Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Bayesian predictive inference of small area proportions under selection bias

by Seongmi Choi, Balgobin Nandram and Dalho Kim

Release date: June 24, 2021

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2021
Vol. 47, No. 1, pp. 91-122
Statistics Canada, Catalogue No. 12-001-X

Bayesian predictive inference of small area proportions under selection bias

Seongmi Choi, Balgobin Nandram and Dalho Kim[1]

Abstract

In a previous paper, we developed a model to make inference about small area proportions under selection bias in which the binary responses and the selection probabilities are correlated. This is the homogeneous nonignorable selection model; nonignorable selection means that the selection probabilities and the binary responses are correlated. The homogeneous nonignorable selection model was shown to perform better than a baseline ignorable selection model. However, one limitation of the homogeneous nonignorable selection model is that the distributions of the selection probabilities are assumed to be identical across areas. Therefore, we introduce a more general model, the heterogeneous nonignorable selection model, in which the selection probabilities are not identically distributed over areas. We used Markov chain Monte Carlo methods to fit the three models. We illustrate our methodology and compare our models using an example on severe activity limitation of the U.S. National Health Interview Survey. We also perform a simulation study to demonstrate that our heterogeneous nonignorable selection model is needed when there is moderate to strong selection bias.

Key Words: Biserial correlation; Metropolis-Hastings algorithm; Nonignorable selection model; Official statistics; Selection probabilities.

1. Introduction

In many complex sample surveys, individuals are sampled with differential selection probabilities. For binary responses, if the proportion of positive responses among the sampled individuals differs substantially from those among the nonsampled individuals, there is a selection bias. In some cases a selection bias is obtained by design (e.g., probability proportional to size (PPS) sampling) and we can take care of the selection bias, but in other problems, this is not the case. For example, in non-probability samples the selection probabilities are unknown. In our problem, we assume that the survey weights or selection probabilities can help us to understand the selection bias. We want to make inference about the finite population proportions of the small areas when a possibly biased sample is available from each area. Choi, Nandram and Kim (2017), henceforth CNK, extended the model of Nandram, Bhatta, Bhadra and Shen (2013), henceforth NBBS, who studied a single area, to accommodate inference about small areas. While the CNK model assumes that distribution of the selection probabilities is the same across areas, our new contribution is to assume that the distributions of the selection probabilities over areas are different, but they share an effect.

There are two types of models that can be considered when making inference about small areas. First, we can use an ignorable selection model in which the response variable is not related to the selection probabilities. Such a model will not adjust for the selection bias, and will produce biased estimates if there

1. Seongmi Choi, Health Insurance Policy Research Istitute, National Health Insurance Service, Wonju, Gangwon 26464, South Korea. E-mail: edisil1202@gmail.com; Balgobin Nandram, Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, USA. E-mail: balnan@wpi.edu; Dalho Kim, Department of Statistics, Kyungpook National University, Daegu 41566, South Korea. E-mail: dalkim@knu.ac.kr.

are no other covariates to act like the selection probabilities. We assume that the only available information is the sampled responses, sampled selection probabilities and area identifiers. Second, in a nonignorable selection model, the response variables are related to the selection probabilities. In our study on binary responses, the distribution of the selection probabilities for the "yes" responses is different from that for the "no" responses, thereby making the response variables and selection probabilities correlated. Official statistics are obtained from many complex surveys, and these surveys are conveniently designed with selection bias (PPS sampling is usually a part of a complex survey design). This study is important in the construction of official statistics from complex surveys.

It is pertinent to give a brief discussion on recent developments. For continuous responses, Pfeffermann (1988), Pfeffermann (1993), Sverchkov and Pfeffermann (2004) and Pfeffermann, Krieger and Rinott (1998) specified a relation between the survey weights and the response variable. Chambers, Dorfman and Wang (1998) assumed that the selection probabilities are related to the continuous responses. A related study within the Bayesian paradigm is given by Ma, Sedransk, Nandram and Chen (2018). There is also an application of this method to calculate the total gas consumption in the US using a PPS sample; see Nandram, Choi, Shen and Burgos (2006). Chen, Elliott and Little (2010) used penalized spline to make a Bayesian predictive inference for PPS sampling. However, because these approaches require some information about the nonsampled selection probabilities, they are difficult to use. The nonsampled selection probabilities are not available to secondary data analysts and we continue to assume that these nonsampled selection probabilities are not available in our work. Pfeffermann and Sverchkov (2007) extended the work of Sverchkov and Pfeffermann (2004) to accommodate small areas with informative sampling of areas and within selected areas. Like Chen et al. (2010), Opsomer, Glaeskens, Ranalli, Kauermann and Breidt (2008) used a penalized spline regression to construct a small-area model, which includes the selection probabilities, in a non-parametric manner. While all these works are for continuous response in small areas, we analyze binary data in this paper.

Our approach to selection bias problems has been to adjust the sample part of a population model. That is, a model is constructed for the entire population, the model is then adjusted to accommodate the sample with the selection probabilities, and the superpopulation parameters are then estimated. Once this is done for the sample, prediction is done for the entire finite population. This approach to selection bias was described nicely by Malec, Davis and Cao (1999). NBBS reviewed and provided a full Bayesian analyis to the method of Malec et al. (1999); Nandram (2007) provided a surrogate sampling procedure within the spirit of Malec et al. (1999). There are many extensions to this approach to accommodate small areas. For example, to analyze continuous responses, Nandram and Choi (2010) included the selection probabilities into a full Bayesian nonignorable model in order to analyze continuous body mass index data, which were discretized in an elegant manner. Malec et al. (1999) used a hierarchical Bayesian model to accommodate a selection mechanism for binary data. As stated by NBBS, there are two possible problems with this model, "First, the θ_{uy} are only weakly identified. Second, the parameters θ_{uy} are never known, and in a Bayesian framework these must also be stochastic. In this paper, in a single attempt, we show how to solve these two problems (weak identifiability and stochastic parameters) for a biased sample drawn from a binary population using information from the survey weights (or selection probabilities)”. Using a nonignorable selection model, NBBS showed how to correct for these two problems for an individual-area model.

Recently, there has been some interesting Bayesian activities for PPS sampling with continuous responses. Zangeneh and Little (2015) used a Bayesian bootstrap to model the measures of size (related to the selection probabilities) and a spline regression of the responses conditional on the measures of size. For the same problem, Si, Pillai and Gelman (2015) poststratified the selection probabilities and used a Gaussian process to model the responses. These are good approaches to Bayesian analyzes of PPS sampling and can be applied to our problem. However, our approach is different because we do not model the nonsampled selection probabilities, rather we adjust a population model.

CNK studied a special case for an extension of NBBS model within the small area context. They assumed that the sample selection probabilities have the same support over the set π_u^* over areas, and the distribution of the selection probabilities given the binary response y_{ij} is

$$P\left(\pi_{ij} = \pi_u^* \,\middle|\, \boldsymbol{\theta}, y_{ij} = y\right) = \theta_{uy}, \; u = 1, \ldots, U, \; y = 0, 1, \; j = 1, \ldots, n_i, \; i = 1, \ldots, \ell.$$

This model assumes that the sample selection probabilities have the same support and each area has the same distribution for the selection probabilities; homogeneous nonignorable selection model. However, the homogeneity assumption is strong and might not be true for most of real settings. In practice, the distribution of the sample selection probabilities can be very different across domains due to sampling designs. The distribution of the sample selection probabilities given $\mathbf{y}$ can also vary a lot across domains. In this paper, we consider a model under a heterogeneity assumption, which is that the sample selection probabilities have different supports and different distributions over areas; this is the heterogeneous nonignorable selection model.

CNK used a simulation study to show how different a baseline ignorable selection model and the homogeneous nonignorable selection model can be when there is selection bias. But it is well known that design information needs to be included when the sample is not randomly selected. In this paper, we use a simulation study to assess the performance of our heterogeneous nonignorable selection model. We draw data from the homogeneous nonignorable selection model and heterogeneous nonignorable selection model respectively, and fit the three models. Then we compare the performance of the models using several measures.

In this paper, we consider the problem of making inference about the finite population proportions of the small areas when there is likely to be a different selection bias by areas. Specifically we extend the homogeneous nonignorable selection model of CNK to accommodate selection probabilities that have different supports in different areas. In Section 2, we give a review of the ignorable selection model and the homogeneous nonignorable selection model, which were studied previously. In Section 3, we show

how to adjust a Bayesian ignorable selection model to incorporate the selection bias in a small area framework when the sample selection probabilities have different distributions by areas. We also describe how to perform the computation. In Section 4, we provide results of an illustrative example on severe activity limitation in the 1995 National Health Interview Survey. We also provide a simulation study to assess the performance of the heterogeneous nonignorable selection model. Section 5 has summary and concluding remarks.

2. Review

This section gives a review of the ignorable selection (IS) model discussed by NBBS and the homogeneous nonignorable selection (HoS) model discussed by CNK. We assume that there are ℓ areas and the population size of the i^{th} area is $N_i,\ i = 1, \ldots, \ell,$ which are known. We consider binary responses.

The samples taken from each area can be biased in that the proportion of positive responses among the sampled units may be different from the proportion of positive responses among the nonsampled units. A sample of $n_i << N_i$ is taken from the i^{th} area, and j^{th} unit within the i^{th} area is taken with selection probabilities $\pi_{ij},\ j = 1, \ldots, N_i,\ i = 1, \ldots, \ell.$ As common to many problems of this kind, the selection probabilities are observed only for the sampled values. (Only the sampled selection probabilities are presented to secondary data users.) Design scientists adjust the selection probabilities to take care of nonresponse and other nonsampling errors. But the selection probabilities are a major part of the survey weights and we assume that the selection probabilities are approximately the reciprocal of the survey weights in our application.

Let $y_{i1}, \ldots, y_{iN_i},\ i = 1, \ldots, \ell,$ denote the binary responses in the ℓ areas. A biased sample S_i of size n_i is available from the i^{th} area together with the selection probabilities of the sample. Denote the sampled values by $y_{i1}, \ldots, y_{in_i},\ i = 1, \ldots, \ell,$ and the set of nonsampled values by $\bar{S}_i.$ Inference is required for the finite population proportion for each area. Let $P_i = \sum_{j=1}^{N_i} y_{ij} \big/ N_i$ denote the small area proportion of the i^{th} area and $\hat{p}_i = \sum_{j=1}^{n_i} y_{ij} \big/ n_i$ denote the corresponding sample proportions. Clearly, $\hat{p}_i$ can be a biased estimator of $P_i,$ the proportion of ones in the finite population. In design-based survey analysis, P_i are fixed unknown quantities, but in the Bayesian paradigm, P_i are random variables, which are to be predicted (i.e., Bayesian predictive inference).

We review the IS model in Section 2.1 and the HoS model in Section 2.2.

2.1 Ignorable selection model

A standard ignorable selection model for the binary variables $y_{ij},\ j = 1, \ldots, N_i,\ i = 1, \ldots, \ell,$ is

$$y_{ij} \mid p_i \overset{\text{ind}}{\sim} \text{Bernoulli}(p_i),$$

$$p_i \mid \mu, \tau \overset{\text{iid}}{\sim} \text{Beta}(\mu\tau, (1-\mu)\tau)$$

and

$$p(\mu, \tau) = \frac{1}{(1+\tau)^2}, \tau \geq 0 (0 < \mu < 1)$$

(i.e., a priori μ and τ are independent). Here the prior for τ is proper and noninformative.

