameters [10, 17]. The Bayesian approach also yields better inferences for small-sample problems where frequentist solutions are not available [10]. An important consideration with model-based approaches is how best to specify the model; models induce subjectivity - if the model is seriously misspecified then it can yield inferences

that are worse than design-based inferences [10, 11]. Recent work on model-dependent strategies has focused on avoiding misspecification of the models by using smooth regression [17]. The fact is, models are needed regardless to handle non-sampling errors. Its strength is that it provides a flexible unified approach for all survey problems such as non-sampling errors, complex sampling design, outliers, small area models, and combining data from diverse data sources [11].

This dissertation focuses on the development and application of a flexible Bayesian modeling approach to survey inference about the population proportion in data challenged by non-response and under-coverage. We then extend this model to allow reliable estimates of survey quantities in small sub-populations. Finally, we apply the proposed methodology to survey data to estimate the prevalence of mental health related outcomes in a unique survey population.

1.3 Introduction to Motivating Data

Survey research in the field of public health can provide a great deal of information disease prevalence and exposure to known and potential risk factors. The data motivating this research is from the Ohio Army National Guard Mental Health Initiative (OANG MHI) study, a survey of active reserve guards in the OANG conducted by the University Hospitals Cleveland Medical Center and the University of Toledo with cooperation from the OANG. This study aims to examine the prevalence and risk factors of psychiatric disorders among active reserve members registered with the ONG between June 2008 and February 2009 with the goal of improving the psychological well-being of reserve guards during combat and post-deployment. The survey primarily included questions on psychopathology, general health history, substance use and related behaviors, and social

support. In particular, pre-, peri-, and post-deployment measures of social support, traumatic events, and preparedness were taken. Mental health research in this population has suggested that reserve forces have greater risk of long-term psychopathology associated with deployment. This, coupled with the increased deployment of reserve guards, make the study of mental health consequences of war in army reserve guards essential, particularly among at-risk, traditionally small sub-populations such as women and racial minorities. Reserve guards who did not have complete and correct postal address and telephone information listed with the OANG, were unable to receive information about the study and participate in the telephone interview and as such, contribute to sampling frame under-coverage. Further, guards choosing to opt-out of the study result in survey non-response and thereby, complicate the analysis of the survey data. To correct for non-response and coverage bias, appropriate methods are needed so that findings can be generalized from the survey sample to the target population. Methods for correcting potential selection bias in the presence of non-response and under-coverage utilize auxiliary data from a variety of sources such as administrative files and the census. For the OANG MHI study, auxiliary data was obtained from administrative files for the target population. This data included the continuous measure of number of years in service. As such, we propose a flexible Bayesian modeling approach to survey inference that is robust to model misspecification to estimate overall population proportion when there is continuous auxiliary information and extend it to domain estimation. We then apply the proposed methods to the OANG MHI study as a data illustration.

The second survey study we use in this dissertation is the 2015 Washington Heights Community Survey (WHCS). This study provides an extensive cross-sectional view of a unique community in New York City that is predominantly Hispanic and low-income.

However, the WHCS is challenged by complex design and non-response with an American Association for Public Opinion Research (AAPOR) Response Rate of 16.8%. Further, the WHCS provides final survey weights for sampled units which incorporate complex study design as well as adjustments for non-sampling errors. These weights should not be taken to be inversely proportional to each unit's probability of inclusion into the sample. As such, we utilize a finite population Bayesian bootstrap (FPBB) procedure to generate synthetic data for non-sampled units. We assess the performance of this modified method as compared to the standard weighted and model-based approaches in simulation.

1.4 Outline of Chapters

In Chapter 2 we present in more detail the goal and motivation of the research. We describe the proposed Bayesian penalized spline regression model which include a penalized spline on the log-transformed survey weights to allow flexible association between auxiliary data and survey outcomes. We show by simulations its performance as compared to common weighting and model-based approaches in the presence of continuous auxiliary information and complex population association. We also apply the proposed method to the Ohio Army National Guard Mental Health Initiative (OANG MHI) 2008-2009 survey data to estimate the prevalence of current and lifetime temper dysregulation disorder (TDD) among reserve guards. Chapter 3 extends the proposed model to include factor-by-curve interaction with the aim of improving survey inference in sub-populations with small sample size. We describe the proposed stratified multilevel regression model and assess its performance in estimating prevalence in sub-populations using a simulation study. We then apply the stratified modeling approach to the OANG MHI survey data

to estimate current and lifetime TDD in domains defined by gender and race. Chapter 4 further extends the proposed modeling framework to the setting where individual auxiliary information for the target population are not available. Instead, we use auxiliary information about the population margins from the U.S. Census and implement a finite population Bayesian bootstrap approach to complete estimation of current and undiagnosed depression in Hispanics/Latinos of different national backgrounds using data from the 2015 Washington Heights Community Survey. In Chapter 5, we conclude with a summary and some thoughts on possible applications and future work.

Chapter 2

Using administrative data to improve survey inference for population proportions

2.1 Introduction

The US military includes over 1.4 million full-time or active soldiers from the US Army, Navy, Marine Corp, Air Force, and Coast Guard as well as over 1.2 million part-time or reserve soldiers from the Army, Navy, Marine, Air Force and Coast Guard Reserve. While both active and reserve soldiers receive similar training and equipment, reserve service members generally serve one weekend a month and 15 days annually while maintaining full-time civilian life including employment or academic studies in contrast to the active soldiers who are full-time employed by the federal government. Deployment of the reserve component is usually in response to a domestic or international crisis; this exposes reserve members to a range of potentially traumatic events such as war

and natural disaster [19, 20]. Following deployment, reserve soldiers face unique readjustment challenges that have been documented to increase their psychiatric disorder burden when compared to active component counterparts [20–24]. As such, an improved understanding of the mental health of reserve soldiers is warranted.

The work of this chapter is motivated by analyzing the Ohio Army National Guard Mental Health Initiative (OANG MHI) study data. The OANG MHI serves as a national model for examining the prevalence and risk factors of mental health related outcomes, such as psychopathology and substance use behaviors, among National Guard service members with the aim of identifying potential areas of intervention that can be modified during the course of deployment to improve the psychological well-being of soldiers [25, 26]. The target population for this study was all service members of the OANG between June 2008 and February 2009. All members with address information listed with the Guard were notified of the study via mailed letter and opt-out card. While some members chose to opt-out of the study, others refused participation when contacted, were ineligible to participate in the study, or were not contacted before the cohort closed [27]. Furthermore, service members with no or incorrect telephone numbers could not be contacted to complete the 60-minute structured computer-assisted telephone interview [27]. As such, the statistical analysis of the OANG MHI study data is complicated by non-sampling errors such as survey non-response and sampling frame under-coverage.

A de-identified administrative dataset was obtained from the OANG containing individual level data for all the service members in the target population ($N = 10,778$). This included information on gender (male, female), race (white, non-white), rank (enlisted, officer), age (17-24 years, 25-34 years, 35 years and older), and number of years in service. Table 2.1 shows the distribution of these auxiliary variables in the population and in the unweighted survey sample. An essential question is how to best use these

administrative data to improve survey inference for the prevalence of mental and mood disorders among the OANG service members.

A common approach that uses the population auxiliary information to improve survey inference is to weight the sample data via post-stratification or raking [3, 14, 28, 29]. Both methods calibrate the weighted distributions of discrete auxiliary variables in the sample to the distributions of these variables in the population, requiring joint population distributions of the auxiliary variables in the use of post-stratification and marginal population distributions in the use of raking. Although weighting adjustments are easy to implement, weighted estimators can be unstable in the presence of extremely variable weights [6, 7, 10]. Regression models represent another framework for correcting selection bias due to non-response and sampling frame under-coverage [7, 30]. Specifically, Little (1993) considered a basic normal post-stratification model for continuous survey outcomes, assuming distinct mean and variance in each post-stratification and using non-informative Jeffreys prior for the post-stratum means and variances. Gelman and his colleagues proposed multilevel regression with post-stratification by including categorical auxiliary variables and their interactions as covariates [15, 30, 31].

Both the aforementioned weighting and model-based methods use discrete population auxiliary information. To apply to the OANG MHI study, the continuous years of service variable has to be categorized. This might lead to loss of important information if the continuous years of service variable is strongly associated with both response and survey outcomes of interest. Recent extensions to model-based approaches have incorporated penalized spline regression in various settings to protect against model misspecification [32–35]. Generally, penalized splines are an easy to implement mechanism, robust to model misspecification with a flexible mean structure that can be used

within a regression model without being overly concerned with the selection of number and position of knots [36].

In this chapter, we propose a Bayesian penalized spline regression model for improving survey inference challenged by non-response and under-coverage. We present this proposed approach alongside existing methods for improving survey inference challenged by non-response and under-coverage in Section 2, and conduct a simulation study to compare the described methods in Section 3. We then apply the proposed method to estimate the prevalence of current temper or disruptive mood dysregulation disorder (TDD) among OANG service members in Section 4 and conclude the chapter in Section 5.

2.2 Methods

2.2.1 Setting and notation

Consider a target population of N units with binary survey outcome variable Y taking a value of 0 or 1. Of interest is estimation and inference of the population proportion, $\theta = \bar{Y}$. Let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_{p-1})^T$ be $(p-1)$ discrete auxiliary variables and Z_p be a continuous auxiliary variable, and both $\mathbf{Z}$ and Z_p are observed for all the units in the population. Values of Z_p can be grouped using m cut-off values to create a discrete variable Z_p^* with $(m+1)$ groups. Let s denote a sample of size n selected from the population with survey outcome values, $y_1 \ldots y_n$. We assume that the population data are de-identified and thus the sample data cannot be linked to the population data on the individual level.

2.2.2 Design-based approaches

2.2.2.1 Post-stratification

We partition the finite population into J disjoint and exhaustive cells or post-strata defined by the joint distribution of $\mathbf{Z}$ and Z_p^*, with N_j population size in cell j, $j = 1 \dots J$, where $\sum_{j=1}^{J} N_j = N$ and $N_j > 0$. Let $\bar{Y}_j$ be the population proportion of Y within cell j, the finite population proportion can then be written as

$$\theta = \bar{Y} = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{\sum_{j=1}^{J} N_j \bar{Y}_j}{\sum_{j=1}^{J} N_j}. \tag{2.1}$$

We can then divide the sample similarly into J cells, with n_j being the sample size in cell j and $\sum_{j=1}^{J} n_j = n$, and with $\bar{y}_j$ being the sample proportion of Y in cell j. Assuming that sample units in each post-stratification cell is a simple random sample of the population in that cell, the population proportion using the sample s can be estimated using

$$\hat{\theta}^w = \frac{\sum_{j=1}^{J} N_j \bar{y}_j}{\sum_{j=1}^{J} N_j}. \tag{2.2}$$

Let $w_j = N_j/n_j$, we can re-write (2.2) as follows:

$$\hat{\theta}^w = \frac{\sum_{j=1}^{J} \sum_{i=1}^{n_j} w_j y_{ij}}{\sum_{j=1}^{J} \sum_{i=1}^{n_j} w_j} \tag{2.3}$$

where w_j is the post-stratification weight assigned to all sample units in cell j, $j = 1, \dots, J$. The post-stratification estimator in (2.3) requires both N_j and n_j to be known for all post-strata and could yield a very large w_j when n_j is small [3]. It is a challenge to find harmony between having many post-strata such that units in the same cells are homogeneous, and maintaining adequate sample sizes within each post-stratum to

avoid extremely large weights due to small sample sizes [7, 30]. In order to compute a weight for empty sample cells, some post-strata have to be collapsed. The choice of what margins to collapse or which cells to pool is somewhat arbitrary and contradicts the goal of including all auxiliary variables that affect inclusion or response so that the assumption of simple random sampling within post-strata is reasonable [7].

Holt and Smith (1979) proposed the following to estimate the conditional variance of the post-stratified mean,

$$\hat{V} = \sum_{j=1}^{J} (\frac{N_j}{N})^2 (1 - \frac{n_j}{N_j}) \frac{\hat{\sigma}^2}{n_j}$$

where $\hat{\sigma}^2$ is the variance of Y in post-stratum j and $1 - n_j/N_j$ is a finite population correction. An alternative weighting adjustment known as raking can also be used to compute weights when sample post-strata are empty or sparse by using the marginal distribution of the auxiliary variables instead of their joint distribution. Additionally, in cases where the population joint distributions are not available, raking is a common substitute to create weights using population margins.

2.2.2.2 Raking

A related weighting adjustment known as raking can be used to obtain weights w_j to replace w_j in (2.3). The raking method of weighting adjustment utilizes the more commonly available marginal population distributions of the auxiliary variables. Raking weights are obtained via an iterative proportional fitting procedure that begins by modifying sample weights to the marginal distribution of the first auxiliary variable to obtain adjusted sample weights. These adjusted weights are then updated to conform to the marginal distribution of the second auxiliary variable. This process of updating the

adjusted weights is carried out for each auxiliary variable and the first iteration ends when the weights are adjusted using the last auxiliary variable. Subsequent iterations are performed until the weights conform to the marginal distributions of all the auxiliary variables, i.e. the algorithm converges [14, 29].

Unlike post-stratification where the weighted sample distributions of the auxiliary variables conform to their joint distributions in the population, with raking adjustments, weighting results in weighted sample distributions of the auxiliary variables that conform to the marginal population distributions. Although raking can be used to address situations where sample post-strata are sparse or empty by using marginal instead of joint distributions of the auxiliary variable, raking can also be difficult to converge or introduce considerable variation in weights when there are many auxiliary variables or there exist empty or sparse margins [6, 15, 30]. Neither raking or post-stratification methods are ideal when working with continuous auxiliary variables which have to be categorized for use.

2.2.2.3 Response propensity weighting

Rosenbaum and Rubin [37] proposed a response propensity weighting approach to handle non-response that allows continuous auxiliary variables as an alternative to using the distributions of auxiliary variables for weighting adjustment. Here, a regression model for response conditioned on auxiliary variable information is used to estimate the response propensity. The inverse of the predicted response propensities in the sample can then be used as weights in (2.3). However, the effect of this weighting adjustment in reducing non-response bias largely relies on correct specification of the response propensity model, and extreme weights can be generated as a result of very small response propensities [38, 39].

2.2.3 Design-based model-assisted approach

The generalized regression (GReg) estimator provides a model-assisted framework for improving the inefficiency of weighted estimation in the presence of variable weights while reducing bias due to non-response [16]. The GReg estimator for finite population proportions combines the predicted values of the outcome of interest $\hat{y}_i$ from a suitable model, and the response propensity weighted estimator for the residuals $r_i = y_i - \hat{y}_i$ of the sampled units,

$$\hat{\theta}^{gr} = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i + (\sum_{i \in s} r_i/\pi_i)(\sum_{i \in s} 1/\pi_i)^{-1} \tag{2.4}$$

where the estimated population size $\sum_{i \in s} 1/\pi_i$ is the bias calibration term for the residuals r_i and π_i is the response propensity for sample unit i [32]. Variance estimation for the estimator in (2.4) is based on the following expression:

$$\hat{V} = \frac{1}{N^2}\sum_{i \in s} \phi_i (\frac{r_i}{\pi_i} - \hat{b})^2,$$

$$\hat{b} = \frac{\sum_{i \in s} \phi_i \frac{r_i}{\pi_i}}{\sum_{i \in s} \phi_i}$$

with $\phi_i = \frac{n}{n-i}(1 - \pi_i)$ [40, 41]. The GReg estimator is robust to model misspecification if either the response propensity model or the predictive model is correctly specified. We introduce a predictive model framework that can be used to obtain the predictions $\hat{y}_i$ in the following section.

2.2.4 Bayesian modeling approaches

Population proportions can be estimated using prediction models, where survey outcomes are first regressed on auxiliary information using the sample data, and the fitted

model is then used to yield predictions for the survey outcomes among the non-sampled units in the population using the observed auxiliary information for the non-sampled units [6, 7, 30]. Little [7] contrasts this predictive modeling approach with the weighting approaches using Figure 2.1. The model-based predictive estimator of θ is

$$\hat{\theta}^m = \frac{\sum_{j=1}^{J} \left(\sum_{i \in s} y_{ij} + \sum_{i \notin s} \hat{y}_{ij} \right)}{N}, \tag{2.5}$$

where y_{ij} denotes the observed survey outcomes in the sample, and $\hat{y}_{ij}$ denotes the predicted survey outcomes for the non-sampled units based on a prediction model. Equation (2.6) can be re-written to accommodate prediction in data that is unlinked between the population and sample.

$$\hat{\theta}^m = N^{-1} \sum_{j=1}^{J} \left(\sum_{i \in s} (y_{ij} - \hat{y}_{ij}) + \sum_{i=1}^{N} \hat{y}_{ij} \right). \tag{2.6}$$

A number of prediction models are suggested in the literature for aggregate auxiliary data [7, 15, 30, 31].

For continuous outcomes, Little (1993) considered a random intercept regression model for survey inference in the presence of non-sampling errors that assumes a distinct mean and variance for the survey outcome value in each post-stratification cell. The basic normal post-stratification model (BNPM) is proposed as a unified and complete Bayesian model-based foundation for survey inference, particularly inferences from small samples. The model is written

$$(y_i | j, \mu_j, \sigma_j^2) \sim G(\mu_j, \sigma_j^2);$$

where y_i is the value of continuous outcome Y for sample unit i , j identifies the post-stratum, and $G(a, b)$ is a normal distribution with mean a and variance b. If the post-stratum means can be regarded as exchangeable, they can be modeled as coming from a common distribution, yielding a random effects model. In large samples, inference is insensitive to the form of the prior but in the presence of small post-strata, the prior for the model results in partial pooling across post-strata [7].

An alternative model known as the multilevel regression for post-stratification (MRP) model was proposed for estimation of a population mean for a binary survey outcome [30]. MRP extends BNPM by taking advantage of hierarchical structures that may exist in the post-stratification categories to improve the efficiency in overall estimation and the precision of small area estimates [30]. Model-based methods have been shown to yield efficient statistical inference in the presence of dispersed weights when models are well constructed [6, 7]. However, an important consideration with model-based approaches is how best to specify the model. Models induce subjectivity; if the model is seriously misspecified then it can yield inferences that are worse than inferences from weighting methods [4, 10, 11]. This is particularly a concern in the presence of continuous auxiliary information and complex population associations as model-dependent approaches may even perform poorly in large samples where the population model is not correctly specified as even small deviations from the assumed model that are not easily detectable through routine model checking can cause serious problems [18]. Recent work on flexible model-based approaches to avoid misspecification for complex survey designs may offer benefits with respect to survey inference in the presence of non-sampling errors [15, 32, 34, 35].

2.2.5 Bayesian penalized spline regression on weights

In the missing data literature, Little and An (2004) proposed a doubly robust approach to inference about a binary outcome using a penalized spline of propensity prediction model. The logistic regression model for the outcome of interest included the logit-transformed estimated response propensity score using a spline term. We extend this model to the survey sampling setting and propose a Bayesian penalized spline regression model for robust survey inference in the presence of non-response and under-coverage.

Let $p_{ij} = Pr(y_{ij} = 1)$ be the probability that y_{ij} takes a value of 1 for unit i in post-stratum j. We propose a logistic model that assumes a different mean for Y in each post-stratum and includes the survey weights as covariates using a smooth regression:

$$\text{logit}(p_{ij}) = \beta_{0j} + s(w_{ij}), \tag{2.7}$$

$$\beta_{0j} \sim N(\beta_0, \tau^2),$$

where $s(w_{ij})$ is a smooth function of w_{ij} and can be modeled using a spline or Kernel function. Here, I use a penalized spline to model this association [36]. Specifically,

$$s(w_{ij}) = \sum_{p=1}^{P} \beta_p w_{ij}^p + \sum_{k=1}^{K} b_k (w_{ij} - m_k)_+^P, \tag{2.8}$$

$$b_k \sim N(0, \tau_b^2),$$

where w_{ij} can be w_{ij}, w_{ij}^{-1}, $\log(w_{ij})$, or other such appropriate transformation. The constants $m_1 < m_2 < \ldots < m_K$ are K pre-selected fixed knots. A function $(w_{ij} - m_k)_+^P$ is called a truncated spline basis function of degree P, with $(w_{ij} - m_k)_+ = w_{ij} - m_k$ if $w_{ij} > m_k$ and 0 otherwise. Values of p=1,2, and 3 define linear, quadratic, and

cubic penalized splines, respectively. By specifying a normal distribution for b_k, the influence of the L knots are constrained, which is equivalent to smoothing the splines via the penalized likelihood. The smooth function in the proposed model allows a flexible association between the survey outcomes and the weights, and thus it protects against potential model misspecification.

We complete the fully Bayesian modeling specification in equations (2.7) and (2.8) by assuming non-informative prior and hyperprior distributions for the model parameters with

$$\beta_1, \boldsymbol{\beta}_2 \propto 1 \text{ and } \tau, \sigma \sim Cauchy_+(0, 3),$$

where $Cauchy_+(0, 3)$ denotes the positive part of a Cauchy distribution with center 0 and scale 3 [4]. Using Markov chain Monte Carlo (MCMC) simulations, we can obtain posterior draws of θ with

$$\hat{\theta}^{(m,d)} = N^{-1} \sum_{j=1}^{J} \left(\sum_{i \in s} (y_{ij} - \hat{y}_{ij}^{(d)}) + \sum_{i=1}^{N} \hat{y}_{ij}^{(d)} \right),$$

where $\hat{y}_{ij}^{(d)}$ is the predicted binary response for the i^{th} unit in the j^{th} post-stratum obtained from the posterior predictive distribution of y_{ij} in the d^{th} draw, $d = 1, \ldots, D$.

The average of the predictive estimates simulates the estimate of the population proportion $\hat{\theta}^m = D^{-1} \sum_{d=1}^{D} \hat{\theta}^{(m,d)}$. Credible intervals for the population proportion can be formed by splitting the tail areas of the posterior predictive distributions equally between the upper and lower endpoints.

The Bayesian modeling approaches are executed in RStan, an R interface to Stan [42]. Stan is an open-source package for obtaining Bayesian inference using the No-U-Turn Sampler (NUTS), a variant of Hamiltonian Monte Carlo (HMC) [43]. HMC avoids

the random walk behavior by using the gradient of the log-posterior [44]. It converges more quickly than the simpler Markov chain Monte Carlo (MCMC) methods such as random-walk Metropolis [45] and Gibbs [46] sampling. NUTS improves on HMC by eliminating the need to choose the number-of-steps parameter required by HMC and costly tuning runs, which makes it suitable for applications in user friendly Bayesian inference package. We monitor the convergence of our MCMC simulations using the convergence measure $\hat{R}$ that suggests the chains mix well if close to one [42]. We keep 7,500 draws from 3 MCMC chains after 2,500 warm-up draws for each chain. RStan code for the proposed method can be found in Appendix A.

2.3 Simulation study

2.3.1 Design

A simulation study was conducted to assess the performance of the proposed Bayesian multilevel penalized spline regression model along with existing weighting and modeling approaches for inference about a population proportion under the conditions of highly dispersed survey weights and model misspecification. Auxiliary information $\mathbf{Z} = (Z_1, Z_2, Z_3)$ was simulated for a population of size $N = 3,000$ in the form of three correlated binary variables Z_1, Z_2 and Z_3 with the respective marginal probabilities of 0.7, 0.4, and 0.2 and a binary correlation of 0.1, and one continuous variable $Z_4 \sim 0.5z_1 + 0.5z_2 + \text{logNormal}(0.2, 1)$.

We conducted two sets of simulations. In the first simulation, we generated the survey outcome Y, such that $Pr(Y = 1) \approx 0.20$ and $\text{logit}(Pr(Y_{ij} = 1)) = -4.95 + 1.35z_{1j} + (3.15z_{4ij} - 0.75z_{4ij}^2)\text{I}(z_{4ij} \leq 4) + 0.75\text{I}(z_{4ij} > 4)$ for unit i in the post-stratum

j. Let R_{ij} be the response indicator with 1 for inclusion in the sample and 0 otherwise. Samples were selected so that the probability of response $\pi_{ij} = Pr(R_{ij} = 1)$, with $\text{logit}(\pi_{ij}) = -3.2 + z_{1j} + z_{2j} - 1.5z_{3j} + z_{4j} - 0.03z_{4j}^2$ or $\text{logit}(\pi_{ij}) = -5.2 + 2z_{1j} + 2z_{2j} - 3z_{3j} + 2z_{4j} - 0.06z_{4j}^2$. Both of the response models resulted in a sample size of approximately 725, but the second response model yielded bigger variation in survey weights for assessing the performance of various methods in the presence of dispersed weights. In simulation two, the outcome Y was generated such that there was a non-additive association using $\text{logit}(Pr(Y_{ij} = 1)) = -1.4 - z_{1j} - 1.2z_4 + 1.5z_{1j}z_{4j} + 0.05z_{1j}z_{4j}^2$. The response model also involved complex, non-additive association with $\text{logit}(\pi_{ij} = 1) = -2.5 - z_{1j} + z_{2j} - 1.5z_{3j} + z_{1j}z_{4j} - 0.03z_{4j}^2$. This model resulted in samples of size approximately 725. Simulation two assessed the performance of the approaches in the presence of misspecification.

For each generated outcome and sampling scenario, 500 replicates of simulation were obtained and population proportion of $Y = 1$ was estimated. Results are presented for twelve estimators including five weighted estimators, and seven model-based estimators. The weighted estimators included the unweighted estimator (UW), the post-stratification (PS) and raking (R4) estimators where values of Z_4 were categorized using quartiles, the raking estimator where values of Z_4 were categorized at deciles (R10), and the propensity weighted estimator (PR) which included the main effects of auxiliary information $\mathbf{Z}$ as well as a linear spline on Z_4 using 15 equally spaced knots across the range of Z_4. I use quartiles and deciles of Z_4 in post-stratification and raking estimation to avoid sparse and empty sample cells, but increase the number of cut points to obtain 20 equally-sized groups for use in the propensity response model to better model the linear spline. The seven prediction modeling estimators included

1. A basic post-stratification model (Basic PS) that assumes a different mean for Y in each post-stratum: $\text{logit}(p_{ij}) = \beta_{0j}$ with $\beta_{0j} \sim N(\beta_0, \tau^2)$ and prior distributions $\beta_0 \propto 1$ and $\tau \sim Cauchy_+(0, 3)$.

2. A basic post-stratification model that incorporates the main effects of the discrete auxiliary variables and a smooth function of Z_4 using a cubic penalized spline on 15 equally spaced knots (Spline-Cov). Values of Z_4 were categorized into 20 equally-sized groups.

3. The generalized regression estimator (GReg) based on the response propensity model and Spline-Cov prediction model.

4. The proposed penalized spline model using the log of the PS weights for w_j (Spline-PS) and placing 15 knots equally spaced across the range of w_j for estimating the cubic spline function.

5. The proposed penalized spline model using the log of the raking weights (R4) for w_{ij} (Spline-R4) and placing 15 knots equally spaced across the range of w_{ij} for estimating the cubic spline function.

6. The proposed penalized spline model using the log of the raking weights (R10) for w_{ij} (Spline-R10) and placing 15 knots equally spaced across the range of w_{ij} for estimating the cubic spline function.

7. The proposed penalized spline model using the log of the PR weights for w_{ij} (Spline-PR) and placing 15 knots equally spaced across the range of w_{ij} for estimating the cubic spline function.

We evaluated the performance of the estimators using absolute empirical bias, root mean squared error (RMSE), and coverage rate of the 95% confidence or credible

interval (CI). Let $\hat{\theta}_t$ be an estimate of the population proportion θ in the t^{th} simulation, $t = 1, \ldots, 500$. The absolute empirical bias and RMSE are defined as follow,

$$\text{Absolute bias} = \Big| \frac{1}{500} \sum_{t=1}^{500} (\hat{\theta}_t - \theta) \Big|,$$

$$\text{RMSE} = \sqrt{\frac{1}{500} \sum_{t=1}^{500} (\hat{\theta}_t - \theta)^2}.$$

Estimators with smaller absolute bias, smaller RMSE, and 95% CI coverage rate close to the nominal level are desired.