For notational convenience, let $\mathbf{p} = (p_1, \ldots, p_\ell)$, $\mathbf{y} = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{\ell 1}, \ldots, y_{\ell n_\ell})$ and $s_i = \sum_{j=1}^{n_i} y_{ij}$. Then Bayesian predictive inference of P_i can be performed based on the following posterior distribution,

$$p(y_{ij} \mid \mathbf{y}) = \int p(y_{ij} \mid p_i, \mathbf{y}) \pi(p_i \mid \mathbf{y}) dp_i = \int p(y_{ij} \mid p_i) \pi(p_i \mid \mathbf{y}) dp_i, \ j = n_i + 1, \ldots, N_i, \ i = 1, \ldots, \ell,$$

where $y_{ij} \mid p$ are independent, $\pi(p_i \mid \mathbf{y}) = \int \pi(p_i \mid \mu, \tau, \mathbf{y}) \pi(\mu, \tau \mid \mathbf{y}) d\mu d\tau$ and $p_i \mid \mu, \tau, \mathbf{y}$ are independent with $p_i \mid \mu, \tau, \mathbf{y} \sim \text{Beta}(s_i + \mu\tau, n_i - s_i + (1-\mu)\tau)$.

2.2 Homogeneous nonignorable selection model

CNK studied a special case for an extension of NBBS model within the small area context. They assumed that the sample selection probabilities $(\pi_{i1}, \ldots, \pi_{in_i})$ have the same support over the set $\pi_u^*, u = 1, \ldots, U$ for $i = 1, \ldots, \ell$.

Letting $\boldsymbol{\theta} = (\theta_{10}, \ldots, \theta_{U0}, \theta_{11}, \ldots, \theta_{U1}) = (\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)$, (say) and $\boldsymbol{\pi} = (\pi_{11}, \ldots, \pi_{1n_1}, \ldots, \pi_{\ell 1}, \ldots, \pi_{\ell n_\ell})$, the distribution of the selection probabilities, given the binary response y_{ij}, is

$$P(\pi_{ij} = \pi_u^* \mid \boldsymbol{\theta}, y_{ij} = y) = \theta_{uy}, \ u = 1, \ldots, U, \ y = 0, 1, \ j = 1, \ldots, n_i, \ i = 1, \ldots, \ell$$

and

$$y_{ij} \mid p_i \overset{\text{iid}}{\sim} \text{Bernoulli}(p_i), \ j = 1, \ldots, n_i, \ i = 1, \ldots, \ell.$$

It is worth noting here that the θ_{uy} are not selection probabilities.

Note that the sample selection probabilities have the same support and distribution by areas. However, this assumption is strong and might not be true for most of real settings. In practice, the distribution of the sample selection probabilities can be very different across domains due to sampling designs.

To proceed, they make a one-to-one transformation from p_i to q_i via

$$q_i = \frac{a_1 p_i}{a_1 p_i + a_0 (1 - p_i)}, \ i = 1, \ldots, \ell.$$

Then, they assumed that $\mathbf{q}$, $\boldsymbol{\theta}_0$, $\boldsymbol{\theta}_1$ are independent with

$$q_i \mid \mu, \tau_1 \overset{\text{iid}}{\sim} \text{Beta}(\mu\tau_1, (1-\mu)\tau_1), \ i = 1, \ldots, \ell,$$

$$\boldsymbol{\theta}_0 \mid \tau_0 \sim \text{Dirichlet}\left(\boldsymbol{\theta}_0^{(0)} \tau_0\right) \quad \text{and} \quad \boldsymbol{\theta}_1 \mid \tau_0 \sim \text{Dirichlet}\left(\boldsymbol{\theta}_1^{(0)} \tau_0\right),$$

where $\boldsymbol{\theta}_0^{(0)}$ and $\boldsymbol{\theta}_1^{(0)}$ are to be specified.

A priori they assumed

$$p(\mu, \tau_0, \tau_1) = \frac{1}{(1+\tau_0)^2} \frac{1}{(1+\tau_1)^2}, \ 0 < \mu < 1, \ \tau_0, \tau_1 \geq 0.$$

Once they drew a sample from the posterior $\pi(\mathbf{q}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \mu, \tau_0, \tau_1 \mid \mathbf{y})$, by retransforming from q_i to p_i,

$$p_i = \frac{a_0 q_i}{a_0 q + a_1 (1 - q_i)},$$

they obtained a sample from the posterior distribution of $\pi(\mathbf{p}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \mu, \tau_0, \tau_1 \mid \mathbf{y})$.

Finally, one can draw the entire finite population values, y_{ij}, from Bernoulli(p_i), $i = 1, \ldots, \ell$, $j = 1, \ldots, n_i$ independently, using estimated $\mathbf{p}$, and can make an inference about the small area proportions. The distribution of the sample selection probabilities given $\mathbf{y}$ can also vary a lot across domains. Therefore, in Section 3, we consider a model which has different supports and distributions for the sample selection probabilities by areas.

3. Heterogeneous nonignorable selection model

In Section 3.1, we describe the heterogeneous nonignorable selection (HeS) model. We also show how to perform the computations in Section 3.2. We show how to fit this model and how to make inference about the small area proportions. Inference about the small area proportions under the HeS model is obtained using surrogate samples (Nandram, 2007).

3.1 Methodology

We assume that the sample selection probabilities $\left(\pi_{i1}, \ldots, \pi_{in_i}\right)$ have the different supports over the set π_{iu}^*, $u = 1, \ldots, U_i$, for $i = 1, \ldots, \ell$. That is, π_{ij}, $j = 1, \ldots, n_i$ have a histogram where the midpoints of the categories are the π_{iu}^*, for $i = 1, \ldots, \ell$. These π_{iu}^* are assumed known and the π_{ij} are assumed to be random quantities. For notational convenience, let $\boldsymbol{\theta}_i = \left(\theta_{i10}, \ldots, \theta_{iU_i0}, \theta_{i11}, \ldots, \theta_{iU_i1}\right) = (\boldsymbol{\theta}_{i0}, \boldsymbol{\theta}_{i1})$ (say), $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_\ell)$, and $\boldsymbol{\pi} = \left(\pi_{11}, \ldots, \pi_{1n_1}, \ldots, \pi_{\ell 1}, \ldots, \pi_{\ell n_\ell}\right)$. The distribution of the selection probabilities, given the binary response y_{ij}, is

$$P\left(\pi_{ij} = \pi_{iu}^* \mid \boldsymbol{\theta}, y_{ij} = y\right) = \theta_{iuy}, \ u = 1, \ldots, U_i, \ y = 0, 1, \ j = 1, \ldots, n_i, \ i = 1, \ldots, \ell$$

and

$$y_{ij} \mid p_i \overset{\text{iid}}{\sim} \text{Bernoulli}(p_i), \ j = 1, \ldots, N_i, \ i = 1, \ldots, \ell.$$

Again, it is worth noting that the θ_{iuy} are not selection probabilities.

To accommodate the sample selection scheme, we assume that θ_{iu0} and θ_{iu1}, for $i = 1, \ldots, \ell$, are different. Note that we consider the heterogeneity assumption for the sample selection probabilities. We replace the homogeneity assumption with the heterogeneity assumption for the sample selection probabilities of the HoS model, so that the sample selection probabilities have different supports and the distributions of the selection probabilities are different by areas.

Let δ_{ij}, π_{ij} and y_{ij} denote the selection indicator, the selection probability and the binary response of the j^{th} unit in the i^{th} small area in the population respectively. Essentially, NBBS postulated that the $(\delta_{ij}, \pi_{ij}, y_{ij})$ within the i^{th} small area are independently distributed with

$$\begin{aligned} P\left(\delta_{ij} = \delta, \pi_{ij} = \pi_{iu}^{*}, y_{ij} = y \mid \boldsymbol{\theta}_i, p_i\right) &= P_1\left(\delta_{ij} = \delta \mid \pi_{ij} = \pi_{iu}^{*}\right) P_2\left(\pi_{ij} = \pi_{iu}^{*} \mid y_{ij} = y, \boldsymbol{\theta}_i\right) P_3\left(y_{ij} = y \mid p_i\right) \\ &= \left(\pi_{iu}^{*}\right)^{\delta} \left(1 - \pi_{iu}^{*}\right)^{1-\delta} \theta_{iuy} p_i^{y} \left(1 - p_i\right)^{1-y}, \\ &\quad \delta = 0, 1,\ u = 1, \ldots, U_i,\ y = 0, 1,\ j = 1, \ldots, n_i,\ i = 1, \ldots, \ell. \end{aligned}$$

Thus, there is a joint probability mass function for the selection indicator and the response indicator. Therefore, the model that NBBS specified is a nonignorable selection model (i.e., NBBS assumed that the selection mechanism is selection not at random (SNAR)). Since there are no data when $\delta = 0$ (i.e., π and y are both unobserved), NBBS used the conditional probability mass function

$$P\left(\pi_{ij} = \pi_{iu}^{*}, y_{ij} = y \mid \delta_{ij} = 1, \boldsymbol{\theta}_i, p_i\right) = \frac{\pi_{iu}^{*} \theta_{iuy} p_i^{y} \left(1 - p_i\right)^{1-y}}{\sum_{y=0}^{1} \sum_{u=1}^{U_i} \pi_{iu}^{*} \theta_{iuy} p_i^{y} \left(1 - p_i\right)^{1-y}},\ i = 1, \ldots, \ell;$$

see the probability mass function in (4) of NBBS.

We have the data (π_{ij}, y_{ij}), $j = 1, \ldots, n_i$, $i = 1, \ldots, \ell$. Since the sampling units are independent, the likelihood function is given by

$$\prod_{i=1}^{\ell} \prod_{j=1}^{n_i} P\left(\pi_{ij} = \pi_{iu}^{*}, Y_{ij} = y_{ij} \mid \delta_{ij} = 1, \boldsymbol{\theta}_i, \mathbf{p}\right) = \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \frac{\pi_{iu}^{*} \theta_{iuy_{ij}} p_i^{y_{ij}} \left(1 - p_i\right)^{1-y_{ij}}}{\sum_{y_{ij}=0}^{1} \sum_{u=1}^{U_i} \pi_{iu}^{*} \theta_{iuy_{ij}} p_i^{y_{ij}} \left(1 - p_i\right)^{1-y_{ij}}},$$

where π_{iu}^{*}, $u = 1, \ldots, U_i$, $i = 1, \ldots, \ell$ are known. The likelihood function can be rewritten as

$$\begin{aligned} P(\mathbf{y}, \boldsymbol{\pi} \mid \boldsymbol{\theta}, \mathbf{p}) &= \frac{\prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \pi_{iu}^{*} \theta_{iuy_{ij}} \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} p_i^{y_{ij}} \left(1 - p_i\right)^{1-y_{ij}}}{\prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \sum_{y_{ij}=0}^{1} \sum_{u=1}^{U_i} \pi_{iu}^{*} \theta_{iuy_{ij}} p_i^{y_{ij}} \left(1 - p_i\right)^{1-y_{ij}}} \\ &= \frac{\prod_{i=1}^{\ell} \prod_{u=1}^{U_i} \left(\pi_{iu}^{*} \theta_{iu0}\right)^{g_{iu0}} \prod_{i=1}^{\ell} \prod_{u=1}^{U_i} \left(\pi_{iu}^{*} \theta_{iu1}\right)^{g_{iu1}} \prod_{i=1}^{\ell} p_i^{s_i} \left(1 - p_i\right)^{n_i - s_i}}{\prod_{i=1}^{\ell} \left[p_i \sum_{u=1}^{U_i} \pi_{iu}^{*} \theta_{iu1} + \left(1 - p_i\right) \sum_{u=1}^{U_i} \pi_{iu}^{*} \theta_{iu0}\right]^{n_i}}, \end{aligned}$$

where $s_i = \sum_{j=1}^{n_i} y_{ij}$, g_{iu0} is the cell count for category u at $y = 0$ and g_{iu1} is the cell count for category u at $y = 1$ under the area i. Note that $\sum_{u=1}^{U_i} g_{iu0} = n_i - s_i$, $\sum_{u=1}^{U_i} g_{iu1} = s_i$ and $\sum_{u=1}^{U_i} \left(g_{iu0} + g_{iu1}\right) = n_i$. This likelihood includes the selection bias.