2.3.2 Results

Table 2.2 presents the absolute empirical bias (x100), RMSE (x100), interval width (x100), and coverage rate of the 95% CI (x100) for twelve estimators in simulation one with low variation in the weights. The UW estimator, that ignores the associations between the auxiliary variables and both response and the survey outcome, performs worse than all other estimators in this setting with large bias and RMSE, and poor coverage rate. The weighted estimators yield smaller bias and RMSE as well as improved CI coverage as compared to UW estimation. Further, using the continuous Z_4 auxiliary variable in raking weighted estimation categorized into deciles (R10) versus quartiles (R4) reduces the bias with no penalty of loss in precision. The prediction modeling approaches don't offer huge gains of efficiency over the weighting approaches in the low variation in weights setting of simulation one. The Basic PS model-based estimator reduces bias and RMSE as compared to the UW estimator but has higher RMSE than the other weighted estimators. However, the Spline-Cov model, which is correctly specified for the outcome, has minimally lower RMSE compared with weighting approaches. The

model-assisted GReg estimator performs similar to the prediction modeling approaches; both the PR and the Spline-Cov models are correctly specified for the response and survey outcome, respectively. Our proposed approach including a cubic spline on the log-transformed R10 survey weights (Spline-R10) performs best in this setting with relatively low bias, lowest RMSE, and closest to nominal coverage as compared to all other estimators. Note that the bias and RMSE of the proposed approach using the log-transformed R4 survey weights (Spline-R4) are further reduced by using the R10 weight (Spline-R10). Using the R10 raking weights, which utilize more of the continuous information in Z_4, does a better job of capturing the non-linear association between the outcome and Z_4.

Table 2.3 presents simulation one results for the setting of high variation in the weights. Despite the fact that all the estimators perform worse in the presence of more variation in the weights, the prediction modeling estimators show greater gains of improved efficiency when compared to the weighting estimators in this setting compared to the low variation setting. Weighted estimators can be inefficient in the presence of highly variable weights and modeling approaches offer greatest benefits of improved efficiency in this case. The GReg estimator does not perform well in this setting as the PR model induces very high variation in the weights. However, GReg estimation does have lower RMSE as compared to PR weighted estimation. Although the newly proposed model-based estimators show improved efficiency as compared to the weighting approaches, they have comparable RMSE as the existing prediction modeling approaches. The Spline-R10 model performs best overall among the proposed models with respect to bias, RMSE, and coverage of 95% CI. The Spline-Cov model, which is correctly specified for the outcome, shows the greatest gains in efficiency among all approaches in this scenario of simulation one.

Table 2.4 compares the results for simulation two with non-additive association between the outcome and continuous auxiliary variable Z_4 so that our prediction modeling approaches are misspecified for the outcome. Among the prediction modeling approaches, the proposed Spline-R10 and Spline-PR models perform best. In this situation when models for the outcome are misspecified, the proposed Spline-R10 estimator has lower bias and RMSE, as well as better coverage than the Spline-Cov estimator. Further, the doubly-robust GReg estimator does not perform well here as the PR model is misspecified for response and the Spline-Cov model is misspecified for the outcome.

2.4 Data application

Of interest, is inference about the prevalence of lifetime temper or disruptive mood dysregulation disorder (TDD) among OANG service members. The de-identified, individual level auxiliary information obtained from administrative records for the target population included information on gender, race, rank, age group, and number of years of service and was unlinked to the survey data. The outcome of interest, presence of TDD, is available for all $n = 2,616$ service members completing the survey. Figures 2.2 and 2.3 show the bivariate association between probability of response, prevalence of TDD and the auxiliary variables. The auxiliary variables are both associated with response and the outcome. Further, in Figure 2.2 there is non-linear association between years of service and both the response mechanism and the outcome association. Therefore, because of potential selection bias from non-response and frame under-coverage, the auxiliary data can be used to improve inference about TDD in the OANG study sample. However, using adjustment methods developed for categorical data may not be

the optimal choice as we observe complex association between years of service and the outcome of interest in the sample data.

To facilitate computation of survey weights for weighting estimation, the continuous years of service measure was categorized according to population percentiles. A major challenge is to maintain adequate sample sizes within each post-stratum to avoid empty or sparse post-strata and thus extreme weights. As such, for post-stratification, years of service was categorized according to population quartiles. Cross-classification of the auxiliary variable information for post-stratification resulted in 2 x 2 x 2 x 3 x 4 = 96 cells or post-strata. However only 77 of these cells were non-zero in the population. Further, only 65 cells from the population data were represented in the sample. Raking weights were computed using the marginal distributions of the auxiliary variables with years of service categorized into 16 groups using population percentiles of years of service to better capture the continuous information in the variable. The maximum raking weight was 11.7. It is important to note that the weights in the OANG MHI study data are less variable than the weights in our simulation study. They are also less variable that those typically found in community surveys.

We use the proposed Bayesian predictive modeling approach by fitting a logistic regression for the survey outcome with random intercepts for the 65 post-strata formed by the cross-classification of the auxiliary variables and raking weights as covariates using a linear penalized spline. We keep 15,000 draws from 3 MCMC chains after 5,000 warm-up draws for each chain. Visual inspection of traceplots of the parameters in the model show that the chains converge to the posterior after around 6,000 iterations. Further the $\hat{R}$ convergence measure was close to 1 indicating the chains mix-well. The model-based estimate of lifetime TDD, given by the mean of the posterior predictive estimates, was 20.9% (19.4%, 22.4%).

2.5 Conclusion

In this chapter we proposed a flexible Bayesian model for survey inference that protects against model misspecification in the presence of continuous auxiliary information. We assessed the performance of our proposed method and existing weighting and model-based approaches via simulation study. Our simulations confirmed established findings on the importance of using appropriate statistical analyses to adjust for non-sampling errors such as non-response and under-coverage [3, 28, 47]. Both the weighting and model-based approaches yield population estimates that are more accurate than the estimates without any adjustments in all simulation settings. Moreover, the simulation study showed that our proposed model-based approach using the raking weights as a covariate outperforms the weighting approach, yielding more efficient estimates and closer to the nominal level 95% CI, in the presence of highly dispersed weights. The Spline-Cov model also performs well in the setting of high variation in the weights. In this setting, both our proposed model and the Spline-Cov model are correctly specified for the outcome.

Importantly, our proposed model is robust to misspecification of the model for the survey outcome. By using a penalized spline on the survey weights, we allow flexible association between the auxiliary variable information and the survey outcome. On the other hand, the Spline-Cov model did not perform as well as our proposed method in the case that there is model misspecification, with increased bias and RMSE as well as poorer coverage.

Utilizing the continuous auxiliary variable information helped to improve the precision and efficiency of estimation. Using the raking R10 weights in our proposed model-based estimation (Spline-R10) consistently resulted in reduced bias, improved efficiency,

and better confidence coverage compared to the Spline-R4 estimator. With respect to using the R10 weights in weighted estimation, the R10 weighted estimator consistently reduced bias of estimation as compared to the R4 weighted estimator. However, in the presence of highly variable weights, the R10 weighted estimator was more inefficient than the R4 weighted estimator.

In conclusion, our study promotes the use of a model-based approach that includes the raking weights as a covariate using penalized splines for survey inference in the presence of highly dispersed weights and in the presence of continuous auxiliary information which can have complex associations with survey outcomes in the population. Future work will need to consider the performance of the proposed approach for inference in sub-populations.

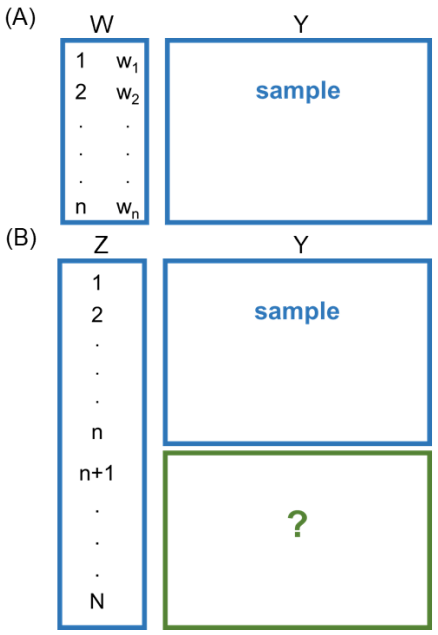

FIGURE 2.1: Survey inference using a weighting approach where a weight is computed for each sampled unit using auxiliary information (A) and a predictive modeling perspective (B).

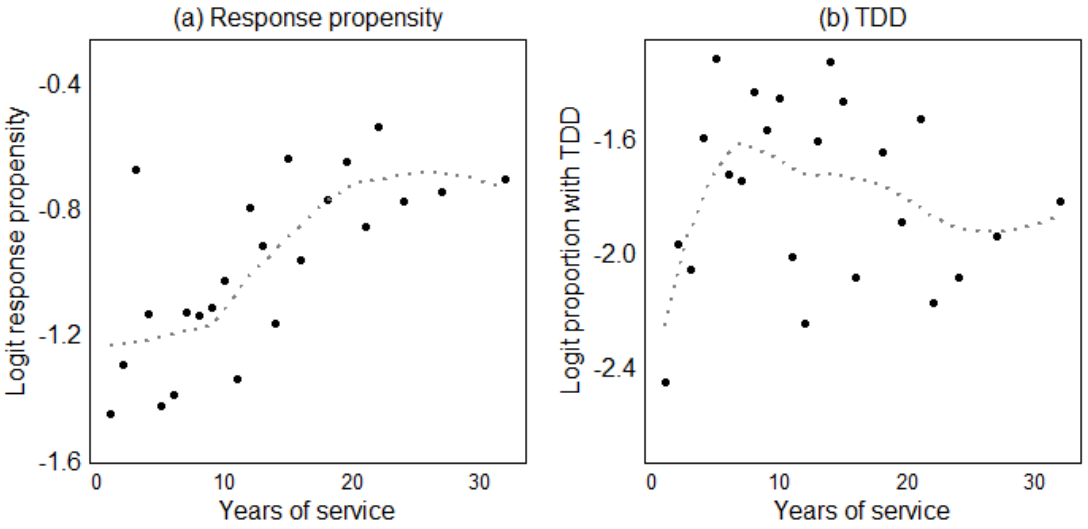

FIGURE 2.2: Bivariate association between probability of responding to survey, probability of having lifetime temper or disruptive mood dysregulation disorder (TDD) and continuous years of service, Ohio National Guard Mental Health Initiative (OANG MHI) Study, 2008-2009. Panel (a) shows the association between the logit response propensity and years of service. Panel (b) shows the association between the logit of having lifetime temper or disruptive mood dysregulation disorder (TDD) and years of service in the unweighted sample.

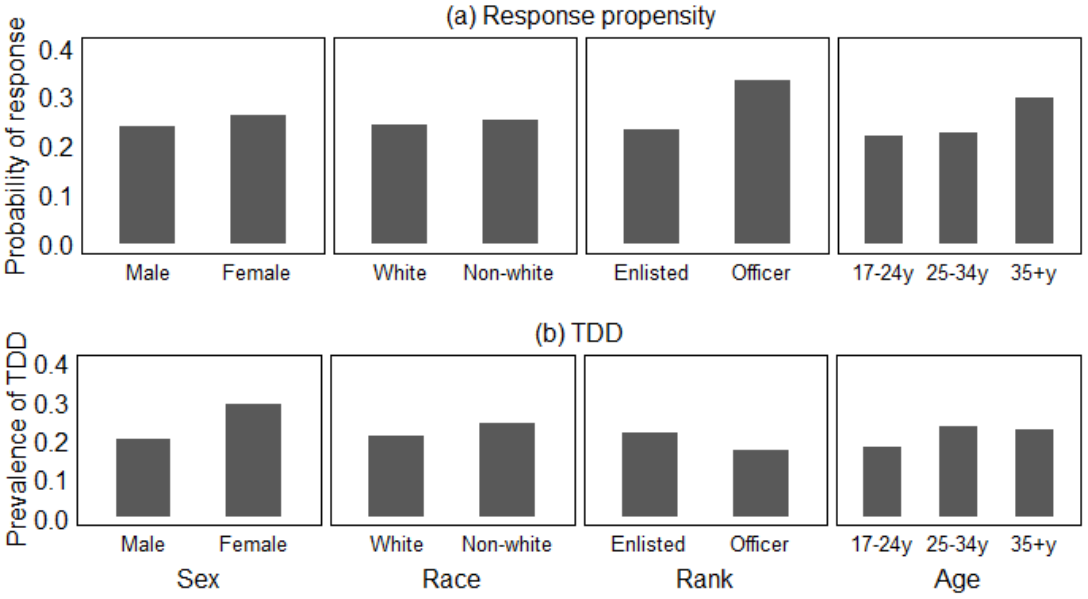

FIGURE 2.3: Bivariate association between probability of responding to survey, probability of having lifetime temper or disruptive mood dysregulation disorder (TDD) and discrete auxiliary variables, Ohio National Guard Mental Health Initiative (OANG MHI) Study, 2008-2009. Panel (a) shows the association for probability of response. Panel (b) shows the association for prevalence of TDD in the unweighted sample.

TABLE 2.1: Distribution of auxiliary variable information in population and unweighted survey sample, Ohio Army National Guard Mental Health Initiative (OANG MHI) Study, 2008-2009.

Measure	Population n(%)	Unweighted sample n(%)
Sex		
Male	9,293(86.2)	2,228(85.2)
Female	1,485(13.8)	388(14.8)
Race		
White	8,761(81.3)	2,298(87.8)
Non-white	2,017(18.7)	318(12.2)
Rank		
Enlisted	9,750(90.5)	2,274(86.9)
Officer	1,028(9.5)	342(13.1)
Marital Status		
Married	4,154(38.5)	1,230(47.0)
Not married	6,624(61.4)	1,386(53.0)
Age group		
17-24 years	4,043(37.5)	881(33.7)
25-34 years	3,746(34.8)	850(32.5)
35 years and older	2,989(27.7)	885(33.8)
Years of service, mean(SD) in years	8.6(7.9)	10.1(8.4)

Table 2.2: Comparison of absolute bias, root mean squared error (RMSE), interval width, and coverage rate of 95 % confidence or credible interval (CI) for estimators from simulation one with low variation in the weights.

Estimator	Bias	RMSE	Width	Coverage
Weighting				
UW	6.6	6.7	5.1	0.0
P-S	0.4	2.5	7.4	84.3
R4	0.6	2.4	8.2	90.2
R10	0.1	2.4	8.4	92.4
PR	0.1	2.8	11.3	96.0
Prediction model				
Basic PS	1.9	2.8	7.9	83.0
Spline-Cov	1.0	2.3	8.0	91.4
Doubly robust				
GReg	1.0	2.3	9.8	96.0
Proposed method				
Spline-PS	0.4	2.5	8.9	91.2
Spline-R4	0.7	2.5	8.7	90.8
Spline-R10	0.2	2.2	8.5	95.0
Spline-PR	0.0	2.5	9.1	93.8

TABLE 2.3: Comparison of absolute bias, root mean squared error (RMSE), interval width, and coverage rate of 95 % confidence or credible interval (CI) for estimators from simulation one with high variation in the weights.