Let $a_{iy} = \sum_{u=1}^{U_i} \pi_{iu}^* \theta_{iuy}$, $y = 0, 1$, $i = 1, \ldots, \ell$. Differences between a_{i0} and a_{i1} for some i indicate that there is selection bias. The likelihood function can be expressed as

$$P(\mathbf{y}, \boldsymbol{\pi} \mid \boldsymbol{\theta}, \mathbf{p}) = \prod_{i=1}^{\ell} \prod_{u=1}^{U_i} (\pi_{iu}^* \theta_{iu0})^{g_{iu0}} \prod_{i=1}^{\ell} \prod_{u=1}^{U_i} (\pi_{iu}^* \theta_{iu1})^{g_{iu1}} \prod_{i=1}^{\ell} \frac{p_i^{s_i} (1 - p_i)^{n_i - s_i}}{[a_{i1} p_i + a_{i0} (1 - p_i)]^{n_i}}.$$

We make a one-to-one transformation from p_i to q_i via

$$q_i = \frac{a_{i1} p_i}{a_{i1} p_i + a_{i0} (1 - p_i)}.$$

Note that if θ_{iuy} does not depend on y, q_i and p_i are the same. In this case the likelihood would be the same as the ignorable case. Let $\mathbf{q} = (q_1, \ldots, q_\ell)$. Then the likelihood function can be expressed as

$$P(\mathbf{y}, \boldsymbol{\pi} \mid \boldsymbol{\theta}, \mathbf{q}) = \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i} (\pi_{iu}^* \theta_{iu0})^{g_{iu0}}}{\left(\sum_{u=1}^{U_i} \pi_{iu}^* \theta_{iu0}\right)^{n_i - s_i}} \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i} (\pi_{iu}^* \theta_{iu1})^{g_{iu1}}}{\left(\sum_{u=1}^{U_i} \pi_{iu}^* \theta_{iu1}\right)^{s_i}} \prod_{i=1}^{\ell} q_i^{s_i} (1 - q_i)^{n_i - s_i}.$$

We assume that $\mathbf{q}$, $\boldsymbol{\theta}_{i0}$, $\boldsymbol{\theta}_{i1}$ are independent, and we take

$$q_i \mid \mu, \tau_1 \overset{\text{iid}}{\sim} \text{Beta}(\mu\tau_1, (1 - \mu)\tau_1), \ i = 1, \ldots, \ell,$$

$$\boldsymbol{\theta}_{i0} \mid \tau_0 \overset{\text{ind}}{\sim} \text{Dirichlet}(\boldsymbol{\theta}_{i0}^{(0)} \tau_0) \quad \text{and} \quad \boldsymbol{\theta}_{i1} \mid \tau_0 \overset{\text{ind}}{\sim} \text{Dirichlet}(\boldsymbol{\theta}_{i1}^{(0)} \tau_0),$$

where $\boldsymbol{\theta}_{i0}^{(0)}$ and $\boldsymbol{\theta}_{i1}^{(0)}$ are to be specified. Recall that $\mathbf{x} \mid \boldsymbol{\mu}, \tau \sim \text{Dirichlet}(\boldsymbol{\mu}\tau)$ has the density $f(\mathbf{x} \mid \boldsymbol{\mu}, \tau) = \frac{\prod_{i=1}^{k} x_i^{\mu_i \tau - 1}}{D(\boldsymbol{\mu}\tau)}$, $0 < x_i < 1$, $\sum_{i=1}^{k} x_i = 1$, where $D(\boldsymbol{\mu}\tau) = \prod_{i=1}^{k} \Gamma(\mu_i \tau) / \Gamma(\tau)$, $0 < \mu_i < 1$, $\sum_{i=1}^{k} \mu_i = 1$, and $\tau > 0$.

Finally, a priori we assume

$$p(\mu, \tau_0, \tau_1) = \frac{1}{(1 + \tau_0)^2} \frac{1}{(1 + \tau_1)^2}, \ 0 < \mu < 1, \tau_0, \tau_1 \geq 0.$$

Of course one can use a half Cauchy density but these are very similar. This latter prior is used to avoid the difficulties associated with improper priors of the form $p(\tau) \propto 1/\tau$ (e.g., see Gelman, 2006).

Hence, the joint prior density of $\mathbf{q}$, $\boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1$ is

$$\pi(\mathbf{q}, \boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1) \propto \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i} \theta_{iu0}^{\theta_{iu0}^{(0)} \tau_0 - 1}}{D(\boldsymbol{\theta}_{i0}^{(0)} \tau_0)} \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i} \theta_{iu1}^{\theta_{iu1}^{(0)} \tau_0 - 1}}{D(\boldsymbol{\theta}_{i1}^{(0)} \tau_0)}$$
$$\times \frac{\prod_{i=1}^{\ell} q_i^{\mu\tau_1 - 1} (1 - q_i)^{(1-\mu)\tau_1 - 1}}{[B(\mu\tau_1, (1 - \mu)\tau_1)]^{\ell}} \frac{1}{(1 + \tau_0)^2} \frac{1}{(1 + \tau_1)^2}.$$

Using Bayes' theorem, the joint posterior density of $\mathbf{q}$, $\boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1$, given the data $\mathbf{y}$, is

$$\pi(\mathbf{q}, \boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1 \,|\, \mathbf{y}) \propto \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i} (\pi_{iu}^* \theta_{iu0})^{g_{iu0}}}{\left(\sum_{u=1}^{U_i} \pi_{iu}^* \theta_{iu0}\right)^{n_i - s_i}} \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i} (\pi_{iu}^* \theta_{iu1})^{g_{iu1}}}{\left(\sum_{u=1}^{U_i} \pi_{iu}^* \theta_{iu1}\right)^{s_i}}$$
$$\times \prod_{i=1}^{\ell} q_i^{s_i} (1 - q_i)^{n_i - s_i} \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i} \theta_{iu0}^{\theta_{iu0}^{(0)} \tau_0 - 1}}{D(\boldsymbol{\theta}_{i0}^{(0)} \tau_0)} \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i} \theta_{iu1}^{\theta_{iu1}^{(0)} \tau_0 - 1}}{D(\boldsymbol{\theta}_{i1}^{(0)} \tau_0)}$$
$$\times \frac{\prod_{i=1}^{\ell} q_i^{\mu\tau_1 - 1} (1 - q_i)^{(1-\mu)\tau_1 - 1}}{[B(\mu\tau_1, (1 - \mu)\tau_1)]^{\ell}} \frac{1}{(1 + \tau_0)^2} \frac{1}{(1 + \tau_1)^2}.$$

To improve the computations, we use the more optimal Rao-Blackwellization to get the posterior density of $\mathbf{q}$ (hence $\mathbf{p}$). Given the data, the joint posterior density can be expressed as

$$\pi(\mathbf{q}, \boldsymbol{\theta}, \mu, \tau_0, \tau_1 \,|\, \mathbf{y}) = \pi(\mathbf{q} \,|\, \boldsymbol{\theta}, \mu, \tau_0, \tau_1, \mathbf{y}) \pi(\boldsymbol{\theta}, \mu, \tau_0, \tau_1 \,|\, \mathbf{y}).$$

3.2 Computations

By integrating out $\mathbf{q}$ from the joint posterior of $\mathbf{q}, \boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1$ given $\mathbf{y}$, we get the marginal joint posterior density of $\boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1$ given $\mathbf{y}$,

$$\pi(\boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1 \,|\, \mathbf{y}) \propto \frac{\prod_{i=1}^{\ell} B(s_i + \mu\tau_1, n_i - s_i + (1 - \mu)\tau_1)}{[B(\mu\tau_1, (1 - \mu)\tau_1)]^{\ell}}$$
$$\times \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i} \theta_{iu0}^{g_{iu0} + \theta_{iu0}^{(0)} \tau_0 - 1}}{\left(\sum_{u=1}^{U_i} \pi_{iu}^* \theta_{iu0}\right)^{n_i - s_i}} \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i} \theta_{iu1}^{g_{iu1} + \theta_{iu1}^{(0)} \tau_0 - 1}}{\left(\sum_{u=1}^{U_i} \pi_{iu}^* \theta_{iu1}\right)^{s_i}}$$
$$\times \frac{1}{\prod_{i=1}^{\ell} D(\boldsymbol{\theta}_{i0}^{(0)} \tau_0)} \frac{1}{\prod_{i=1}^{\ell} D(\boldsymbol{\theta}_{i1}^{(0)} \tau_0)} \frac{1}{(1 + \tau_0)^2} \frac{1}{(1 + \tau_1)^2}.$$

The conditional posterior density of $\mathbf{q}$ is given by

$$\pi(\mathbf{q} \,|\, \boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1, \mathbf{y}) \propto \prod_{i=1}^{\ell} q_i^{s_i + \mu\tau_1 - 1} (1 - q_i)^{n_i - s_i + (1-\mu)\tau_1 - 1},$$

which we can sample directly.

We obtain a sample of $\mathbf{q}$ in the following manner: We draw each element of $(\boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1)$ from the conditional posterior density of $\boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1 \,|\, \mathbf{y}$ using the Metropolis-Hastings algorithm and a grid method, and then we draw $\mathbf{q}$ from the conditional posterior density of $\mathbf{q} \,|\, \boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1, \mathbf{y}$.

We have monitored the convergence of the Metropolis-Hastings sampler using trace plots, autocorrelation plots and Geweke test of stationarity, which showed satisfactory performance.

The conditional posterior densities needed to execute the Metropolis-Hastings sampler are

$$\pi\left(\boldsymbol{\theta}_{i0} \mid \boldsymbol{\theta}_{j0},\ j \neq i,\ \boldsymbol{\theta}_{i1},\ i=1,\ldots,\ell,\ \mu,\ \tau_0,\ \tau_1,\ \mathbf{y}\right) \propto \frac{\prod_{u=1}^{U_i} \theta_{iu0}^{g_{iu0}+\theta_{iu0}^{(0)}\tau_0-1}}{\left(\sum_{u=1}^{U_i} \pi_{iu}^{*}\theta_{iu0}\right)^{n_i-s_i}},\ i=1,\ldots,\ell,$$

$$\pi\left(\boldsymbol{\theta}_{i1} \mid \boldsymbol{\theta}_{j1},\ j \neq i,\ \boldsymbol{\theta}_{i0},\ i=1,\ldots,\ell,\ \mu,\ \tau_0,\ \tau_1,\ \mathbf{y}\right) \propto \frac{\prod_{u=1}^{U_i} \theta_{iu1}^{g_{iu1}+\theta_{iu1}^{(0)}\tau_0-1}}{\left(\sum_{u=1}^{U_i} \pi_{iu}^{*}\theta_{iu1}\right)^{s_i}},\ i=1,\ldots,\ell,$$

$$\pi\left(\mu \mid \boldsymbol{\theta}_{10},\ldots,\boldsymbol{\theta}_{\ell 0},\boldsymbol{\theta}_{11},\ldots,\boldsymbol{\theta}_{\ell 1},\tau_0,\tau_1,\mathbf{y}\right) \propto \frac{\prod_{i=1}^{\ell} B\left(s_i+\mu\tau_1,\ n_i-s_i+(1-\mu)\tau_1\right)}{\left[B\left(\mu\tau_1,(1-\mu)\tau_1\right)\right]^{\ell}},$$

and transforming τ_0 and τ_1 to respectively $\rho_0 = 1/(1+\tau_0)$ and $\rho_1 = 1/(1+\tau_1)$,

$$\pi\left(\rho_0 \mid \boldsymbol{\theta}_{10},\ldots,\boldsymbol{\theta}_{\ell 0},\boldsymbol{\theta}_{11},\ldots,\boldsymbol{\theta}_{\ell 1},\mu,\tau_1,\mathbf{y}\right) \propto \left[\prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i}\theta_{iu0}^{\theta_{iu0}^{(0)}\tau_0-1}}{D\left(\boldsymbol{\theta}_{i0}^{(0)}\tau_0\right)} \prod_{i=1}^{\ell} \frac{\prod_{u=1}^{U_i}\theta_{iu1}^{\theta_{iu1}^{(0)}\tau_0-1}}{D\left(\boldsymbol{\theta}_{i1}^{(0)}\tau_0\right)}\right]_{\tau_0=(1-\rho_0)/\rho_0},$$

$$\pi\left(\rho_1 \mid \boldsymbol{\theta}_{10},\ldots,\boldsymbol{\theta}_{\ell 0},\boldsymbol{\theta}_{11},\ldots,\boldsymbol{\theta}_{\ell 1},\mu,\rho_0,\mathbf{y}\right) \propto \left[\frac{\prod_{i=1}^{\ell} B\left(s_i+\mu\tau_1,\ n_i-s_i+(1-\mu)\tau_1\right)}{\left[B\left(\mu\tau_1,(1-\mu)\tau_1\right)\right]^{\ell}}\right]_{\tau_1=(1-\rho_1)/\rho_1}.$$

In above formula, the $\boldsymbol{\theta}_{i0}$, $\boldsymbol{\theta}_{i1}$ and μ are conditionally independent.

We use Metropolis steps to sample $\boldsymbol{\theta}_{i0},\ i=1,\ldots,\ell$ and $\boldsymbol{\theta}_{i1},\ i=1,\ldots,\ell$ from conditional distributions $\pi\left(\boldsymbol{\theta}_{i0} \mid \boldsymbol{\theta}_{j0},\ j\neq i,\ \boldsymbol{\theta}_{i1},\ i=1,\ldots,\ell,\ \mu,\ \tau_0,\ \tau_1,\ \mathbf{y}\right)$ and $\pi\left(\boldsymbol{\theta}_{i1} \mid \boldsymbol{\theta}_{j1},\ j\neq i,\ \boldsymbol{\theta}_{i0},\ i=1,\ldots,\ell,\right.$ $\left.\mu,\ \tau_0,\ \tau_1,\ \mathbf{y}\right)$, respectively. Let $\tau_{i0}^{*} = n_i - s_i + \tau_0$, $\theta_{iu0}^{*} = \left(g_{iu0}+\theta_{iu0}^{(0)}\tau_0\right)/\tau_{i0}^{*}$, $\tau_{i1}^{*} = s_i+\tau_0$, and $\theta_{iu1}^{*} = \left(g_{iu1}+\theta_{iu1}^{(0)}\tau_0\right)/\tau_{i1}^{*}$. For notational convenience, let $\boldsymbol{\theta}_{i0}^{*} = \left(\theta_{i10}^{*},\ldots,\theta_{iU_i0}^{*}\right)$ and $\boldsymbol{\theta}_{i1}^{*} = \left(\theta_{i11}^{*},\ldots,\theta_{iU_i1}^{*}\right)$. We choose overdispersed Dirichlet distributions as proposal densities for $\boldsymbol{\theta}_{i0}$ and $\boldsymbol{\theta}_{i1}$, $i=1,\ldots,\ell$. In fact, the proposal densities are $\text{Dirichlet}\left(\boldsymbol{\theta}_{i0}^{*}\tilde{\tau}_{i0}\right)$ and $\text{Dirichlet}\left(\boldsymbol{\theta}_{i1}^{*}\tilde{\tau}_{i1}\right)$, where $\tilde{\tau}_{i0} = \tau_{i0}^{*}/\kappa_{i0}$ and $\tilde{\tau}_{i1} = \tau_{i1}^{*}/\kappa_{i1}$, for all $i=1,\ldots,\ell$. Note that as the κ_{i0} and κ_{i1} increase, the dispersion tends to increase in the Dirichlet distribution.

Assuming that the Metropolis-Hastings sampler is at $\boldsymbol{\theta}_{i0}^{(r)}$, then the probability of accepting $\boldsymbol{\theta}_{i0}^{(r+1)}$ is

$$U_{r,r+1}^{(0)} = \min\left\{\frac{\psi\left(\boldsymbol{\theta}_{i0}^{(r+1)},\mathbf{y}\right)}{\psi\left(\boldsymbol{\theta}_{i0}^{(r)},\mathbf{y}\right)},1\right\},$$

where $\psi\left(\boldsymbol{\theta}_{i0},\mathbf{y}\right) = \frac{\prod_{u=1}^{U_i}\theta_{iu0}^{\theta_{iu0}^{*}\tau_{i0}^{*}(1-1/\kappa_{i0})}}{\left(\sum_{u=1}^{U_i}\pi_{iu}^{*}\theta_{iu0}\right)^{n_i-s_i}}$, for all $i=1,\ldots,\ell$. Also, if we assume that the Metropolis-Hastings sampler is at $\boldsymbol{\theta}_{i1}^{(r)}$, then the probability of accepting $\boldsymbol{\theta}_{i1}^{(r+1)}$ is

$$U_{r,r+1}^{(1)} = \min\left\{\frac{\psi\left(\boldsymbol{\theta}_{i1}^{(r+1)},\mathbf{y}\right)}{\psi\left(\boldsymbol{\theta}_{i1}^{(r)},\mathbf{y}\right)},1\right\},$$

where $\psi\left(\boldsymbol{\theta}_{i1},\mathbf{y}\right) = \frac{\prod_{u=1}^{U_i}\theta_{iu1}^{\theta_{iu1}^{*}\tau_{i1}^{*}(1-1/\kappa_{i1})}}{\left(\sum_{u=1}^{U_i}\pi_{iu}^{*}\theta_{iu1}\right)^{s_i}}$, for all $i=1,\ldots,\ell$.

The Metropolis step is obtained as follows: Assume the Markov chain is at $\boldsymbol{\theta}_{i0}^{(r)}$, a random vector $\boldsymbol{\theta}_{i0}^{(r+1)}$ is drawn from the proposal density with properly chosen κ_{i0}, and $U_{r,r+1}^{(0)}$ is computed. A random uniform deviate U in $[0, 1]$ is drawn, and if $U \le U_{r,r+1}^{(0)}$, then random vector $\boldsymbol{\theta}_{i0}^{(r+1)}$ is accepted, otherwise the chain stays at $\boldsymbol{\theta}_{i0}^{(r)}$. This algorithm is applied for all $i = 1, \ldots, \ell$. The Metropolis step is utilized in a manner similar to that for $\boldsymbol{\theta}_{i1}$ for all $i = 1, \ldots, \ell$.

We generate μ, ρ_0 and ρ_1, using the grid method; see CNK. Once we get the sample from the posterior $\pi(\mathbf{q}, \boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1 \,|\, \mathbf{y})$, by retransforming from q_i to p_i,

$$p_i = \frac{a_{i0} q_i}{a_{i0} q + a_{i1}(1 - q_i)},$$

we can draw a sample from the posterior distribution of $\pi(\mathbf{p}, \boldsymbol{\theta}_{10}, \ldots, \boldsymbol{\theta}_{\ell 0}, \boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{\ell 1}, \mu, \tau_0, \tau_1 \,|\, \mathbf{y})$.

Once $\mathbf{p}$ is estimated, we draw the entire finite population values, $y_{i1}, \ldots, y_{iN_i}$, independently from $\text{Bernoulli}(p_i)$, $i = 1, \ldots, \ell$. This is surrogate sampling (e.g., Nandram, 2007). So we have corrected the observed biased sample and replaced it by a surrogate sample for $\mathbf{p}$ that we obtained from the heterogeneous nonignorable selection model. We can obtain a sample of P_i by drawing $\sum_{i=1}^{N_i} y_{ij}$ from $\text{Binomial}(N_i, p_i)$ and by dividing the result by N_i for all $i = 1, \ldots, \ell$.

The selection mechanism is similar to the missing data mechanism. So it is possible to incorporate missing data into our framework, or independently (i.e., on its own) we can assume that the missing data are “missing not at random” and a nonignorable nonresponse model can be used to adjust a population model, see Nandram and Choi (2010).

4. Numerical studies

In Section 4.1, we describe an example on severe activity limitation. We present the results of the IS, HoS and HeS models for the comparison. In Section 4.2, we describe a simulation study to assess the performance of the three models under two kinds of assumptions of the distribution of the sample selection probabilities. That is, data are generated from either the homogeneous nonignorable or the heterogeneous nonignorable selection model and all three models are fit.

4.1 Illustrative example

In our application, we use data from the 1995 National Health Interview Survey (NHIS95). These data were first used by Nandram, Bai and Choi (2011) to estimate change point in activity limitation. NBBS constructed synthetic data arising from a segment of NHIS95 for this study. For adults, 30-80 years old, NBBS analyzed data on severe activity limitation (SAL) for a single area (no pooling), where $y = 1$ if an adult has SAL and $y = 0$ otherwise. CNK used the data from NBBS to perform small area estimation on SAL. We use the data in a manner similar to NBBS but we fit a more general model, the key contribution of this paper.

NBBS formed twelve domains (small areas) by crossing education, sex and race. They have categorized education into three levels (pre-college: L, college: M and post-college: H). Sex (male: M and female: F) and Race (white: W and nonwhite: B) have two levels each. We will continue to call these domains LMW, LMB, LFW, LFB, MMW, MMB, MFW, MFB, HMW, HMB, HFW, HFB (e.g., LMW: white males with pre-college education, LMB: black males with pre-college education, etc.).

NHIS95 used a multistage sample design to draw samples from the population of the United States. Therefore, it is necessary to use an adult's survey weight for accurate analysis. See NBBS and CNK for a discussion of the survey weights. Like NBBS and CNK, we considered the reciprocal of a survey weight as the "selection" probability of each adult. Selection probabilities are a major part of the survey weights. It would have been better if we had the selection probabilities. But this is an approximation we make in our data analysis. For convenience, we have presented the data again in Table 4.1.

Table 4.1
Summaries of the data on severe activity limitation (SAL) including selection probabilities

Domain	n	s	p	CV	$f = n/N$	avg0	avg1	p-value
LMW	174	26	0.149	0.181	0.0000303	0.0004059	0.0003706	0.025*
LMB	31	10	0.323	0.260	0.0000280	0.0003179	0.0003167	0.614
LFW	200	22	0.110	0.201	0.0000325	0.0004523	0.0004381	1.00
LFB	40	9	0.225	0.293	0.0000286	0.0003152	0.0003740	0.910
MMW	760	56	0.074	0.129	0.0000227	0.0002710	0.0002901	0.006*
MMB	151	18	0.119	0.221	0.0000250	0.0002820	0.0002847	0.056
MFW	892	42	0.047	0.151	0.0000233	0.0002891	0.0002415	0.837
MFB	200	21	0.105	0.206	0.0000278	0.0003064	0.0003377	0.469
HMW	756	14	0.019	0.265	0.0000213	0.0002484	0.0001919	0.095
HMB	124	7	0.056	0.367	0.0000238	0.0002702	0.0003506	0.146
HFW	779	22	0.028	0.210	0.0000219	0.0002575	0.0002976	0.072
HFB	168	2	0.012	0.703	0.0000257	0.0002969	0.0001948	1.00

NOTE: Here n is the total sample size, s is the number of adults with SAL, and $p = s/n$; $f = n/N$ is the sampling fraction; CV $= \frac{\sqrt{(1-p)}}{\sqrt{np}}$ is the coefficient of variation; avg0 is the average of the selection probabilities for $y = 0$ (SAL, no) and avg1 is the average of the selection probabilities for $y = 1$ (SAL, yes); p-value corresponds to that of a chi-squared test of equality of $\theta_{iu0} = \theta_{iu1}$, $u = 1, \ldots, U_i$ $(U_i = 5)$; domains are formed by crossing education (L, M, H), sex (M, F) and race (W, B).