Estimator	Bias	RMSE	Width	Coverage
Weighting				
UW	8.6	8.7	4.6	0.0
P-S	1.0	4.6	9.3	63.1
R4	0.5	4.2	13.6	91.4
R10	0.2	4.4	14.0	89.6
PR	0.4	6.5	19.1	86.8
Prediction model				
Basic PS	3.3	4.2	10.4	79.0
Spline-Cov	1.9	3.2	10.7	89.2
Doubly robust				
GReg	1.1	4.6	9.6	78.0
Proposed method				
Spline-PS	0.9	4.6	12.8	80.3
Spline-R4	1.4	4.4	12.9	85.6
Spline-R10	1.1	3.4	11.2	88.7
Spline-PR	0.2	4.7	14.7	89.0

Table 2.4: Comparison of absolute bias, root mean squared error (RMSE), interval width, and non-coverage rate of 95 % confidence or credible interval (CI) for estimators from simulation two.

Estimator	Bias	RMSE	Width	Coverage
Weighting				
UW	20.2	20.2	7.1	0.0
P-S	2.7	3.2	5.4	48.2
R4	5.4	5.6	6.2	3.4
R10	2.7	3.2	6.1	55.0
PR	1.7	2.9	9.0	87.2
Prediction model				
Basic PS	3.9	4.2	6.7	29.1
Spline-Cov	1.9	2.5	6.5	80.4
Doubly robust				
GReg	2.7	4.6	11.1	74.0
Proposed method				
Spline-PS	2.8	3.2	6.3	58.4
Spline-R4	3.1	3.5	6.4	46.7
Spline-R10	1.3	2.0	6.1	89.1
Spline-PR	1.1	1.9	6.2	92.5

Chapter 3

Stratified Bayesian penalized spline model for domain estimation

3.1 Introduction

Since the beginning of the 21st century, the US Army has relied more on its reserve component both domestically and internationally than at any other time since the Korean war [48]. The increasing share of women and racial minorities in the ranks of reserve personnel since the 2000s [49] coupled with the greater risk of long-term psychopathology among US military reserve forces compared to active duty counterparts [21–23, 50], make assessing the mental health among gender and racial sub-populations in the reserve component of the US military essential. The Ohio Army National Guard Mental Health Initiative (OANG MHI) study provides important information about the prevalence and risk factors of mental health related outcomes among National Guard service members.

Although the target population for the OANG MHI study is all serving members of the OANG between June 2008 and February 2009, using the study data for estimation of the prevalence of important mental health indicators in gender by race sub-populations can be very informative as a secondary analysis. In order to conduct estimation at the domain or sub-group level, there must be appropriate data at this level.

Cross-classifying the final sample of 2,616 OANG service members by gender and race, results in 1,996 white and 232 non-white male, and 302 white and 86 non-white female soldiers. While surveys like the OANG MHI study provide a cost-effective way of generating reliable prevalence estimates for aggregates of domains, often they may not have sufficiently sized samples to produce reliable estimates for domains. Further, due to non-sampling errors such as non-response and sampling frame under-coverage, the sample data from the OANG MHI study is subject to selection bias [10, 51–53]. In this setting, estimation using auxiliary data from administrative records, the census, or other large registries, becomes an important tool not only in reducing bias due to non-sampling errors, but in improving domain estimation in sub-groups with small sample size [7, 10, 52, 53]. Using auxiliary data via weighting adjustments such as post-stratification or raking can reduce bias if, conditioned on the auxiliary information, the non-sampling errors are ignorable. However, these estimators can be inefficient in the presence of small sample sizes and although model-based approaches offer advantages of improved efficiency in the presence of sparse data, they can be subject to misspecification [7, 10, 51–53].

The auxiliary information available for the OANG population of 10,778 service members included gender (male, female), race (white, non-white), rank (enlisted, officer), age (17-24 years, 25-34 years, 35 years and older), and number of years in service. Figure 3.1 shows the association between the outcome of interest and continuous years of service

stratified by gender and race sub-groups. The association between the log odds of having lifetime temper or disruptive mood dysregulation disorder (TDD) and years of service differs between groups, and is non-linear within groups. Having a flexible model that is robust to misspecification is particularly important for domain estimation when the survey data is subject to selection bias, as is the case in the OANG MHI study.

In the previous chapter, we proposed a Bayesian penalized spline regression model for survey inference that utilized auxiliary information to address issues of non-response and sampling frame-under-coverage. The proposed model allowed distinct means of survey outcomes in post-strata defined by the joint distribution of auxiliary variables and incorporated survey weights using a penalized spline to allow flexible association with the survey outcome. Simulation study found that the proposed prediction modeling approach performed better than weighting approaches in the presence of highly dispersed weights and was robust to model misspecification. However, it is not clear that this model is appropriate for domain estimation, particularly in domains with small sample size. In the missing data literature, Zhang and Little (2009) proposed a stratified penalized spline of propensity prediction approach for robust inference about conditional means.

In this chapter, we extend their approach to the survey sampling setting and propose an extension to our Bayesian penalized spline model to facilitate robust estimation of sub-population proportions. This stratified regression model builds on our previous approach by including an interaction of the penalized spline of the survey weights and the auxiliary variable defining the domain estimand of interest. The following section describes the notation and proposed stratified model in detail. In Section 3 we conduct a simulation study to demonstrate the performance of the stratified model as compared to the previously proposed penalized spline model as well as common weighting approaches for domain estimation under various conditions. Section 4 applies the proposed method

to estimating the prevalence of lifetime TDD among OANG gender by race sub-groups. We conclude the chapter with general suggestions in Section 5.

3.2 Methods

3.2.1 Setting and notation

For a population of N units, let Y be a binary survey outcome of interest taking a value of 0 or 1. Let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_{p-1})^T$ be $(p-1)$ discrete auxiliary variables and Z_p be a continuous auxiliary variable, and both $\mathbf{Z}$ and Z_p are observed for all the units in the population. Values of Z_p can be grouped using m cut-off values to create a discrete variable Z_p^* with $(m+1)$ groups. Consider that X is a discrete auxiliary variable with $l = 1, \ldots, L$ levels. Let s denote a sample of size n selected from the population with survey outcome values, $y_1 \ldots y_n$. We assume that the population data are de-identified and thus the sample data cannot be linked to the population data on the individual level. Of interest is estimation of the population proportion, $\theta_l = \bar{Y}_l$ within each domain l.

Define the conditioning auxiliary variable,

$$x_{il} = \begin{cases} 1, & \text{if } x_i = l, \\ 0, & \text{otherwise} \end{cases}$$

for $l = 1, \ldots, L$.

Then the population proportion of Y in domain l can be written,

$$\theta_l = \bar{Y}_l = \frac{\sum_{i=1}^{N_l} y_{il}}{N_l} \tag{3.1}$$

where $\bar{Y}_l$ is the mean of Y for population units belonging to domain l and N_l is the population size of domain l.

3.2.2 Stratified Bayesian penalized spline regression on weights

In the proposed stratified model for the survey outcome, we assume a distinct mean within each post-stratum j, which is further assumed to follow a normal distribution with a common variance and estimated using a fully Bayesian approach. We extend this multilevel model to include factor-by-curve interaction by fitting a stratified logit regression for the observed stratum specific prevalence, p_{ij}, using a partially linear model on the survey weight, w_{ij}, that is flexible to accommodate non-linear associations between the auxiliary information and the outcome of interest. Consider the triple $(w_{ij}, x_{ijl}, y_{ij})$ where w_{ij} is the survey weight, x_{ijl} is an indicator variable identifying the domain membership, and y_{ij} represents the binary survey outcome of interest for the i^{th} sampled unit in post-stratum j. A stratified cubic penalized spline model for the observed stratum specific prevalence $p_{ij} = Pr(y_{ij} = 1|z_{ijl}, w_{ij})$ is then written:

$$\begin{aligned}\text{logit}(p_{ij}) = \beta_{0j} + \sum_{p=1}^{P} \beta_p w_{ij}^p + \sum_{k=1}^{K} b_k (w_{ij} - m_k)_+^P + \sum_{l=2}^{L} x_{ijl}(\gamma_{0l} + \sum_{p=1}^{P} \gamma_p w_{ij}^p) \\ + \sum_{l=1}^{L} x_{ijl} \Big\{ \sum_{k=1}^{K} c_k^l (w_{ij} - m_k)_+^P \Big\} \end{aligned} \tag{3.2}$$

$$\beta_{0j} \sim N(\beta_0, \tau^2),$$

$$b_k \sim N(0, \tau_b^2),$$

$$c_k^l \sim N(0, \tau_{cl}^2), l = 1, \ldots, L.$$

where constants $m_1 < m_2 < \ldots < m_K$ are K pre-selected fixed knots for the truncated cubic spline basis function, $(w_{ij} - m_k)_+$, with $(w_{ij} - m_k)_+ = w_{ij} - m_k$ if $w_{ij} > m_k$ and 0 otherwise. By specifying a normal distribution for b_k, the influence of the K knots is constrained, which is equivalent to smoothing the splines via the penalized likelihood. The stratified smooth function in the proposed model allows a flexible association between the survey outcomes and the weights at levels of the conditioning variable, and thus it protects against potential model misspecification. This is analogous to the framework proposed by Coull (2001).

To perform inference, we assume non-informative prior and hyperprior distributions [4] for the parameters of interest $\beta_p, \gamma_p \propto 1$ and $\tau, \tau_b, \tau_{cl} \sim Cauchy_+(0, 3)$, and obtain draws from the posterior distributions of θ_l using Markov chain Monte Carlo (MCMC) simulations via

$$\hat{\theta}_l^{(d)} = N_l^{-1} \sum_{j=1}^{J} \left(\sum_{i \in s} (y_{ijl} - \hat{y}_{ijl}^{(d)}) + \sum_{i=1}^{N_l} \hat{y}_{ijl}^{(d)} \right),$$

where $\hat{y}_{ijl}^{(d)}$ is the predicted binary response for the i^{th} unit in the j^{th} post-stratum of domain l obtained from the posterior predictive distribution of y_{ijl} in the d^{th} draw, $d = 1, \ldots, D$. The average of the predictive estimates simulates the estimate of the conditional population proportion $\hat{\theta}_l = D^{-1} \sum_{d=1}^{D} \hat{\theta}_l^{(d)}$. Credible intervals for the conditional population proportion can be formed by splitting the tail areas of the posterior predictive distributions equally between the upper and lower endpoints.

3.3 Simulation study

3.3.1 Design

The performance of the proposed stratified Bayesian penalized spline regression model for domain estimation as compared to select weighted estimators as well as our previously proposed Bayesian penalized spline regression was evaluated using a simulation study. Auxiliary information $\mathbf{Z} = (Z_1, Z_2, Z_3)$ was simulated for a population of size $N = 3,000$ in the form of three correlated binary variables Z_1, Z_2 and Z_3 with the respective marginal probabilities of 0.7, 0.4, and 0.2 and a binary correlation of 0.1, and one continuous variable $Z_4 \sim 0.5z_1 + 0.5z_2 + \text{logNormal}(0.2, 1)$. Of interest is the estimation and inference about the conditional proportion of $Y = 1$ in sub-populations defined by auxiliary variables.

In our simulation setting the marginal proportion $Pr(Y = 1) \approx 0.20$ and samples of size approximately 725 were obtained. The outcome Y was generated such that there was a non-additive association using $\text{logit}(Pr(Y_i = 1)) = -1.4 - z_1 - 1.2z_4 + 1.5z_1z_4 + 0.05z_1z_{4j}^2$. The response model also involved complex, non-additive association with $\text{logit}(\pi_i = 1) = -2.5 - z_1 + z_2 - 1.5z_3 + z_1z_4 - 0.03z_4^2$. In this simulation, we assessed the performance of the described approaches for estimation of the proportion in the overall population and in sub-populations defined by Z_1 such that group $Z_1 = 0$ had domain size $n \approx 115$ and group $Z_1 = 1$ had size $n \approx 610$. The population associations between the continuous auxiliary variable Z_4 and both the outcome and response at levels of Z_1 are illustrated in Figure 3.2.

We obtained 500 replicates and estimated the proportion of $Y = 1$ in the population and sub-populations defined by Z_1. Results are presented for nine estimators

including the unweighted estimator (UW), four weighted estimators, and four model-based estimators. The weighted estimators included the post-stratification (PS) and raking (R4) estimators where values of Z_4 were categorized using quartiles, the raking estimator where values of Z_4 were categorized at deciles (R10), and the propensity weighted estimator (PR) which included the main effects of auxiliary information $\mathbf{Z}$ as well as a linear spline on Z_4 using 15 equally spaced knots across the range of Z_4. We include four Bayesian model-based estimators, including

1. A penalized spline model using the log-transformed R10 weights (Spline-R10)
2. A penalized spline model using the log-transformed PR weights (Spline-PR)
3. The proposed stratified penalized spline model with interaction between Z_1 and cubic spline function on the log-transformed R10 weights (S-Spline-R10)
4. The proposed stratified penalized spline model with interaction between Z_1 and cubic spline function on the log-transformed PR weights (S-Spline-PR)

In the Bayesian predictive models, similar to Chapter 2, we place 15 knots equally spaced across the range of w_{ij} for estimating the cubic spline function. We evaluate the performance of the estimators using absolute empirical bias, root mean squared error (RMSE), and coverage rate of the 95% confidence or credible interval (CI). Estimators with smaller absolute bias, smaller RMSE, and 95% CI coverage rate close to the nominal level are desired.

3.3.2 Results

We present simulation results for nine estimators of the proportion of $Y = 1$ in the overall population and in domains defined by the cross-classification of auxiliary variables in

Table 3.1. For overall estimation, the weighted estimators are inefficient with poor confidence coverage as compared to the proposed prediction modeling approaches. We observe gains in efficiency and improved confidence coverage associated with using the stratified Bayesian modeling approach (S-Spline-R10 versus Spline-R10, and S-Spline-PR versus Spline-PR). In this setting, the proposed stratified model is correctly specified for the population association in the outcome.

For estimation in domains defined by Z_1, we see that in group $Z_1 = 0$ which is the smaller domain, the UW estimator has lowest RMSE and close to nominal coverage. From Figure 3.2 we see that conditioned on Z_1, the non-response is ignorable with respect to the outcome of interest. Importantly, in this small group we see the proposed stratified spline models have closer to nominal coverage and lower RMSE than Spline-R10 and Spline-PR. In group $Z_1 = 1$ where there is non-linear association between Z_4 and the outcome of interest, the proposed models outperform weighting approaches. In particular, the Spline-R10 and Spline-PR approaches perform well in this setting with low RMSE and close to nominal coverage.

3.4 Data application

Our proposed modeling approaches were used to estimate lifetime TDD prevalence for all OANG service members and for domains defined by gender x race. As mentioned previously, auxiliary information was obtained from administrative files. This included information on gender, race, rank, age, and number of years in service.

In our modeling approach we assume a distinct mean in each post-stratum defined by the cross-classification of auxiliary information and utilize raking weights computed using the the marginal distributions of the auxiliary variables with years of service

categorized into 16 groups using population percentiles of years of service to better capture the continuous information in the variable for both models. Similar to the previous chapter, we keep 15,000 draws from 3 MCMC chains after 5,000 warm-up draws for each chain. We assess model convergence by visual inspection of traceplots of the parameters and using the $\hat{R}$ convergence measure where values close to 1 indicate that the chains mix-well.

The results presented in Table 3.2 show that both methods result in similar estimates for the prevalence of lifetime TDD in the OANG population. However, the interval width associated with the stratified spline model is minimally narrower that that of the spline model. The estimates and interval widths for estimation in the largest sub-population, male white service members, are very similar. For smaller groups, the spline model-based approach produces estimates with smaller interval widths as compared to the stratified spline model.

3.5 Conclusion

In this chapter, we proposed an extension of our Bayesian model for survey inference to a stratified model and applied it to survey inference of proportion in the overall population as well as within sub-populations. In overall population proportion estimation the stratified modeling approach performed well, with lower RMSE than the non-stratified modeling approach. For estimation in domains, we saw that both proposed models outperformed weighted estimators with respect to absolute bias, RMSE, and coverage rate when there was complex associations within domains. In domains of small size, the stratified model showed promise with improved efficiency and better confidence coverage as compared to the non-stratified model. In application, the width

of the interval associated with estimates from the stratified model were wider than those from the non-stratified proposed modeling approach in the smallest sample domins. As such, an important consideration of the proposed stratified approach in domains with small sample size is whether there is sufficient data to support efficient estimation. It is possible that the stratified modeling approach may lead to an overfit model for the survey outcome.

In conclusion, this chapter proposes a stratified modeling approach for domain estimation that includes an interaction between the conditioning variable and a penalized spline on the survey weights. We find that it performs well in small size domains and offers further protection against model misspecification in estimation of population proportion when compared to the non-stratified penalized spline approach.

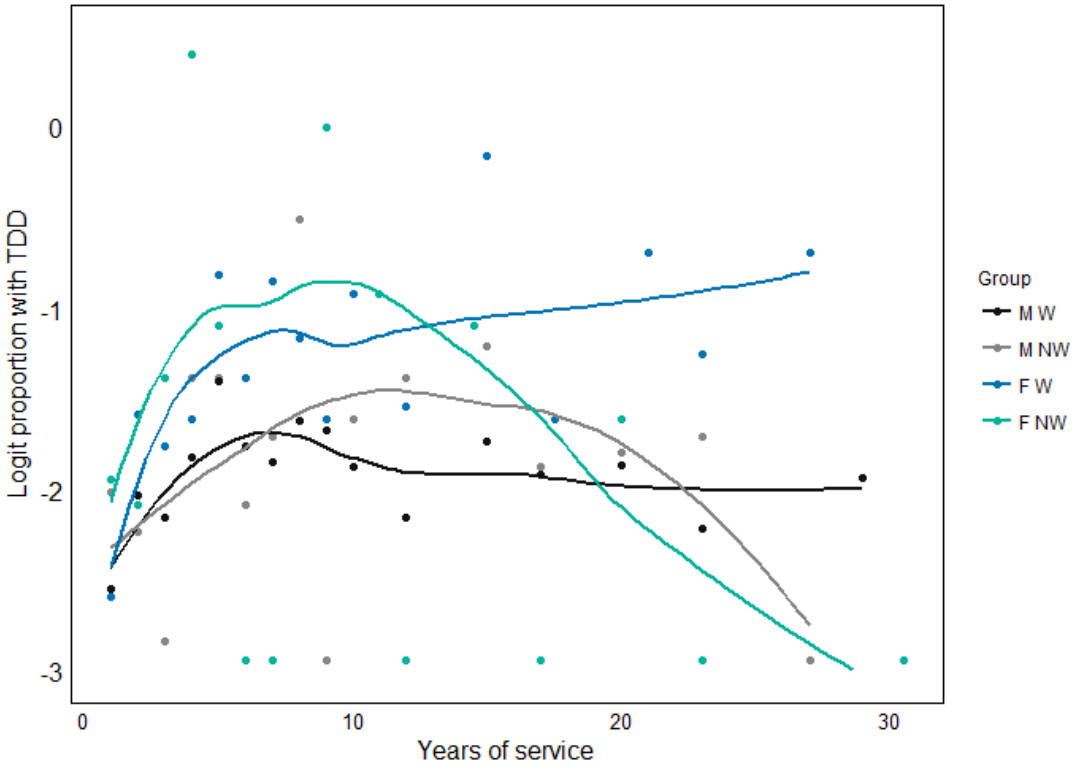

FIGURE 3.1: Association between the logit of having current temper or disruptive mood dysregulation disorder (TDD) and years of service among gender by race domains in the unweighted sample, Ohio National Guard Mental Health Initiative (OANG MHI) Study, 2008-2009.

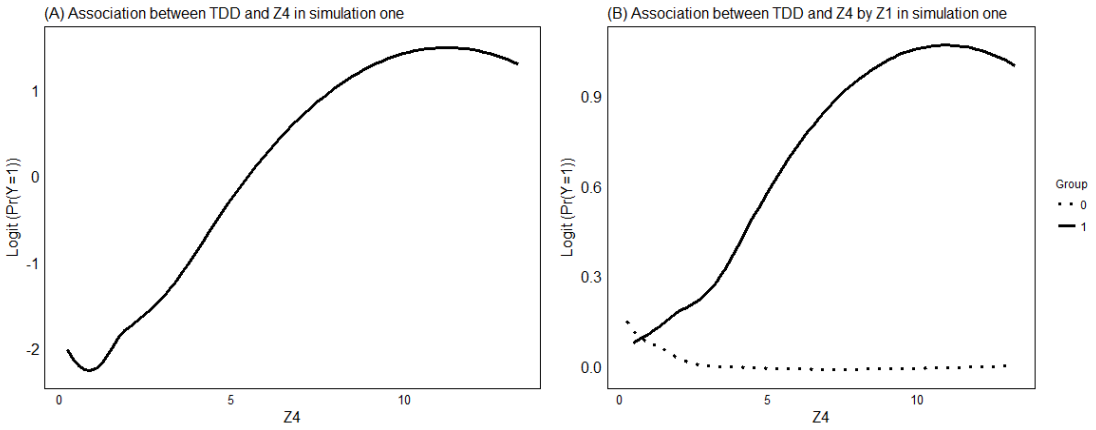

FIGURE 3.2: Association between the logit of $Pr(Y = 1)$ and auxiliary variable Z_4 overall and by categories of Z_1 in simulation one.

TABLE 3.1: Comparison of absolute bias, root mean squared error (RMSE), interval width, and coverage rate of 95 % confidence or credible interval (CI) for estimators from simulation one.

Estimator	All				Group 0				Group 1			
	Bias	RMSE	Width	Cov	Bias	RMSE	Width	Cov	Bias	RMSE	Width	Cov
Weighting												
UW	20.2	20.2	7.1	0.0	0.9	2.4	8.8	85.1	18.5	18.6	7.7	0.0
P-S	2.7	3.2	5.4	48.2	0.1	2.7	8.9	87.3	3.3	3.9	6.6	53.4
R4	5.4	5.6	6.2	3.4	1.5	3.6	11.9	91.9	7.0	7.3	7.1	3.3
R10	2.7	3.2	6.1	55.0	1.3	3.5	11.4	90.8	3.4	3.9	7.0	51.2
PR	1.8	2.9	8.9	87.9	1.0	4.3	14.0	88.6	1.6	2.7	9.3	94.8
Proposed method												
Spline-R10	1.3	2.0	6.1	89.1	1.9	3.2	10.6	93.0	0.0	1.8	6.9	94.8
Spline-PR	1.1	1.9	6.2	92.5	3.1	4.1	11.0	81.4	0.3	1.7	6.9	95.1
S-Spline-R10	1.2	2.0	6.0	89.1	0.2	2.7	10.0	93.2	0.5	2.0	7.3	93.7
S-Spline-PR	0.8	1.8	9.0	94.2	0.4	2.7	9.8	93.2	0.5	2.0	7.8	94.8

TABLE 3.2: Estimated prevalence and 95 % confidence or credible interval (CI) in domains defined by gender and race for lifetime temper or disruptive mood dysregulation disorder (TDD) based on proposed modeling approaches, Ohio Army National Guard Mental Health Initiative (OANG MHI) Study, 2008-2009

Estimator Interval width	All $\hat{p}$ (95 % CI)	Male white $\hat{p}$ (95 % CI)	Male non-white $\hat{p}$ (95 % CI)	Female white $\hat{p}$ (95 % CI)	Female non-white $\hat{p}$ (95 % CI)
Spline-R10	0.21 (0.19,0.22)	0.20 (0.18,0.22)	0.22 (0.17,0.26)	0.26 (0.22,0.30)	0.26 (0.20,0.35)
S-Spline-R10	0.21 (0.19,0.22)	0.20 (0.18,0.21)	0.21 (0.16,0.26)	0.27 (0.23,0.32)	0.28 (0.19,0.38)

Chapter 4

Estimating current and undiagnosed depression among Hispanics/Latinos of different national backgrounds

4.1 Introduction

Depression is a leading cause of disability, contributing to increased health care costs and decreased workplace productivity and quality of life [55–59]. This coupled with the fact that Hispanics in the U.S. are the largest minority group, heterogeneous with regard to place of origin, and potentially different with respect to the diagnosis and treatment of mental health conditions, has increased the demand for within ethnicity estimates of the prevalence of mental health conditions [55, 60–62].

Many epidemiologic studies examining within ethnicity depression prevalence in U.S. Hispanics have focused on Mexican, Cuban, and Puerto Rican subgroups and report 12-month depression prevalence [62–64]. The Hispanic Health and Nutrition Examination Survey (Hispanic HANES), the first population-based health survey of U.S. Hispanics, found significantly higher 12-month major depression prevalence among Puerto Ricans (6.9%) than in Cuban Americans (2.5%) and Mexican Americans (2.8%) [64]. A more recent study examining within Hispanic ethnicity past-year depression in the U.S. found similar rankings with highest prevalence among Puerto Ricans [63]. However, results from González et al. (2010) found higher national 12-month major depression prevalence estimates for Puerto Ricans and Mexican Americans with rates of 11.9%, 8.0%, respectively. Although, these 12-month depression estimates are important, having estimates related to short-term depression symptomology will further inform and enhance medical practice and care provision.

The Hispanic Community Health Study, a 2008-2011 cross-sectional study of Hispanic/Latinos, estimated past-week depression prevalence using the Center for Epidemiological Studies Depression Scale (CES-D) in U.S. Mexicans, Puerto Ricans, Dominicans, Cubans, Central Americans, and South Americans [60]. These findings indicate that, just as with past-year depression, Puerto Ricans (38.0%) had the highest prevalence of past-week depression followed by Cubans (27.9%), Dominicans (27.4%), Central Americans (24.9%), South Americans (24.2%), and then Mexicans (22.3%). However, the authors found rather high levels of short-term depression symptomology with prevalence rates three times greater than estimates reported in previous studies for 12-month depression. Because depression is a debilitating health condition, often co-morbid with serious chronic illness and poor social conditions, identifying depression symptomology in the short-term is important as having undiagnosed and thus untreated depression can

worsen and significantly diminish health [56, 57, 60, 65–67].

We were unable to find current estimates of undiagnosed depression among community-based Hispanic ethnicity populations in the literature. However, a study of patients with co-morbid depression and diabetes found that approximately 14% of depression cases among Hispanics with diabetes went undiagnosed [68]. A more recent cross-sectional analysis of Caribbean-origin Hispanics with poorly controlled diabetes found that while 52.8% of the sample had depression, only 21.4% reported taking antidepressants [66]. Finally, results from the household components of the 2012 and 2013 Medical Expenditure Panel Surveys found that screen-positive depression was nearly 5 times more prevalent among adults in the lowest income group than the highest income group [65].

The 2015 Washington Heights Community Survey (WHCS) was a cross-sectional survey of residents of Washington Heights, New York, that aimed to provide a population based health assessment of a predominantly Hispanic/Latino and low-income neighborhood with high rates of foreign-born residents [69]. The information generated from this survey is extensive and can help to provide a picture of the community in terms of ethnicity and mental health. As the survey assessed short-term clinical depression using the Patient Health Questionnaire-9 (PHQ-9) diagnostic tool [70] and obtained respondent report of ever diagnosed with depression by a health provider (diagnosed depression), it provides a unique opportunity to estimate the prevalence rates of current depression and undiagnosed depression by Hispanic/Latino origin.

4.2 Methods

4.2.1 Data source

The data for this study were drawn from the 2015 WHCS. The WCHS was administered by the Global Research Analytics for Public Health group at Columbia University Medical Center between March and September of 2015 to 2,489 sample households in Washington Heights, New York as part of a community assessment funded by the NewYork-Presbyterian Hospital. Eligible participants had to be 18 years or older, a resident of zip codes 10032 and 10033, and be able to complete the telephone interview in either English or Spanish [69]. The survey included items on neighborhood social and economic conditions, health care access, general health and health conditions and was conducted using both an address-based sample (ABS) and a cell phone random digit dial (RDD) sample. The survey had an American Association for Public Opinion Research (AAPOR) Response Rate of 16.8% [69]. Survey weights were computed to account for sampling design, survey non-response, and sample frame under coverage.

4.2.2 Measures of interest

The current depression status of each respondent was determined using the Patient Health Questionnaire-9 (PHQ-9), a nine-item module from the more comprehensive full PHQ that can be used to diagnose up to eight Diagnostic and Statistical Manuals of Mental Disorders, version 4 (DSM-IV) conditions [70, 71]. The PHQ-9 is used to classify and diagnose major depression via questions that ask the respondent to say how often in the past two weeks they experienced specific feelings such as low interest in daily activities, helplessness, trouble sleeping, low energy, or poor appetite using responses not at all, several days, more than half the days, or nearly every day. As a severity

measure, the PHQ-9 score can range from 0 to 27 as each item is scored from 0 (not at all) to 3 (nearly every day). A PHQ-9 score ≥ 10 was classified as clinical depression in this study [71, 72]. The survey also included self-report of ever being told by a doctor or medical practitioner that respondent has depression. To avoid response bias, these mental health related items were included in the survey as an interactive voice response (IVR) module, so that participants could hear recorded versions of the survey items and respond via their phone keypad, rather than answering to a live interviewer. Undiagnosed depression was defined as having PHQ-9 score ≥ 10 but reporting never being diagnosed with depression by a medical practitioner.