We reduce the sample size from the original data set to increase the effect of small area model. For the HeS model, we order the selection probabilities from smallest to largest $\pi_{(1)}, \ldots, \pi_{(n_i)}$ within each small area. Let the quantiles be $t_{i1} = \pi_{(0.20 n_i)}, t_{i2} = \pi_{(0.40 n_i)}, t_{i3} = \pi_{(0.60 n_i)}, t_{i4} = \pi_{(0.80 n_i)}$ and let $t_{i0} = \pi_{(1)}$ and $t_{i5} = \pi_{(n_i)}$ for $i = 1, \ldots, \ell$. We define $\pi_{iu}^{*} = \left(t_{i(u-1)} + t_{iu}\right)/2$, $u = 1, \ldots, U_i$, $i = 1, \ldots, \ell$ (i.e., the mid point of each quantile within each small area). Note that θ_{iuy} is the proportion of sampled units in the u^{th} quantile conditional on $y = 0, 1$ for $i = 1, \ldots, \ell$. If the θ_{iu0} are considerably different from θ_{iu1} in some areas, there is strong evidence that the sampled values are biased. To specify $\boldsymbol{\theta}_{i0}^{(0)}$ and $\boldsymbol{\theta}_{i1}^{(0)}$ in the prior distributions, we take $\boldsymbol{\theta}_{iy}^{(0)} = \hat{\boldsymbol{\theta}}_{iy}$, the maximum likelihood estimator of $\boldsymbol{\theta}_{iy}$, $y = 0, 1$, $i = 1, \ldots, \ell$, which we call a "MLE" prior (see NBBS).

For our data, mixing is slow, so we draw 120,000 samples and burn in 60,000. Then we take every 30^{th} iterate to obtain a sample of 2,000 iterates for inference. This burn-in period is sufficiently long to get random samples, which is based on the trace plots and Geweke test. We have enough samples since the effective sample size (ESS) of parameters sampled from an MCMC should be less than or equal to 2,000. The correlation is nonsignificant for all parameters. Also, stationarity of our sampler is demonstrated by Geweke test. The results of convergence diagnostic for hyper-parameters are shown in Table 4.8 and Figure 4.2. By taking $\boldsymbol{\kappa}_0 = (7.3, 3, 11, 4.8, 8.5, 9, 11.5, 10, 10, 9.7, 10.2, 11)$ and $\boldsymbol{\kappa}_1 = (3.2, 1.7, 2.2, 1.5, 5, 2.2, 3.3, 2.2, 2.5, 1.5, 1.5, 1.3)$, we obtained acceptance rates between 25% and 45% of Metropolis-Hastings samplers.

The summaries of data are shown in Table 4.1. The averages of the selection probabilities in each area are mostly similar for $y = 0$ and $y = 1$. Since the sample sizes within the domains are not quite large, the sampling fractions are very small. Thus, the simultaneous analysis using a small area model is appropriate. In column 4 of Table 4.1 we have also presented the proportion of individuals with SAL, and we can see that this value is relatively large for low level education. We compared the counts in the two sets of bins from the histograms of the selection probabilities for $y = 0, 1$ in each domain. In fact, this is a test of independence in a $2 \times U_i$ categorical table, and we use a chi-squared test and a Fisher's exact test for equality of $\theta_{iu0} = \theta_{iu1}$, $u = 1, \ldots, U_i$ ($U_i = 5$ cells) in each domain. As is evident from the p-values which are presented in the last column of Table 4.1, selection bias should matter mostly in Domain MMW and perhaps in Domains LMW, MMB, HMW and HFW. As a measure of selection bias, we also calculated the biserial correlation between the y_{ij} and π_{ij} for all areas combined and got a value of 0.033 with a p-value of 0.03.

In Table 4.2, we provide the results of the unpooled (individual analysis) of the finite population proportions under the ignorable selection model and the nonignorable selection model for each domain separately (the nonignorable selection model is the NBBS model). We compare them using the posterior means (PM), the posterior standard deviations (PSD) of PMs, the coefficient of variation (CV) and 95% highest posterior density (HPD) intervals. This is a repetition of the analysis under the NBBS model.

In Table 4.3, we compare summaries of the pooled estimators of the finite population proportions under the IS model, the HoS model and the HeS model for the 12 domains. The effect of the selection bias is seen because the PMs under the IS model and the other models are different. The estimators under the HoS model and the HeS model are similar in some domains because the selection bias is more severe for some domains. The PSDs under the HoS model and the HeS model are bigger than under the IS model in most domains, but the 95% HPD intervals overlap. The effect of the simultaneous analysis is also seen because the PMs for pooled estimators are smoothed relative to PMs for separately analyzed finite populations.

The posterior summaries of $\boldsymbol{\theta}_{i0}$ and $\boldsymbol{\theta}_{i1}$ for $i = 1, \ldots, \ell$ are shown in Tables 4.4. The θ_{iu0} are different from the θ_{iu1} in some areas. It is strong evidence to indicate the presence of selection bias. We also present the posterior summaries of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ under the HoS model in Table 4.5. Note that in Tables 4.4-4.5, these θ_{i0} and θ_{i1} are not selection probabilities. The $\boldsymbol{\theta}$ under the HoS model appears different from those under the HeS model. We present the posterior summaries of a_{i0} and a_{i1} for $i = 1, \ldots, \ell$ in Table 4.7, and a_0 and a_1 of the HoS model in Table 4.6. These values are very small, but its ratio is large in each area. It is also strong evidence of selection bias, albeit these are small sample sizes within areas.

Table 4.2
Comparisons of the unpooled estimators of the finite population proportions under the ignorable and nonignorable selection models for individual domain

Domain	Ignorable Selection Model				Nonignorable Selection Model			
	PM	PSD	CV	CI	PM	PSD	CV	CI
LMW	0.154	0.027	0.175	(0.098, 0.214)	0.188	0.033	0.176	(0.123, 0.265)
LMB	0.333	0.080	0.240	(0.177, 0.525)	0.336	0.083	0.247	(0.170, 0.531)
LFW	0.114	0.022	0.193	(0.068, 0.165)	0.113	0.022	0.195	(0.069, 0.165)
LFB	0.244	0.066	0.270	(0.108, 0.394)	0.215	0.064	0.298	(0.090, 0.361)
MMW	0.075	0.010	0.133	(0.055, 0.100)	0.068	0.009	0.132	(0.050, 0.086)
MMB	0.124	0.026	0.210	(0.073, 0.185)	0.114	0.024	0.211	(0.065, 0.171)
MFW	0.048	0.008	0.167	(0.032, 0.065)	0.052	0.008	0.154	(0.036, 0.072)
MFB	0.108	0.022	0.204	(0.066, 0.161)	0.104	0.021	0.202	(0.060, 0.148)
HMW	0.020	0.005	0.250	(0.008, 0.032)	0.021	0.005	0.238	(0.011, 0.034)
HMB	0.065	0.022	0.338	(0.021, 0.119)	0.046	0.018	0.391	(0.014, 0.089)
HFW	0.030	0.006	0.200	(0.017, 0.045)	0.021	0.005	0.238	(0.011, 0.032)
HFB	0.017	0.010	0.588	(0.001, 0.043)	0.021	0.012	0.571	(0.001, 0.048)

NOTE: Here PM is the posterior mean; PSD is the posterior standard deviation; CV is the coefficient of variation; CI is the 95% HPD interval.

Table 4.3
Comparisons of the pooled estimators of the finite population proportions under the ignorable selection (IS) model, homogeneous nonignorable selection (HoS) model and heterogeneous nonignorable selection (HeS) models by domain

Domain	Model	PM	PSD	CV	CI
LMW	IS	0.148	0.026	0.176	(0.094, 0.210)
	HoS	0.177	0.030	0.169	(0.123, 0.238)
	HeS	0.172	0.037	0.215	(0.099, 0.264)
LMB	IS	0.264	0.071	0.269	(0.109, 0.426)
	HoS	0.310	0.082	0.265	(0.167, 0.483)
	HeS	0.267	0.076	0.285	(0.120, 0.449)
LFW	IS	0.110	0.021	0.191	(0.062, 0.158)
	HoS	0.134	0.026	0.194	(0.088, 0.189)
	HeS	0.116	0.027	0.233	(0.058, 0.179)
LFB	IS	0.198	0.057	0.288	(0.083, 0.326)
	HoS	0.237	0.064	0.270	(0.122, 0.377)
	HeS	0.192	0.060	0.313	(0.074, 0.332)
MMW	IS	0.074	0.009	0.122	(0.054, 0.095)
	HoS	0.091	0.011	0.121	(0.071, 0.115)
	HeS	0.081	0.013	0.160	(0.054, 0.111)
MMB	IS	0.119	0.026	0.218	(0.067, 0.178)
	HoS	0.144	0.030	0.208	(0.092, 0.207)
	HeS	0.120	0.027	0.225	(0.060, 0.183)
MFW	IS	0.048	0.007	0.146	(0.032, 0.064)
	HoS	0.059	0.009	0.153	(0.044, 0.078)
	HeS	0.060	0.011	0.183	(0.036, 0.085)
MFB	IS	0.106	0.021	0.198	(0.062, 0.154)
	HoS	0.128	0.026	0.203	(0.084, 0.183)
	HeS	0.105	0.023	0.219	(0.059, 0.160)
HMW	IS	0.020	0.005	0.250	(0.010, 0.032)
	HoS	0.025	0.006	0.240	(0.014, 0.039)
	HeS	0.027	0.008	0.296	(0.013, 0.047)
HMB	IS	0.062	0.021	0.339	(0.019, 0.111)
	HoS	0.075	0.025	0.333	(0.035, 0.131)
	HeS	0.057	0.020	0.351	(0.019, 0.108)
HFW	IS	0.029	0.006	0.207	(0.017, 0.044)
	HoS	0.037	0.008	0.216	(0.023, 0.053)
	HeS	0.027	0.006	0.222	(0.014, 0.042)
HFB	IS	0.018	0.010	0.556	(0.001, 0.043)
	HoS	0.022	0.012	0.545	(0.005, 0.050)
	HeS	0.020	0.012	0.600	(0.001, 0.048)

Table 4.4
Comparison of θ_{i0} and θ_{i1} using PM, PSD, NSE and 95% HPD interval under the heterogeneous nonignorable selection model