Respondents were classified into Hispanic/Latino sub-ethnicities using a combination of two survey items. The first item asked respondents to self-identify as either Hispanic/Latino, or non-Hispanic/Latino black, white, or other race. The second question asked respondents identifying as Hispanic/Latino to specify their national background as Dominican, Puerto Rican, Mexican, Central American, South American, Cuban, or European/other.

4.2.3 Statistical analysis

We first provide weighted estimates of prevalence of current depression (having PHQ-9 score ≥ 10 vs. < 10) and undiagnosed depression (yes vs. no) in Hispanic ethnicity subgroups. However, since sample size in each sub-ethnicity group can be small (e.g. there are only 32 survey respondents who are Mexican), the traditional weighted estimates of prevalence in each small sub-ethnicity group can be unstable. As such, we utilized our proposed Bayesian regression model-based approach that has been found in simulation to perform better than weighting approaches for survey inference in domains with small sample size.

4.2.3.1 Bayesian models for survey inference

This modeling approach fits a logistic regression for the domain-specific prevalence for respondent i in domain or sub-ethnicity l, $p_{il} = Pr(Y_l = 1)$, using a linear spline on the survey weight, w_{il}, that is flexible to accommodate non-linear associations between the auxiliary information and the outcome of interest. Specifically,

$$\begin{aligned} \text{logit}(p_{il}) &= \beta_{0l} + \beta_1 w_{il} + \sum_{k=1}^{K} b_k (w_{il} - m_k)_+ \\ \beta_{0l} &\sim N(\beta_0, \tau^2) \\ b_k &\sim N(0, \tau_b^2). \end{aligned} \tag{4.1}$$

The constants $m_1 < m_2 < \ldots < m_K$ are K pre-selected fixed knots. To obtain the domain estimates, we assume non-informative prior and hyperprior distribution for the parameters of interest and obtain draws from the posterior distribution of our parameter of interest using MCMC simulations. We can then take the average of the predictive estimates, $\hat{\theta}_l^{(d)} = N_l^{-1} \left(\sum_{i \in s} (y_{il} - \hat{y}_{il}^{(d)}) + \sum_{i=1}^{N_l} \hat{y}_{il}^{(d)} \right)$ to get the estimate of the conditional population proportion, $\hat{\theta}_l = D^{-1} \sum_{d=1}^{D} \hat{\theta}_l^{(d)}$. Here, $\hat{y}_{il}^{(d)}$ is the predicted binary response for the i^{th} unit in domain l obtained from the posterior predictive distribution of y_{il} in the d^{th} draw, $d = 1, \ldots, D$.

4.2.3.2 Finite population Bayesian bootstrap

Because survey weight w_{il} information is unavailable for the non-sampled population units, we cannot compute $\hat{y}_{il}^{(d)}$ directly and use a Bayesian bootstrap procedure for unequal probability sample designs in finite populations to generate synthetic data for prediction [73, 74]. The finite population Bayesian bootstrap (FPBB), developed by Lo

(1988) and extended by Cohen (1997) to accommodate unequal probability of selection survey designs, can be used to generate synthetic data utilizing available sample weights. Let s_i indicate sample unit i, $i = 1, \ldots, n$. The FPBB proceeds by selecting a sample of size $N - n$: $s_1^*, s_2^*, \ldots, s_{N-n}^*$ by drawing s_k^* from $s_1, s_2, \ldots, s_n$ so that s_i is selected with probability

$$\frac{w_i - 1 + l_{i,k-1}(\frac{N-n}{n})}{N - n + (k-1)(\frac{N-n}{n})}$$

where w_i is the survey weight associated with sample unit i, and $l_{i,k-1}$ = number of bootstrap selections of s_i among $s_1^*, s_2^*, \ldots, s_{k-1}^*$. The FPBB population is then formed by $s_1, s_2, \ldots, s_n, s_1^*, s_2^*, \ldots, s_{(N-n)}^*$. Procedure is repeated B times to generate B synthetic populations. Dong et al. (2014) provides a theoretical proof for Cohen 1997 weighted FPBB procedure. The function wtpolyap in the R package *polypost* can be used to obtain draws from a weighted Pólya urn [76].

For each synthetic population b, $b = 1, \ldots, B$, predictive estimates of the domain proportion are obtained via

$$\hat{\theta}_l^{b(d)} = N_l^{-1} \left(\sum_{i \in s} (y_{il} - \hat{y}_{il}^{b(d)}) + \sum_{i=1}^{N_l} \hat{y}_{il}^{b(d)} \right) \tag{4.2}$$

The average of the predictive estimates simulates the Bayesian model-based estimator of the population proportion in b^{th} bootstrap, $\hat{\theta}_l^b = D^{-1} \sum_{d=1}^{D} \hat{\theta}_l^{b(d)}$. The prevalence estimate $\hat{\theta}_l = B^{-1} \sum_{b=1}^{B} \hat{\theta}_l^b$ is the mean of the $\hat{\theta}_l^{(b)}$s and the variance in each group is $(1 + B^{-1})V_B$ where $V_B = (B-1)^{-1} \sum_{b=1}^{B} (\hat{\theta}_l^{(b)} - \hat{\theta}_l)^2$ [73].

In this application, the FPBB procedure is implemented by first drawing a weighted sample of size $N - n = 64,306 - 1,462 = 62,844$ using the provided survey weights to

form FPBB population. Here, $N = 64,306$ is the size of the target population i.e. population size of Hispanic residents of Washington Heights, New York, over 18 years of age residing in zip codes 10032 and 10033 as obtained from the American Community Survey 5-year estimates and n is the number of Hispanics in the WHCS sample. This procedure is implemented $B = 200$ times to produce 200 synthetic FPBB populations.

4.3 Results

Depression data were available for 1,460 Hispanic residents. Table 4.1 shows the distribution of Hispanic residents by national background in Washington Heights, New York. Weighted estimates show that about three-quarters of the Hispanic/Latino residents of Washington Heights are from the Dominican Republic with the remaining 25% comprising of Hispanic/Latinos of South American and Mexican descent. About 85% of respondents report being 18 to 64 years of age.

Figure 4.1 shows the distribution of the survey weights in the WCHS study. The maximum weight in the sample was 229.8. Figure 4.2 shows the bivariate association between the outcomes of interest and survey weights in the 2015 WHCS survey data. Here we see non-linear association between the logit of having current depression and the logit of having undiagnosed depression and survey weight. Using the proposed modeling approach to survey estimation can provide improved efficiency in the presence of highly variable weights as well as being robust to potential misspecification of the model for the outcome of interest.

Table 4.2 shows the prevalence estimates for current depression and undiagnosed depression for Hispanic residents 18 to 64 years and 65 years and older based on the model-based approach. For Hispanic residents aged 18 to 64 years, the prevalence of

current depression is 11.5%. Puerto Rican ethnicity Hispanics had the highest prevalence of current depression and Mexicans the lowest, at 15.0% and 8.4%, respectively. About one in thirteen Hispanics aged 18 to 64 years in Washington Heights have undiagnosed depression. Among Hispanics ages 18 to 64 years, undiagnosed depression is highest among Puerto Ricans; among older adults, Hispanics from the Dominican Republic had the highest rates of undiagnosed depression.

Table 4.3 provides the weighted estimates for current depression and undiagnosed depression. While the ranking of the sub-ethnicity prevalence estimates do not differ, there are some differences in the magnitude between weighting and modeling approaches. Importantly, the standard error associated with the weighted estimate of overall depression in Hispanics in the 18 to 64 years age group is smaller than the standard error for the model-based estimate of overall depression while for those 65 years and older, the standard error associated with the model-based estimate is marginally smaller than that of the weighted estimate. This is not surprising since this survey was designed to produce estimates at aggregated levels as has sufficient sample to provide weighted estimates for all Hispanics that have relatively small standard errors. Examining the Hispanic origin-specific prevalence estimates of current depression, the standard errors associated with the model-based approach are consistently smaller than the weighted survey estimate standard errors. Also of note is that while the survey weighted estimates of current depression among Mexican and European/Other Hispanics 65 years and older are 0.0, the model-based estimates are 14.7% and 13.1%, respectively. This highlights an advantage of using a modeling approach that provide predictions for zero sample population groups. Similar observations can be made with regard to the weighted and model-based estimates for undiagnosed depression.

4.4 Conclusion

In this study, we provide prevalence estimates for current depression and undiagnosed depression in Hispanic residents of a low-income neighborhood in New York City. Further, we compare results from weighted and model-based methods for survey estimation in this study and show the advantages of utilizing flexible regression models in small domain survey estimation. An important aspect of public health research is identifying and addressing health disparities. In order to identify differences in important health indicators between demographic subgroups or domains, there must be appropriate data at the domain level to produce prevalence estimates in these domains. Survey weighted prevalence estimates for subgroups can yield unacceptably large standard errors in the presence of small sample sizes and cannot accommodate prediction in subgroups with no samples [7, 47, 52, 53, 77]. An alternative approach is to utilize regression models for survey estimation. Using regression models for small domain estimation can improve the accuracy of estimation and reduce standard errors [7, 30, 32, 35, 47, 52, 53, 77]. However, correctly specifying the model is essential as misspecified model-based estimates are subject to bias.

Ethnicity-specific prevalence estimates of current depression from this community-based sample were lower than the CES-D based depression prevalence reported in the Hispanic Community Health Study/Study of Latinos [60]. Possible explanations for this finding may lie with the diagnostic tools as well as the samples. Evaluations of both the PHQ-9 and CES-D instruments in patient populations concluded that they are valid screening tools with similar performance [78–80]. Therefore, the differences in prevalence could be attributable to the different populations. The ranking of Hispanic/Latino sub-ethnicities with regard to depression prevalence found in this study was

consistent with previous work [60, 63, 64]. Our study findings advance the epidemiologic literature by providing rates of undiagnosed depression among Hispanics and Hispanic ethnicity subgroups in a community-based sample. Additionally, the fact that over 60% of residents who screened as having current depression had never been diagnosed with depression should highlight the continued need for provider screening and understanding of cultural differences in patient care seeking. Research has shown that while some patients are able to recognize symptoms and seek care, many are unwilling or unable to bring these concerns to the attention of medical practitioners for reasons related to lack of knowledge and culture [62, 81, 82].

While this is an important study, it is critical to note that the target population for the 2015 WHCS is unique with higher proportions of foreign-born and low-income residents than Manhattan and New York City [69] and as such, the results may not be fully generalizable to other Hispanic populations. Please note that, while innovative with respect to content, the survey was cross-sectional in design. Notwithstanding these limitations, the findings from this study have important implications as it relates to the continued promotion of depression screening particularly among low-income minority groups in an effort to improve health outcomes and social well-being.

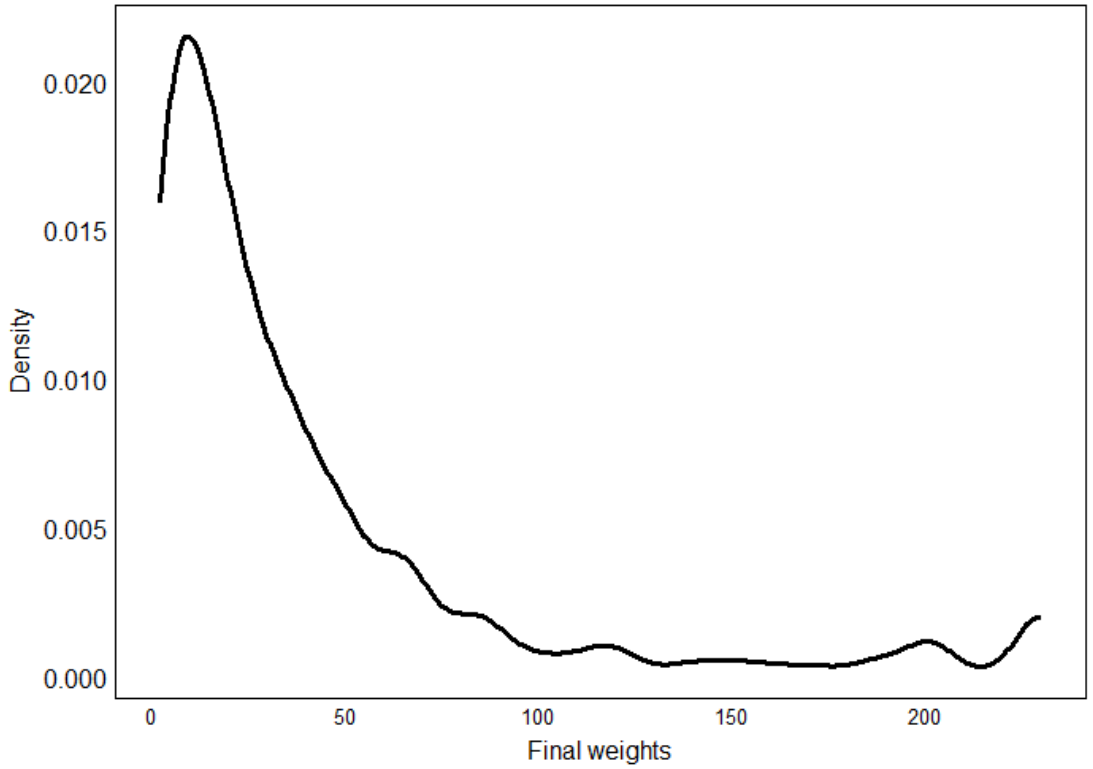

FIGURE 4.1: Distribution of final survey weights, Washington Heights Community Survey (WHCS), 2015.

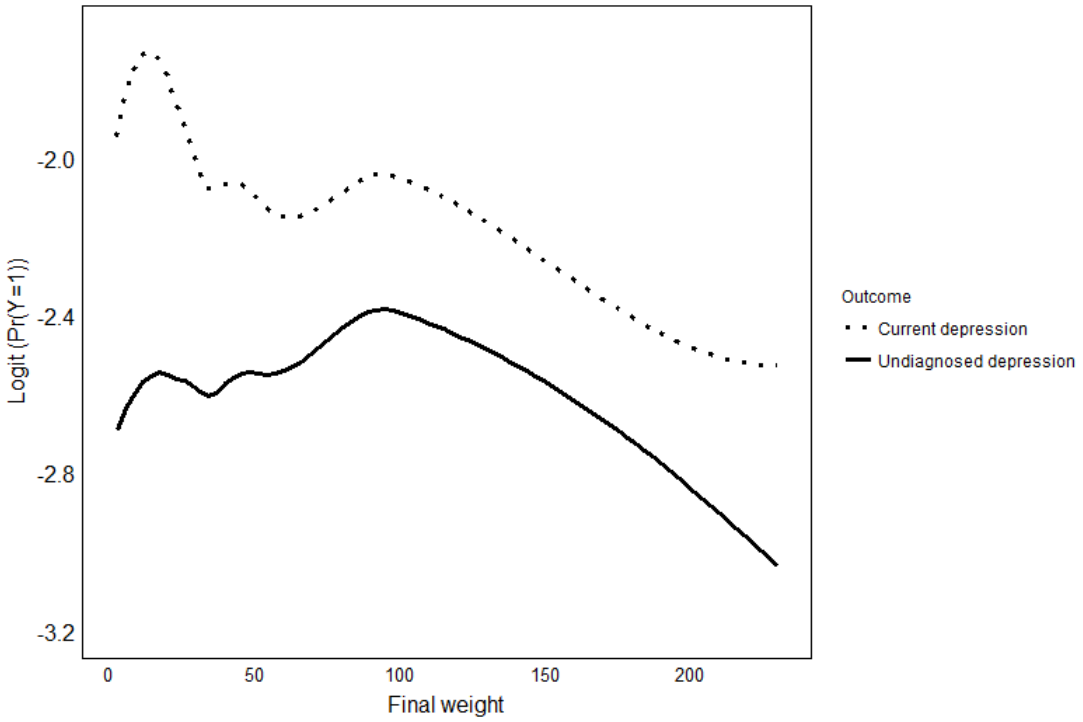

FIGURE 4.2: Bivariate association between probability of having current depression and undiagnosed depression and final survey weights, Washington Heights Community Survey (WHCS), 2015.

TABLE 4.1: Distribution of Hispanic residents by national background in population and survey sample, Washington Heights Community Survey, 2015.

Hispanic origin	Weighted sample, n(%)
All	1,460(100.0)
Dominican	1,081(72.2)
Puerto Rican	113(4.3)
Mexican	32(6.7)
Central American	47(2.3)
South American	103(9.6)
Cuban	49(2.2)
European/other	35(2.6)

TABLE 4.2: Current and undiagnosed depression prevalence estimates by Hispanic sub-ethnicities using model-based estimation ($n = 1,460$), Washington Heights Community Survey, 2015.

Hispanic origin	Current depression, %(SE)	Undiagnosed depression, %(SE)
All	11.7(1.2)	7.4(1.0)
Dominican	11.9(1.2)	7.7(1.0)
Puerto Rican	14.0(2.5)	7.5(1.6)
Mexican	8.5(2.8)	5.8(2.2)
Central American	11.6(2.7)	7.4(2.3)
South American	11.3(2.3)	7.5(1.9)
Cuban	15.3(3.7)	6.9(1.9)
European/other	10.8(2.8)	7.2(2.1)

Current depression is defined as having a PHQ-9 score ≥ 10

TABLE 4.3: Current and undiagnosed depression prevalence estimates by Hispanic sub-ethnicities using survey weighting ($n = 1,460$), Washington Heights Community Survey, 2015.

Hispanic origin	Current depression, %(SE)	Undiagnosed depression, %(SE)
All	11.5(1.2)	7.5(1.0)
Dominican	11.8(1.4)	8.2(1.3)
Puerto Rican	20.4(7.9)	7.5(3.3)
Mexican	1.2(1.2)	1.2(1.2)
Central American	10.8(5.9)	8.6(5.4)
South American	13.0(5.1)	8.6(3.8)
Cuban	14.6(6.9)	0.9(0.7)
European/other	5.7(4.5)	5.1(4.4)

Current depression is defined as having a PHQ-9 score ≥ 10

Chapter 5

Conclusion

In this thesis, we described an important area of research in mental health that motivated our work. Specifically, researchers in the OANG MHI study were interested in improving the well-being of reserve soldiers during and after deployment by examining the prevalence and risk factors of a number of mental health related outcomes using survey data from a sample of active guards. However, the survey data was challenged by non-response and sampling frame under-coverage which can lead to biased inference. Individual level auxiliary variable information for the OANG target population can be used to improve inference in the presence of non-response and under-coverage. Available auxiliary information included demographic measures such as gender, race, age, and rank as well as a continuous measure, number of years in service. This continuous measure can be discretized and used to account for non-response via post-stratification and raking weighting or can be incorporated into a predictive modeling approach similar to the basic post-stratification model or multilevel regression for post-stratification. However, examination of the association between years of service and probability of responding to the survey as well as outcomes of interest, showed strong non-linear associations. As

such, categorizing years of service may result in loss of important information. Response propensity weighting estimation can accommodate continuous auxiliary information in a model for response, as can prediction models for the outcome. However, an important consideration with using regression models for either the response propensity weighting or prediction modeling approaches to survey inference, is how best to specify the model to avoid misspecification. Models induce subjectivity and misspecified models can yield poor inferences.

The aim of this work was to develop a robust predictive modeling approach to survey inference about a population proportion in the presence of non-response and sampling frame under-coverage. In our first project, we proposed a flexible modeling framework for survey inference that is robust to model misspecification by using a penalized spline of the survey weights to model the association between the auxiliary variables and survey outcome. Our second project extended the proposed modeling approach to include factor-by-curve interaction with the aim of improving survey inference in sub-populations or domains defined by categories of the auxiliary information. Being able to assess prevalence in sub-groups defined by sociodemographic characteristics is an important part of epidemiologic research. We find that both our proposed model and the extension to the stratified model, are more efficient than the existing weighting and modeling approaches with closer to nominal coverage. Further, we find improved performance of the stratified penalized spline model for inference in small size domains. We apply both methods to 2008-2009 data from the OANG MHI study which aims at improving the well-being of reserve soldiers during and after deployment.

The third project extends the proposed model to the case where we have limited auxiliary information for the non-sampled population. In this application, we are

interested in estimating current and undiagnosed depression in Hispanic and Latino sub-ethnicities defined by national origin. We utilize data from the 2015 WHCS, a survey of residents of Washington Heights, New York City, that aims to improve targeted public health and clinical interventions in this neighborhood by describing the health and health needs of residents. In this study, limited auxiliary information is available in the form of marginal counts of domain size in the population as well as final survey weights that account for sampling strategy as well as non-sampling errors. In order to apply our proposed modeling approach to the 2015 WHCS data, we implemented a weighted finite population Bayesian bootstrap (FPBB) method to generate synthetic populations for prediction. We find that the modeling approach incorporating the FPBB procedure resulted in domain estimates with smaller standard errors than the survey weighted estimate.

In summary, this dissertation proposed a Bayesian penalized spline prediction modeling framework for survey inference challenged by non-response and under-coverage. The proposed approach reduces inefficiency of inference due to highly dispersed weights as compared to weighted estimators, and yields more robust inference when there is model misspecification as compared to prediction modeling approaches.

5.1 Implications to Public Health

Survey data play an important role in epidemiologic research, facilitating estimation of and inference about health indicators in large finite populations using moderately sized samples. However, surveys are increasingly challenged by a growing problem of non-response and sampling frame under-coverage. Both of these issues can lead to biased inference. In order to address these potential biases and improve the accuracy of survey

estimates, appropriate statistical methods need to be applied. Although traditional weighted estimators work well at correcting bias due to non-response and under-coverage, they can be very inefficient in the presence of highly variable survey weights leading to unstable estimates. Further, prediction models typically used to improved these inefficiencies are subject to model misspecification; this can lead to incorrect inference.

Using our proposed Bayesian penalized spline prediction modeling approach can improve inefficiency in estimation due to highly dispersed weights as compared to weighted estimators. Moreover, our proposed modeling approach yields more robust inference when there is model misspecification as compared to existing weighted and prediction modeling approaches. In our Chapter 2 simulation setting with high variation in weights, gains in reduced RMSE from our proposed method as compared with weighting approaches, ranged from 19% to 48%. In inferential statistics, reducing RMSE, increases the power of the statistical test i.e. we are better able to correctly identify important associations. Further, in research practice, where there is no definitive way to determine whether the specified model for the outcome is correct, utilizing our proposed method which is more robust to model misspecification than existing modeling approaches, is advantageous. In fact, in simulation study, this is where our proposed Bayesian penalized spline model offered greatest improvement with respect to decreased bias and RMSE, as well as improved confidence coverage relative to other approaches.

Quantifying and identifying disparities in health-related outcomes such as burden of disease and health behaviors, is an important aspect of public health research. Improving inference in sub-populations that define geographic or demographic domains of the target population can aid in correctly identifying meaningful differences between groups. Chapter 3 of this thesis focused on domain estimation for survey data with particular attention to estimation in domains with small size relative to population size,

where weighted estimation is often inappropriate and issues relating to model misspecification can become magnified. In our simulation study, we see gains of up to 16% reduction in RMSE for estimation in small size domains comparing the proposed modeling approach with weighting estimation. These results are promising and necessitate further work to assess the performance of the stratified penalized spline approach as domain size and prevalence varies in order to fully define the method's utility.

To implement the Bayesian penalized spline prediction modeling approach proposed here, auxiliary information for the joint distribution of the auxiliary variables is needed. Oftentimes, we have limited information on the non-sampled units. In our last project, we utilized a FPBB procedure to complete survey inference in sub-populations using our proposed modeling approach. Here, the breadth of utility of the proposed Bayesian predictive modeling framework is exemplified in its application to survey inference where limited auxiliary information is available for non-sampled population units via final survey weights and marginal counts of population size in domains of interest. While this work is developmental and requires further assessment via simulation study, it holds much promise for expanding the application of the proposed modeling approach. Many surveys provide minimal auxiliary information on non-sampled units which necessarily prohibits use of the predictive modeling approach as initially proposed. Being able to utilize a bootstrap procedure that can incorporate important survey features such as design and non-response adjustments to facilitate the use of prediction models, will extend the utility of this thesis research.

When interpreting the results of this work, it is important to note that we focused on point estimation of the population proportion in a simple setting which involved non-response and sampling frame under-coverage issues in a survey with a moderate amount of auxiliary data. We further extend to domain estimation and the case where

limited auxiliary information is available on non-sampled population units. There is a need for further work that considers the application of these methods to other types of important outcomes such as count data, to survey settings that involve a large number of auxiliary variables that result in challenges common to high-dimensional data, as well as to surveys with complex design features.

Notwithstanding, there is much promise in applying the proposed Bayesian prediction modeling framework to survey inference when the data are challenged by nonresponse and under-coverage, and particularly when there exists highly variable weights or complex population associations. Not only do results from the proposed model parallel those from weighting approaches in large samples, they can outperform weighting approaches if the model is correctly specified. Moreover, the hierarchical Bayes approach is attractive because of its ability to yield better inferences for small-sample problems.

Appendix A

Bayesian Multilevel Penalized Spline Stan model

```
data {
int<lower=1> n; //number of individuals
int<lower=1> P1; //number of linear predictors & intercept
int<lower=1> P2; //number of varying intercepts
int<lower=1> numknots; //fixed number of knots
int<lower=0,upper=1> Y[n]; //outcome variable
matrix[n,P1] x; //linear predictors & intercept
matrix[n,P2] z; //random intercepts
matrix[n,numknots] spn; // spline terms
}

parameters {
vector[P1] beta;
vector[P2] b;
real b0;
real<lower=0.001> tau;
vector[numknots] bk;
real<lower=0.001> taubk;
}

model {
Y ~ bernoullilogit( x*beta + z*b + spn*bk);
for (k in 1:P2) {
b[k] ~ normal(b0,tau);
}
tau ~ cauchy(0,3);
for (k in 1:numknots)
bk[k] ~ normal(0,taubk);
}
taubk ~ cauchy(0,3);
}
```

Appendix B

Stratified Bayesian Multilevel Penalized Spline Stan model

```
data {
int<lower=1> n; //number of individuals
int<lower=1> P1; //number of linear predictors
int<lower=1> P2; //number of varying intercepts
int<lower=1> L1; //number of dummy variables in interaction variable L-1
int<lower=1> L2; //number of categories in interaction variable L
int<lower=0,upper=1> Y[n]; //outcome variable
rowvector[P1] x[n]; //linear predictors
rowvector[P2] z[n]; //random intercepts
int<lower=0> numknots; //fixed number of knots
rowvector[numknots] spn[n]; //spline terms
rowvector[L1] z1[n]; //linear predictors for interaction variable
rowvector[L2] z2[n]; //linear predictors for interaction variable
}

parameters {
vector[P1] beta;
vector[P2] b;
real b0;
real<lower=0.0001> tau;
vector[numknots] bk;
real<lower=0.0001> taubk;
vector[L1] gamma;
matrix[numknots,L2] ck;
vector<lower=0.0001>[L2] tauck;
}

model {
for (i in 1:n) {
Y[i] ~ bernoullilogit( x[i]*beta + z[i]*b + spn[i]*bk +
z1[i]*gamma + sum( z2[i] .*(spn[i]*ck) ) );
}
for (k in 1:P2) {
```

```
b[k] ~ normal(b0,tau);
}
for (l in 1:numknots) {
bk[l] ~ normal(0,taubk);
for (l2 in 1:L2) {
ck[l,l2] ~ normal(0,tauck[l2]);
tauck[l2] ~ cauchy(0,3);
}
}
tau ~ cauchy(0,3);
taubk ~ cauchy(0,3);
}
```

Bibliography

[1] S.L. Lohr. *Sampling: Design and Analysis*. Pacific Grove: Brooks/Cole Publishing, 1999.

[2] E. L. Korn and B. I. Graubard. *Analysis of Health Surveys*. New York: Wiley, 1999.

[3] J. M. Brick and G. Kalton. Handling missing data in survey research. *Statistical Methods in Medical Research*, 5:215–238, 1996.

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

[5] A. Gelman and J. B. Carlin. Poststratification and weighting adjustments. In R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, editors, *Survey Nonresponse*, pages 289–302. Wiley, New York, 2002.

[6] M. R. Elliott and R. J. A. Little. Model-based alternatives to trimming survey weights. *Journal of Official Statistics*, 16(3):191–209, 2000.

[7] R.J.A. Little. Post-stratification: a modeler's perspective. *Journal of the American Statistical Association*, 88:1001–1012, 1993.

[8] T.G. Gregoire. Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28:1429–1447, 1998.

[9] W.G. Cochran. *Sampling Techniques*. New York: Wiley, 1977.

[10] R.J.A. Little. To model or not to model? competing modes of inference for finite population sampling. *Journal of the American Statistical Association*, 99(466):546–556, 2004.