Domain	u	θ_0				θ_1			
		PM	PSD	CV	CI	PM	PSD	CV	CI
LMW	1	0.260	0.083	0.319	(0.096, 0.447)	0.477	0.109	0.229	(0.252, 0.714)
	2	0.213	0.075	0.352	(0.073, 0.397)	0.195	0.087	0.446	(0.048, 0.401)
	3	0.214	0.070	0.327	(0.084, 0.383)	0.126	0.070	0.556	(0.008, 0.296)
	4	0.189	0.064	0.339	(0.065, 0.342)	0.079	0.052	0.658	(0.003, 0.208)
	5	0.124	0.048	0.387	(0.027, 0.235)	0.121	0.057	0.471	(0.019, 0.252)
LMB	1	0.318	0.098	0.308	(0.101, 0.536)	0.355	0.100	0.282	(0.140, 0.570)
	2	0.279	0.086	0.308	(0.058, 0.422)	0.176	0.080	0.455	(0.020, 0.361)
	3	0.209	0.083	0.397	(0.048, 0.340)	0.148	0.073	0.493	(0.021, 0.324)
	4	0.126	0.065	0.516	(0.015, 0.286)	0.229	0.083	0.362	(0.063, 0.414)
	5	0.114	0.059	0.518	(0.012, 0.254)	0.093	0.054	0.581	(0.007, 0.224)
LFW	1	0.279	0.087	0.312	(0.101, 0.473)	0.301	0.090	0.299	(0.113, 0.500)
	2	0.232	0.083	0.358	(0.075, 0.431)	0.215	0.083	0.386	(0.051, 0.410)
	3	0.194	0.075	0.387	(0.054, 0.371)	0.203	0.079	0.389	(0.056, 0.390)
	4	0.167	0.065	0.389	(0.046, 0.325)	0.191	0.071	0.372	(0.046, 0.348)
	5	0.128	0.052	0.406	(0.025, 0.245)	0.089	0.045	0.506	(0.008, 0.198)
LFB	1	0.308	0.098	0.318	(0.120, 0.538)	0.276	0.094	0.341	(0.084, 0.480)
	2	0.247	0.095	0.385	(0.058, 0.463)	0.183	0.083	0.454	(0.024, 0.377)
	3	0.192	0.085	0.443	(0.032, 0.391)	0.200	0.086	0.430	(0.044, 0.410)
	4	0.158	0.074	0.468	(0.023, 0.326)	0.202	0.081	0.401	(0.050, 0.392)
	5	0.095	0.058	0.611	(0.005, 0.239)	0.139	0.069	0.496	(0.024, 0.308)
MMW	1	0.204	0.043	0.211	(0.116, 0.304)	0.167	0.078	0.467	(0.035, 0.355)
	2	0.221	0.044	0.199	(0.133, 0.326)	0.186	0.081	0.435	(0.039, 0.380)
	3	0.213	0.043	0.202	(0.126, 0.314)	0.198	0.082	0.414	(0.043, 0.396)
	4	0.189	0.040	0.212	(0.110, 0.282)	0.359	0.094	0.262	(0.174, 0.594)
	5	0.173	0.033	0.191	(0.104, 0.248)	0.090	0.047	0.522	(0.013, 0.205)
MMB	1	0.284	0.086	0.303	(0.107, 0.472)	0.248	0.095	0.383	(0.059, 0.458)
	2	0.230	0.079	0.343	(0.088, 0.418)	0.168	0.078	0.464	(0.023, 0.347)
	3	0.196	0.068	0.347	(0.055, 0.351)	0.176	0.078	0.443	(0.034, 0.353)
	4	0.148	0.062	0.419	(0.034, 0.297)	0.339	0.096	0.283	(0.134, 0.552)
	5	0.142	0.053	0.373	(0.038, 0.264)	0.069	0.046	0.667	(0.002, 0.186)
MFW	1	0.212	0.045	0.212	(0.127, 0.323)	0.305	0.093	0.305	(0.110, 0.520)
	2	0.218	0.045	0.206	(0.122, 0.321)	0.248	0.085	0.343	(0.084, 0.446)
	3	0.207	0.043	0.208	(0.117, 0.309)	0.191	0.076	0.398	(0.042, 0.364)
	4	0.201	0.043	0.214	(0.117, 0.311)	0.186	0.073	0.392	(0.049, 0.364)
	5	0.162	0.035	0.216	(0.090, 0.246)	0.071	0.038	0.535	(0.006, 0.161)
MFB	1	0.260	0.077	0.296	(0.110, 0.427)	0.331	0.093	0.281	(0.155, 0.568)
	2	0.229	0.073	0.319	(0.071, 0.387)	0.155	0.070	0.452	(0.029, 0.327)
	3	0.203	0.068	0.335	(0.067, 0.356)	0.156	0.069	0.442	(0.032, 0.319)
	4	0.179	0.061	0.341	(0.058, 0.317)	0.189	0.072	0.381	(0.050, 0.362)
	5	0.129	0.049	0.380	(0.032, 0.242)	0.168	0.064	0.381	(0.039, 0.319)
HMW	1	0.224	0.045	0.201	(0.138, 0.333)	0.277	0.105	0.379	(0.079, 0.531)
	2	0.206	0.044	0.214	(0.119, 0.313)	0.282	0.104	0.369	(0.074, 0.513)
	3	0.193	0.044	0.228	(0.104, 0.295)	0.257	0.099	0.385	(0.071, 0.491)
	4	0.210	0.044	0.210	(0.121, 0.312)	0.162	0.079	0.488	(0.025, 0.350)
	5	0.166	0.034	0.205	(0.099, 0.247)	0.022	0.026	1.182	(0.000, 0.097)
HMB	1	0.296	0.097	0.328	(0.095, 0.504)	0.265	0.099	0.374	(0.072, 0.493)
	2	0.223	0.085	0.381	(0.062, 0.423)	0.203	0.091	0.448	(0.027, 0.417)
	3	0.204	0.078	0.382	(0.058, 0.391)	0.125	0.073	0.584	(0.004, 0.305)
	4	0.166	0.070	0.422	(0.040, 0.323)	0.266	0.091	0.342	(0.091, 0.467)
	5	0.112	0.050	0.446	(0.026, 0.236)	0.142	0.069	0.486	(0.021, 0.307)
HFW	1	0.222	0.045	0.203	(0.126, 0.322)	0.191	0.078	0.408	(0.038, 0.367)
	2	0.217	0.047	0.217	(0.125, 0.329)	0.184	0.073	0.397	(0.039, 0.356)
	3	0.207	0.044	0.213	(0.110, 0.304)	0.169	0.071	0.420	(0.030, 0.332)
	4	0.199	0.043	0.216	(0.106, 0.296)	0.260	0.081	0.312	(0.097, 0.442)
	5	0.155	0.034	0.219	(0.079, 0.229)	0.196	0.050	0.255	(0.081, 0.348)
HFB	1	0.274	0.084	0.307	(0.116, 0.470)	0.366	0.114	0.311	(0.136, 0.631)
	2	0.223	0.075	0.336	(0.072, 0.400)	0.189	0.093	0.492	(0.012, 0.400)
	3	0.198	0.071	0.359	(0.060, 0.364)	0.212	0.095	0.448	(0.018, 0.421)
	4	0.177	0.064	0.362	(0.049, 0.317)	0.178	0.088	0.494	(0.024, 0.384)
	5	0.128	0.050	0.391	(0.035, 0.246)	0.055	0.050	0.909	(0.000, 0.194)

NOTE: u indicates the five intervals for the selection probabilities.

Table 4.5
Comparison of θ_0 and θ_1 using PM, PSD, NSE and 95% CI under the homogeneous nonignorable selection model

u	θ_0				θ_1			
	PM	**PSD**	**CV**	**CI**	**PM**	**PSD**	**CV**	**CI**
1	0.384	0.011	0.029	(0.363, 0.405)	0.632	0.036	0.057	(0.559, 0.700)
2	0.222	0.008	0.036	(0.206, 0.237)	0.160	0.026	0.163	(0.114, 0.217)
3	0.189	0.007	0.037	(0.176, 0.203)	0.112	0.018	0.161	(0.080, 0.150)
4	0.153	0.006	0.039	(0.142, 0.164)	0.060	0.009	0.150	(0.045, 0.079)
5	0.053	0.002	0.038	(0.049, 0.057)	0.036	0.005	0.139	(0.027, 0.047)

NOTE: u indicates the five intervals for the selection probabilities.

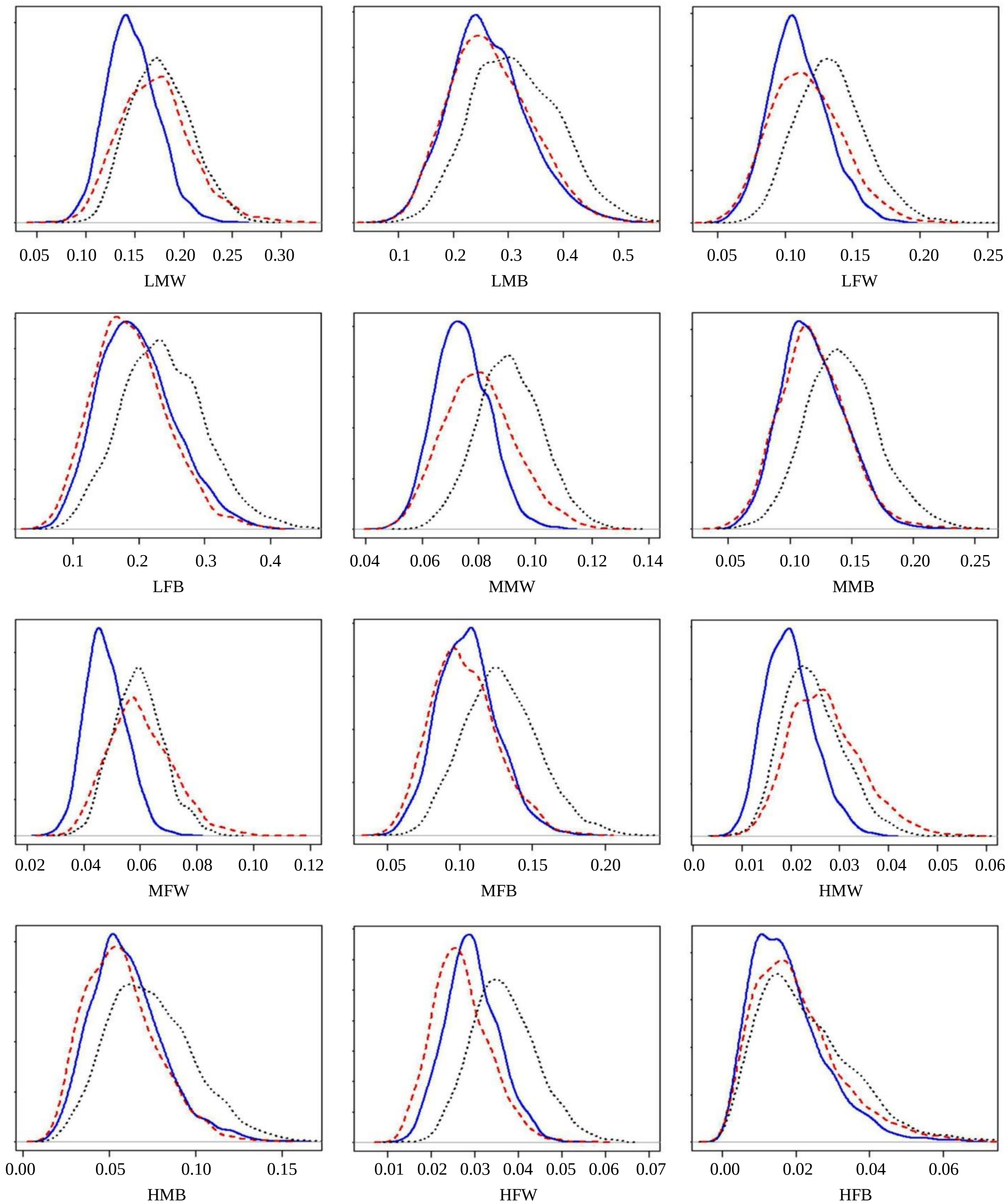
Table 4.6
Comparison of a_0 and a_1 using PM and PSD and NSE and 95% CI under the homogeneous nonignorable selection model

a_y	PM	PSD	CV	CI
a_0	0.000222	0.000002	0.009009	(0.000218, 0.000227)
a_1	0.000177	0.000006	0.033898	(0.000164, 0.000191)

Table 4.7
Comparison of a_{i0} and a_{i1} using PM, PSD, NSE and 95% HPD interval under the heterogeneous nonignorable selection model for each area

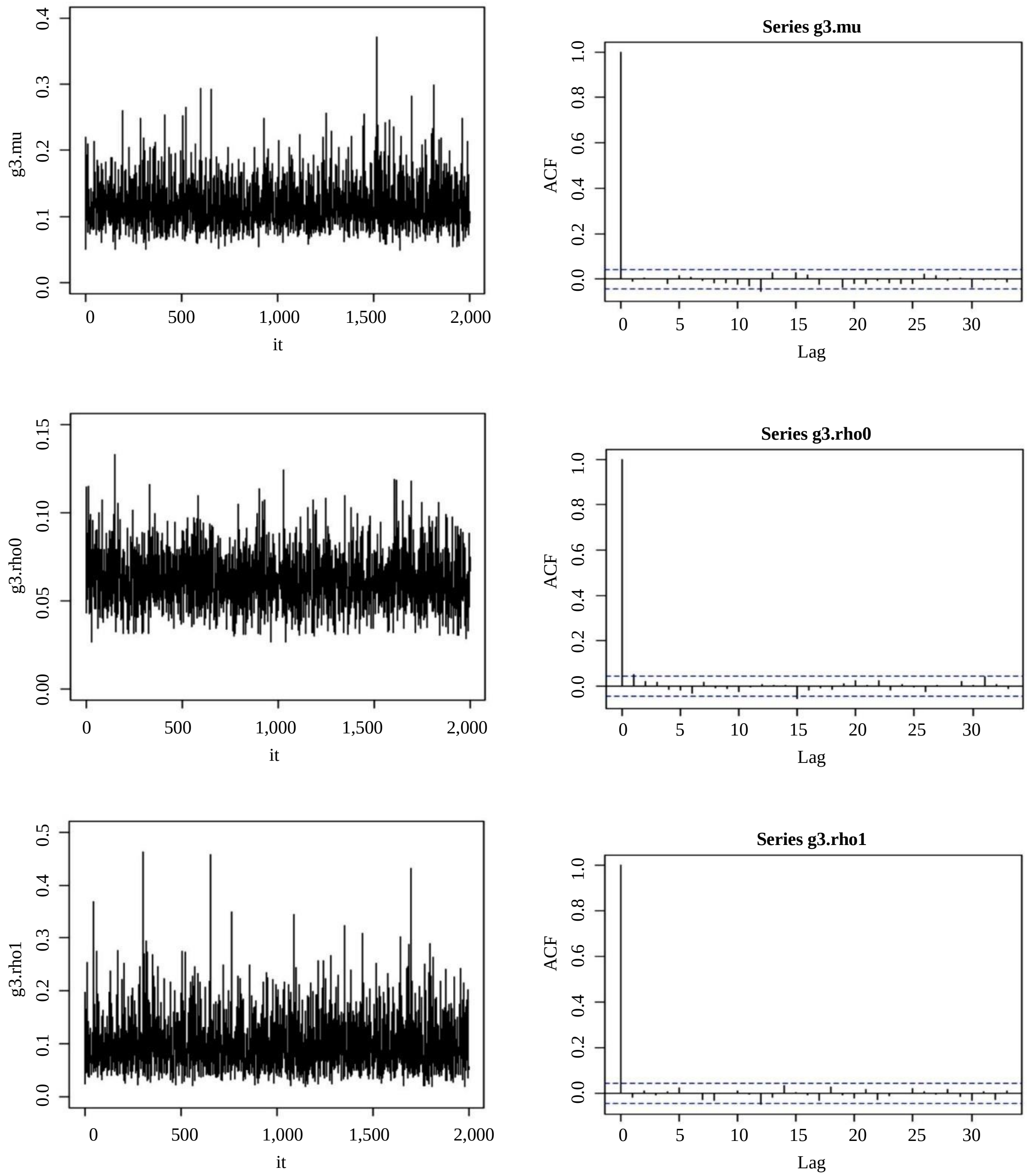
Domain	a_{iy}	PM	PSD	CV	CI
LMW	a_0	0.000367	0.000034	0.092643	(0.000296, 0.000441)
	a_1	0.000307	0.000040	0.130293	(0.000225, 0.000400)
LMB	a_0	0.000283	0.000021	0.074205	(0.000241, 0.000332)
	a_1	0.000282	0.000020	0.070922	(0.000240, 0.000331)
LFW	a_0	0.000411	0.000042	0.102190	(0.000317, 0.000504)
	a_1	0.000391	0.000040	0.102302	(0.000307, 0.000489)
LFB	a_0	0.000294	0.000022	0.074830	(0.000247, 0.000342)
	a_1	0.000314	0.000023	0.073248	(0.000268, 0.000369)
MMW	a_0	0.000286	0.000017	0.059441	(0.000248, 0.000323)
	a_1	0.000264	0.000025	0.094697	(0.000209, 0.000327)
MMB	a_0	0.000262	0.000018	0.068702	(0.000223, 0.000304)
	a_1	0.000260	0.000018	0.069231	(0.000224, 0.000304)
MFW	a_0	0.000313	0.000023	0.073482	(0.000263, 0.000362)
	a_1	0.000247	0.000026	0.105263	(0.000194, 0.000312)
MFB	a_0	0.000289	0.000017	0.058824	(0.000253, 0.000332)
	a_1	0.000294	0.000022	0.074830	(0.000246, 0.000345)
HMW	a_0	0.000263	0.000016	0.060837	(0.000228, 0.000298)
	a_1	0.000192	0.000016	0.083333	(0.000157, 0.000234)
HMB	a_0	0.000250	0.000022	0.088000	(0.000201, 0.000299)
	a_1	0.000270	0.000027	0.100000	(0.000212, 0.000335)
HFW	a_0	0.000293	0.000021	0.071672	(0.000246, 0.000340)
	a_1	0.000324	0.000037	0.114198	(0.000248, 0.000410)
HFB	a_0	0.000278	0.000021	0.075540	(0.000234, 0.000324)
	a_1	0.000246	0.000024	0.097561	(0.000196, 0.000303)

Figure 4.1 The posterior densities of the finite population proportions under the three models for the SAL data with 12 areas.



NOTE: Here the solid line represents the densities of P under the IS model, the dotted line represents that under the HoS model, and the dashed line represents that under the HeS model.

Figure 4.2 Trace plot and Autocorrelation plot for μ, ρ_0, ρ_1.



We present the posterior densities of the finite population proportions in Figure 4.1. The solid line represents the density of P under the IS model, the dotted line represents that under the HoS model, and the dashed line represents that under the HeS model for each domain. We can see that the plots of IS model are shifted due to the effect of the selection bias.

Table 4.8
Geweke convergence diagnostic (Z-score) and effective sample size (ESS) for μ, ρ_0, ρ_1

parameter	Geweke's statistic	p -value	ESS
μ	-0.4549	0.6492	2,000
ρ_0	1.7750	0.0759	1,809
ρ_1	-1.0331	0.3015	2,000

Our example shows that it is important to include a component for the selection mechanism into a model when a biased sample is available, otherwise there is likely to be misleading estimates of the small area proportions. Because the distribution of the sample selection probabilities is a little different across domains, the heterogeneity assumption is more reasonable for our numerical example.

4.2 Simulation study

We perform a simulation study to assess the accuracy of our model. Specifically, we consider two situations, which are homogeneous or heterogeneous distributions of the sample selection probabilities to show how different the IS, HoS and HeS models can be. We also show that when the correlation between the binary responses and the selection probabilities is strong, the IS model can perform badly.

To perform a simulation study, we generate $\ell = 12$ finite populations with i^{th} finite population having N_i = 1,000 units and selection probability π_{ij}, $j = 1, \ldots,$ 1,000, $i = 1, \ldots, 12$. Then samples of size n_i = 50, $i = 1, \ldots, 12$ are taken from each area. We have generated 100 data sets. The outline of the data collection is as follows.

Step 1. Generate $p_i \sim U(0.2,\, 0.7)$, $i = 1, \ldots, \ell$. These values are true small population proportions.

Step 2. Set $\tau = 100$, $f_i = n_i / N_i$, $a =$ 0.975, $\mu_0 = af$, $\mu_1 = f/a$, $i = 1, \ldots, \ell$.

Step 3. Generate $u \sim U(0,\, 1)$. If $u \le p_i$ then we set $y_{ij} = 1$; otherwise set $y_{ij} = 0$, $j = 1, \ldots, N_i$, $i = 1, \ldots, \ell$.

Step 4. If $y_{ij} = 1$ then we generate $\pi_{ij} \sim \text{Beta}(\mu_1 \tau,\, (1 - \mu_1)\tau)$; if $y_{ij} = 0$ then we generate $\pi_{ij} \sim \text{Beta}(\mu_0 \tau,\, (1 - \mu_0)\tau)$ for $j = 1, \ldots, N_i$, $i = 1, \ldots, \ell$.

Step 5. Sample n_i units by systematic PPS sampling with probabilities $\frac{n_{ij}\pi_{ij}}{\sum_{i=1}^{N_i} \pi_{ij}}$.

We control the biserial correlation between the binary responses and the selection probabilities by changing the a values in Step 2. For example, if we set $a =$ 0.975, 0.95, 0.9, 0.8, 0.7 in the simulation scheme, then the biserial correlations are respectably $\rho =$ 0.05, 0.1, 0.2, 0.4, 0.7. We generate data having several biserial correlations, and categorized them as three levels (Low: $\rho <$ 0.3, Medium: 0.3 $\le \rho <$ 0.6, High: $\rho \ge$ 0.6). These correspond to weak, medium and strong selection bias.

We can also control the heterogeneity or homogeneity assumptions by changing a support of the sample selection probabilities. For homogeneity assumption, we generate π_{ij} of Step 4 from similar supports for all areas. On the other hand, we set up a support of π_{ij} from a different interval by area to make heterogeneous distributions of the sample selection probabilities.

To compare the performance of three models, we compute several frequentist measures. First, we calculate the finite population proportion $P_i^{(h)}$, the posterior mean $\mathrm{PM}_i^{(h)}$ and the posterior standard deviation $\mathrm{PSD}_i^{(h)}$, $i = 1, \ldots, \ell$, $h = 1, \ldots, 100$. Then we compute the absolute bias $\mathrm{AB}_i^{(h)} = \left| \mathrm{PM}_i^{(h)} - P_i^{(h)} \right|$ and the root mean squared error $\mathrm{RMSE}_i^{(h)} = \sqrt{\mathrm{PSD}_i^{(h)2} + \mathrm{AB}_i^{(h)2}}$, $i = 1, \ldots, \ell$, $h = 1, \ldots, 100$. Using these frequentist quantities we obtain $\mathrm{AB}_i = \frac{1}{100} \sum_{h=1}^{100} \mathrm{AB}_i^{(h)}$ and $\mathrm{RMSE}_i = \frac{1}{100} \sum_{h=1}^{100} \mathrm{RMSE}_i^{(h)}$. Also we compute the 95% highest posterior density (HPD) interval for each of the 100 simulated runs in each area. Then we look at the width $\left(W_i^{(h)}\right)$ and the HPD incidence $\left(I_i^{(h)}\right)$. Let $I_i^{(h)} = 1$ if the 95% HPD interval contains the true value P_i and let $I_i^{(h)} = 0$ otherwise. Then we calculate the coverage $C_i = \sum_{h=1}^{100} I_i^{(h)} / 100$ and $W_i = \sum_{h=1}^{100} W_i^{(h)} / 100$.

The biserial correlation (i.e., Pearson correlation coefficient) ρ between $\mathbf{y}$ and $\boldsymbol{\pi}$ is calculated since the variable $\mathbf{y}$ is binary (e.g., see Cox, 1974). The summaries based on AB, RMSE, coverage and width of the 95% HPD intervals for three correlation cases under the homogeneity assumption are shown in Table 4.9. Under this assumption, the performances of the HoS model is better than the HeS model in some domains, but the difference is very small. Both of them are better than the IS model in the sense of the closeness to the true P_i. In particular, we can see that as the correlation between $\mathbf{y}$ and $\boldsymbol{\pi}$ increases, there is greater discrepancy between the IS model and the others.

Table 4.10 shows the summaries for three correlation cases under the heterogeneity assumption. From this table, we find that the HeS model is well behaved in the sense that it is closer to true P_i than the other models when the effect of the selection bias is moderate to strong. It has smaller bias, smaller mean square error and better coverage.

As the biserial correlation increases there are increased disparities between the IS model and the HeS model. We present some plots of the posterior distributions of the finite population proportions for low, medium, high correlation cases under the homogeneity assumption (Figures 4.3-4.5). It is worth noting that there are large differences of the distributions of the finite population means as ρ increases. There are no differences in the posterior densities of the HoS model and the HeS model under the homogeneity assumption. The posterior distributions under the HeS model departs from the right of the distribution under the IS model, with little changes in their spreads for all domains.

Similarly, we present some plots of the posterior distributions of the finite population proportions for low, medium, high correlation cases under the heterogeneity assumption (Figures 4.6-4.8). The posterior densities under the HeS model are different from the others for all domains.