[11] R.J.A. Little. Survey sampling: past controversies, current orthodoxy, and future paradigms. In X. Lin, D. L. Banks, C. Genest, G. Molenberghs, D.W. Scott, and J.-L. Wang, editors, *Past, Present and Future of Statistical Science*, COPSS 50th Anniversary Volume, pages 413–425. CRC Press, Florida, 2014.

[12] Swensson B. Wretman J. Särndal, C-E. *Model Assisted Survey Sampling*. New York: Springer-Verlag, 1992.

[13] C.-E. Särndal. Design-based and model-based inference in survey sampling. *Scandanavian Journal of Statistics*, 5:27–52, 1978.

[14] J.C. Deville, C.E. Sarndal, and O. Sautory. Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88(423):1013–1020, 1993.

[15] L. C. Lazzeroni and R. J. A. Little. Random-effects models for smoothing post-stratification weights. *Journal of Official Statistics*, 14(1):61–78, 1998.

[16] R. Lehtonen and A. Veijanen. Logistic generalized regression estimators. *Survey Methodology*, 24(1):51–55, 1998.

[17] R.M. Royall. The model based (prediction) approach to finite population sampling theory. In M. Ghosh and P.K. Pathak, editors, *Current Issues in Statistical Inference: Essays in honor of D Basu*, pages 225–240. Institute of Mathematical Statistics, Hayward, 1992.

[18] J.N.K. Rao. Impact of frequentist and bayesian method on survey sampling practice: a selective appraisal. *Statistical Science*, 26(2):240–256, 2011.

[19] G.H. Cohen, D.S. Fink, L. Sampson, and S. Galea. Mental health among reserve component military service members and veterans. *Epidemiologic Reviews*, 37:7–22, 2015.

[20] C.S. Fullerton, R.J. Ursano, and L. Wang. Acute stress disorder, posttraumatic stress disorder, and depression in disaster or rescue workers. *American Journal of Psychiatry*, 161:1370–1376, 2004.

[21] J. Griffith. Decades of transition for us reserves: changing demands on reserve identity and mental well-being. *International Review of Psychiatry*, 23:181–191, 2011.

[22] J.L. Thomas, J.E. Wilk, L.A. Riviere, D. McGurk, C.A. Castro, and C.W. Hoge. Prevalence of mental health problems and functional impairment among active component and national guard soldiers 3 and 12 months following combat in iraq. *Archives of General Psychiatry*, 67(6):614–623, 2010.

[23] L.W. Castaneda, M.C. Harrell, D.M. Varda, K.C. Hall, M.K. Beckett, and S. Stern. Deployment experiences of guard and reserve families: implications for support and retention. *RAND, National Defense Research Institute*, 2008.

[24] J. Stuhltrager. Send in the guard: the national guard response to natural disasters. *Natural Resources and Environment*, 20(4):21–26, 2006.

[25] E. Goldmann, J.R. Calabrese, M.R. Prescott, M. Tamburrino, I. Liberzon, R. Slembarsk, E. Shirley, T. Fine, T. Goto, K. Wilson, et al. Potentially modifiable pre-,

peri-, and post-deployment characteristics associated with deployment-related post-traumatic stress disorder among ohio army national guard soldiers. *Annals of Epidemiologys*, 22:71–78, 2011.

[26] M. B. Tamburrino, P. Chan, M. Prescott, J. Calabrese, I. Liberzon, R. Slembarski, E. Shirley, T. Fine, T. Goto, et al. Baseline prevalence of axis i diagnosis in the ohio army national guard. *Psychiatry Research*, 226:142–148, 2015.

[27] J. Calabrese, M. Prescott, M. B. Tamburrino, I. Liberzon, R. Slembarski, E. Goldmann, E. Shirley, T. Fine, T. Goto, K. Wilson, et al. Ptsd comorbidity and suicidal ideation associated with ptsd within the ohio army national guard. *Journal of Clinical Psychiatry*, 72(8):1072–1078, 2011.

[28] D. Holt and T.M.F. Smith. Post stratification. *Journal of the Royal Statistical Society. Series A*, 142(1):33–46, 1979.

[29] W. E. Deming and F. F. Stephan. On a least squares adjustment of a sampled frequency table when the expected margin totals are known. *The Annals of Mathematics and Statistics*, 11:427–444, 1940.

[30] A. Gelman and T.C. Little. Postratification into many categories using hierarchical logistic regression. *Survey Methodology*, 23(2):127–135, March 1998.

[31] D. K. Park, A. Gelman, and J. Bafumi. Bayesian multilevel estimation with post-stratification: state-level estimates from national polls. *Political Analysis*, 12:375–385, 2004.

[32] Q. Chen, M.R. Elliott, and R.J.A. Little. Basyesian penalized spline model-based inference for finite population proportion in unequal probability sampling. *Survey Methodology*, 36:23–34, 2010.

[33] G. Zhang and R. Little. Extensions of the penalized spline of propensity prediction method of imputation. *Biometrics*, 65:911–918, 2009.

[34] R. Little and H. An. Robust likelihood-based analysis of multivariate data with missing values. *Statistica Sinica*, 14:949–968, 2004.

[35] H. Zheng and R.J.A. Little. Penalized spline model-based estimation of the finite populations total from probability-proportional-to-size samples. *Journal of Official Statistics*, 19(2):99–117, 2003.

[36] P. H. C. Eilers and B. D. Marx. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2):89–121, 1996.

[37] P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[38] B. K. Lee, J. Lessler, and E. A. Stuart. Weight trimming and propensity score weighting. *PLoS ONE*, 6(3):e18174 doi:10.1371/journal.pone.0018174, 2011.

[39] M. Schonlau, A. van Soest, A. Kapteyn, and M. Couper. Selection bias in web surveys and the use of propensity scores. *Sociological Methods & Research*, 37(3): 291–318, 2009.

[40] A. Matei and Y. Tille. Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size. *Journal of Official Statistics*, 21(4):543–570, 2005.

[41] D. Haziza, F. Mecatti, and J.N.K. Rao. Evaluation of some approximate variance estimators under the rao-sampford unequal probability sampling design. *International Journal of Statistics*, 66(1):91–108, 2008.

[42] The Stan Development Team. Rstan: the r interface to stan, version 2.5.0. http://mc-stan.org/rstan.html, 2014.

[43] M.D. Hoffman and A. Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 25: 1351–1381, 2013.

[44] R.M. Neal. Mcmc using hamiltonian dynamics. In S. Brooks, A. Gelman, G.L. Jones, and X-L. Meng, editors, *Handbook of Markov Chain Monte Carlo*, pages 113–162. Chapman and Hall/CRC Press, 2011.

[45] N. Metropolis, A. Rosenbluth, M. Rosenbluth, M. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21: 1087–1092, 1953.

[46] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

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

[48] C.M. Schnaubelt, R.S. Cohen, M. Dunigan, G. Gentile, J.L. Hastings, J. Klimas, J.P. Marquis, A.G. Schaefer, B. Triezenberg, and M.D. Ziegler. Sustaining the army's reserve components as an operational force. *RAND, National Defense Research Institute*, 2017.

[49] DOD. 2015 demographics: profile of the military community. *Department of Defense, Office of the Deputy Assistant Secretary of Defense for Military Community and Family Policy*, 2015.

[50] I.G. Jacobson, M.A. Ryan, T.I. Hopper, T.C. Smith, P.J. Amoroso, E.J. Boyko, G.D. Gacksetter, T.S. Wells, and N.S. Bell. Alcohol use and alcohol-related problems before and after military combat deployment. *Journal of the American Medical Association*, 300:663–675, 2008.

[51] J.N.K. Rao and I. Molina. *Small Area Estimation*. Hoboken: John Wiley and Sons, Inc, 2 edition, 2015.

[52] D. Pfeffermann. New important developments in small area estimation. *Statistical Science*, 28(1):40–68, 2013.

[53] M. Ghosh and J. N. K. Rao. Small area estimation: an appraisal. *Statistical Science*, 9(1):55–76, 1994.

[54] B.A. Coull, D. Ruppert, and M.P. Wand. Simple incorporation of interactions into additive models. *Biometrics*, 57(2):539–545, 2001.

[55] E. López, A. Steiner, K. Manier, B. Vanle, J. Parisi, T. Dang, T. Chang, S. Ganjian, J. Mirocha, I. Danovitch, and W. Ishak. Quality of life and functioning of hispanic patients with major depressive disorder before and after treatment. *Journal of Affective Disorders*, 225:117–122, 2018.

[56] S.Z. Williams, G. Chung, and P. Muennig. Undiagnosed depression: a community diagnosis. *SSM Population Health*, 3:633–638, 2017.

[57] L. Egede, K. Bishu, R. Walker, and C. Dismuke. Impact of diagnosed depression on healthcare costs in adults with and without diabetes: United states, 2004-2011. *Journal of Affective Disorders*, 195:119–126, 2016.

[58] D. Chisholm, K. Sweeny, P. Sheehan, B. Rasmussen, F. Smith, P. Cuijpers, and S. Saxena. Scaling-up treatment of depression and anxiety: a global return on investment analysis. *Lancet Psychiatry*, 3(5):415–424, 2016.

[59] M. Jackson-Triche, J. Sullivan, W. Wells, K. andRogers, P. Camp, and R. Mazel. Depression and health-related quality of life in ethnic minorities seeking care in general medical settings. *Journal of Affective Disorders*, 58:89–97, 2000.

[60] S. Wassertheil-Smoller, E. Arredondo, J. Cai, S. Castaneda, J. Choca, L. Gallo, M. Jung, Lee-Rey E. LaVange, L., T. Mosley Jr, F. Penedo, D. Santistaban, and P. Zee. Depression, anxiety, antidepressant use, and cardiovascular disease among hispanic mean and women of different national backgrounds: results from the hispanic community health study/study of latinos. *Annals of Epidemiology*, 24:822–830, 2014.

[61] M. Alegria, N. Mulvaney-Day, M. Torres, A. Polo, Z. Cao, and G. Canino. Prevalence of psychiatric disorders across latino subgroups in the united states. *Amercan Journal of Public Health*, 97(1):68–75, 2007.

[62] R. Lewis-Fernández, A. Das, C. Alfonso, M. Weissman, and M. Olfson. Depression in us hispanics: diagnostic and management considerations in family practice. *The Journal of the American Board of Family Practice*, 18(4):282–396, 2005.

[63] H. González, W. Vega, D. Williams, W. Tarraf, B.T. West, and H. Neighbors. Depression care in the united states. *Archives of General Psychiatry*, 67(1):37–46, 2010.

[64] Council on Scientific Affairs. Hispanic health in the united states. *Journal of the American Medical Association*, 265:248–252, 1991.

[65] M. Olfson, C. Blanco, and S. Marcus. Treatment of adult depression in the united states. *JAMA Internal Medicine*, 176(10):1482–1491, 2016.

[66] D. March, J. Luchsinger, J. Teresi, J. Eimicke, S. Findley, O. Carrasquillo, and W. Palmas. High rates of depressive symptoms in low income urban hispanics of caribbean origin with poorly controlled diabetes: correlates and risk factors. *Journal of health Care for the Poor and Undeserved*, 25(1):321–331, 2014.

[67] S. Sahai-Srivastava and L. Zheng. Undiagnosed depression and its correlates in a predominantly hispanic neurology clinic. *Clinical Neurology and Neurosurgery*, 113: 623–625, 2011.

[68] C. Li, E. Ford, G. Zhao, I. Ahluwalia, W. Pearson, and A. Mokdad. Prevalence and correlates of undiagnosed depression among u.s. adults with diabetes: the behavioral risk factor surveillance system, 2006. *Diabetes Research and Clinical Practice*, 83:268–279, 2009.

[69] S.Z. Williams, E. Austin, Z. Rosen, Q. Chen, E. Siegel, D. Hernandez, and P. Muennig. Washington heights: a community diagnosis, 2016.

[70] K. Kroenke and R. Spitzer. The phq-9: a new depression disgnostic and severity measure. *Psychiatric Annals*, 32:509–515, 2002.

[71] K. Kroenke, R. Spitzer, and J. Williams. The phq-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16:606–613, 2001.

[72] L. Manea, S. Gilbody, and D. McMillan. Optimal cut-off score for diagnosing depression with the patient health questionairre (phq-9): a meta-analysis. *The Canadian Medical Association Journal*, 184(3):E191–E196, 2012.

[73] Q. Dong, M. Elliott, and T. Raghunathan. A nonparametric method to generate synthetic populations to adjust for complex sampling design features. *Survey Methodology*, 40(1):29–46, 2014.

[74] The bayesian bootstrap and multiple imputation for unequal probability sample designs. *JSM Proceedings, Survey Research Methods Section (Anaheim, CA)*, 1997.

[75] A.Y. Lo. A bayesian bootstrap for a finite population. *Annals of Statistics*, 16(4): 1684–1695, 1988.

[76] G. Meeden, R. Lazar, and C. Geyer. polyapost: Simulating from the polya posterior, 2017.

[77] D. Pfeffermann. Small area estimation: new developments and directions. *International Statistical Review*, 70(1):125–143, 2002.

[78] D. Amtmann, J. Kim, H. Chung, Askew R. Bamer, A., S. Wu, K. Cook, and K. Johnson. Comparing cesd-10, phq-9, and promis depression instruments in individuals with multiple sclerosis. *Rehabilitation Psychology*, 59(2):220–229, 2014.

[79] J. Williams, E. Hirsch, K. Anderson, A. Bush, S. Goldsten, S. Grill, S. Lehmann, J. Little, R. Margolis, J. Palanci, G. Pontone, H. Weiss, P. Rabins, and L. Marsh. A comparison of nine scales to detect depression in parkinson disease: which scale to use? *Neurology*, 78:998–1006, 2012.

[80] K. Milette, M. Hudson, M. Baron, B. Thombs, and Canadian Scleroderma Research Group. Comparison of the phq-9 and ces-d depression scales in systemic

sclerosis: internal consistency reliability, convergent validity and clinical correlates. *Rheumatology*, 49:789–796, 2010.

[81] R. Epstein, P. Duberstein, M. Feldman, A. Rochlen, R. Bell, R. Kravtiz, Becker J. Cipri, C., P. Bamonti, and D. Paterniti. "i didn't know what was wrong:" how people with undiagnosed depression recognize, name, and explain their distress. *Journal of General Internal Medicine*, 25(9):954–961, 2010.

[82] L. Goldman, N. Nielsen, and H. Champion. Awareness, diagnosis, and treatment of depression. *Journal of General Internal Medicine*, 14:569–580, 1999.