Also, we have selected two values of $\ell = 12, 24,$ the number of areas, to see changes by increasing ℓ. We compute AB, RMSE, coverage and width of 95% HPD intervals for each domain, then we average these values over all domains. We present the results under the homogeneity and the heterogeneity assumptions in respectively Tables 4.11 and 4.12. The results are consistent as ℓ increases.

The simulation study shows that as the biserial correlation between the binary responses and the selection probabilities increases, there is greater discrepancy among the IS model, the HoS moel and the HeS model. That is, as the correlation is stronger, the effect of the selection bias becomes larger. The HeS model is better than the other models when the each area has the different distribution for the sample selection probabilities under moderate to strong selection bias.

Table 4.9
Comparisons of estimates based on absolute bias (AB), root posterior mean squared error (RMSE), coverage (C) and width (W) of 95% HPD intervals for low (L), medium (M), high (H) correlation cases under the homogeneity assumption

area	ρ	AB			RMSE			W			C		
		IS	HoS	HeS	IS	HoS	HeS	IS	HoS	HeS	IS	HoS	HeS
1	L	0.061	0.044	0.051	0.087	0.076	0.087	0.241	0.248	0.281	0.90	0.98	0.98
	M	0.134	0.046	0.057	0.149	0.075	0.092	0.259	0.238	0.293	0.50	0.98	0.97
	H	0.188	0.047	0.050	0.198	0.075	0.089	0.268	0.239	0.296	0.16	0.99	0.99
2	L	0.050	0.053	0.051	0.081	0.087	0.093	0.257	0.277	0.315	0.98	0.98	0.98
	M	0.103	0.043	0.042	0.122	0.081	0.091	0.260	0.279	0.331	0.72	0.96	1.0
	H	0.166	0.045	0.051	0.177	0.084	0.100	0.254	0.288	0.350	0.24	0.99	1.0
3	L	0.051	0.065	0.063	0.083	0.098	0.102	0.262	0.286	0.320	0.97	0.90	0.98
	M	0.084	0.064	0.058	0.107	0.099	0.103	0.258	0.295	0.342	0.81	0.95	0.98
	H	0.132	0.056	0.059	0.146	0.094	0.108	0.250	0.304	0.361	0.45	0.97	0.97
4	L	0.058	0.084	0.085	0.091	0.114	0.120	0.269	0.297	0.333	0.91	0.88	0.89
	M	0.052	0.078	0.070	0.083	0.110	0.113	0.258	0.311	0.355	0.97	0.93	0.98
	H	0.112	0.066	0.072	0.128	0.105	0.118	0.243	0.325	0.372	0.57	0.97	0.98
5	L	0.039	0.037	0.041	0.073	0.074	0.083	0.250	0.263	0.296	0.99	0.99	1.0
	M	0.118	0.037	0.043	0.135	0.074	0.088	0.261	0.261	0.315	0.64	0.98	1.0
	H	0.193	0.050	0.055	0.202	0.083	0.100	0.261	0.269	0.336	0.14	0.99	0.99
6	L	0.052	0.040	0.046	0.081	0.075	0.085	0.245	0.256	0.289	0.96	0.96	0.99
	M	0.125	0.041	0.049	0.142	0.073	0.089	0.260	0.249	0.301	0.52	0.98	0.98
	H	0.186	0.048	0.056	0.197	0.078	0.096	0.263	0.253	0.316	0.18	0.99	0.98
7	L	0.041	0.053	0.058	0.076	0.088	0.098	0.261	0.281	0.316	0.97	0.95	0.96
	M	0.096	0.049	0.045	0.116	0.087	0.095	0.259	0.289	0.340	0.75	0.94	0.99
	H	0.145	0.046	0.050	0.158	0.086	0.102	0.254	0.293	0.356	0.36	0.96	0.98
8	L	0.044	0.039	0.045	0.076	0.075	0.085	0.249	0.261	0.295	0.99	0.99	1.0
	M	0.118	0.039	0.041	0.135	0.074	0.086	0.261	0.258	0.309	0.62	0.99	0.99
	H	0.184	0.049	0.058	0.195	0.081	0.101	0.262	0.262	0.329	0.24	0.98	0.97
9	L	0.047	0.043	0.046	0.077	0.078	0.088	0.253	0.268	0.301	0.98	0.98	0.99
	M	0.116	0.041	0.042	0.133	0.076	0.089	0.260	0.265	0.322	0.64	1.0	1.0
	H	0.184	0.050	0.059	0.194	0.084	0.103	0.258	0.274	0.338	0.21	0.98	0.97
10	L	0.062	0.046	0.059	0.088	0.078	0.095	0.246	0.257	0.296	0.94	0.98	0.97
	M	0.135	0.041	0.050	0.149	0.073	0.090	0.260	0.245	0.300	0.48	0.99	0.99
	H	0.191	0.055	0.058	0.202	0.082	0.097	0.264	0.247	0.310	0.17	0.99	1.0
11	L	0.053	0.070	0.074	0.085	0.103	0.112	0.265	0.292	0.327	0.96	0.92	0.95
	M	0.070	0.065	0.053	0.096	0.100	0.101	0.258	0.303	0.348	0.88	0.95	0.99
	H	0.131	0.058	0.058	0.145	0.098	0.109	0.245	0.313	0.368	0.46	0.94	0.97
12	L	0.047	0.043	0.043	0.077	0.077	0.086	0.250	0.264	0.306	0.97	0.99	0.99
	M	0.122	0.047	0.048	0.139	0.080	0.091	0.260	0.263	0.317	0.56	0.97	0.99
	H	0.173	0.044	0.050	0.184	0.078	0.095	0.261	0.266	0.333	0.27	0.99	0.99

Table 4.10
Comparisons of estimates based on absolute bias (AB), root posterior mean squared error (RMSE), coverage (C) and width (W) of 95% HPD intervals for low (L), medium (M), high (H) correlation cases under the heterogeneity assumption

area	ρ	AB			RMSE			W			C		
		IS	HoS	HeS	IS	HoS	HeS	IS	HoS	HeS	IS	HoS	HeS
1	L	0.050	0.052	0.047	0.078	0.085	0.084	0.215	0.235	0.251	0.91	0.92	0.94
	M	0.106	0.101	0.041	0.125	0.123	0.084	0.229	0.248	0.268	0.55	0.67	0.99
	H	0.163	0.145	0.056	0.176	0.162	0.109	0.242	0.261	0.343	0.16	0.39	0.99
2	L	0.045	0.047	0.050	0.079	0.085	0.094	0.231	0.252	0.287	0.95	0.97	0.99
	M	0.087	0.092	0.049	0.109	0.116	0.105	0.231	0.250	0.345	0.75	0.78	1.0
	H	0.161	0.174	0.090	0.173	0.187	0.145	0.221	0.235	0.382	0.26	0.23	0.89
3	L	0.056	0.058	0.068	0.088	0.094	0.108	0.235	0.259	0.287	0.91	0.92	0.91
	M	0.085	0.080	0.062	0.111	0.111	0.114	0.239	0.264	0.345	0.77	0.86	0.97
	H	0.151	0.133	0.060	0.165	0.152	0.129	0.233	0.261	0.405	0.32	0.55	0.97
4	L	0.056	0.059	0.056	0.088	0.095	0.099	0.240	0.263	0.290	0.92	0.93	0.96
	M	0.117	0.123	0.053	0.135	0.144	0.108	0.242	0.265	0.341	0.57	0.59	0.99
	H	0.175	0.194	0.081	0.188	0.207	0.140	0.245	0.267	0.404	0.23	0.20	0.96
5	L	0.047	0.049	0.047	0.078	0.084	0.089	0.224	0.245	0.273	0.98	0.97	0.99
	M	0.105	0.098	0.052	0.126	0.124	0.102	0.239	0.263	0.313	0.60	0.71	0.98
	H	0.164	0.144	0.071	0.178	0.163	0.130	0.246	0.271	0.399	0.28	0.46	1.0
6	L	0.045	0.051	0.059	0.076	0.084	0.097	0.218	0.238	0.273	0.95	0.95	0.95
	M	0.060	0.063	0.058	0.086	0.091	0.109	0.211	0.227	0.331	0.87	0.90	0.97
	H	0.103	0.113	0.055	0.118	0.128	0.112	0.195	0.203	0.353	0.51	0.53	0.99
7	L	0.043	0.046	0.056	0.079	0.085	0.098	0.234	0.257	0.284	0.97	0.98	0.97
	M	0.093	0.086	0.051	0.117	0.116	0.107	0.241	0.265	0.337	0.73	0.81	0.98
	H	0.152	0.134	0.065	0.168	0.155	0.129	0.240	0.266	0.400	0.31	0.53	0.98
8	L	0.045	0.048	0.060	0.077	0.083	0.098	0.223	0.244	0.277	0.96	0.99	0.96
	M	0.066	0.069	0.053	0.092	0.097	0.109	0.217	0.235	0.338	0.87	0.90	0.98
	H	0.103	0.115	0.064	0.121	0.132	0.123	0.207	0.217	0.375	0.56	0.48	0.99
9	L	0.048	0.050	0.045	0.080	0.086	0.088	0.227	0.249	0.275	0.98	0.97	0.99
	M	0.103	0.096	0.052	0.125	0.123	0.101	0.241	0.264	0.312	0.63	0.70	0.99
	H	0.166	0.146	0.057	0.180	0.166	0.120	0.245	0.272	0.382	0.24	0.42	0.99
10	L	0.046	0.047	0.057	0.076	0.080	0.095	0.214	0.235	0.272	0.92	0.96	0.95
	M	0.051	0.053	0.062	0.078	0.082	0.110	0.207	0.222	0.327	0.93	0.94	0.97
	H	0.093	0.102	0.054	0.108	0.117	0.110	0.191	0.198	0.348	0.61	0.60	1.0
11	L	0.057	0.058	0.071	0.088	0.094	0.110	0.237	0.260	0.293	0.91	0.95	0.92
	M	0.075	0.068	0.059	0.102	0.102	0.112	0.238	0.260	0.344	0.79	0.86	0.99
	H	0.136	0.120	0.067	0.152	0.142	0.128	0.229	0.257	0.398	0.43	0.58	0.99
12	L	0.044	0.046	0.051	0.076	0.082	0.092	0.222	0.243	0.277	0.98	0.96	1.0
	M	0.072	0.076	0.050	0.097	0.103	0.105	0.219	0.236	0.337	0.79	0.82	1.0
	H	0.115	0.127	0.064	0.131	0.143	0.123	0.211	0.222	0.369	0.51	0.45	0.98

Table 4.11
Comparison of the three models using absolute bias (AB), root posterior mean squared error (RMSE), coverage (C) and width (W) of 95% HPD intervals under the homogeneity assumption

ρ	ℓ	Model	AB	RMSE	W	C
Low	12	IS	$0.0503_{0.0075}$	$0.0812_{0.0056}$	$0.2539_{0.0086}$	$0.9600_{0.0292}$
		HoS	$0.0514_{0.0145}$	$0.0853_{0.0131}$	$0.2709_{0.0155}$	$0.9583_{0.0381}$
		HeS	$0.0551_{0.0133}$	$0.0943_{0.0117}$	$0.3062_{0.0161}$	$0.9733_{0.0303}$
	24	IS	$0.0486_{0.0040}$	$0.0809_{0.0029}$	$0.2594_{0.0123}$	$0.9721_{0.0177}$
		HoS	$0.0509_{0.0101}$	$0.0861_{0.0103}$	$0.2778_{0.0215}$	$0.9713_{0.0201}$
		HeS	$0.0568_{0.0105}$	$0.0975_{0.0101}$	$0.3166_{0.0210}$	$0.9792_{0.0177}$
Medium	12	IS	$0.1060_{0.0261}$	$0.1254_{0.0212}$	$0.2594_{0.0011}$	$0.6742_{0.1556}$
		HoS	$0.0491_{0.0127}$	$0.0835_{0.0126}$	$0.2714_{0.0237}$	$0.9683_{0.0221}$
		HeS	$0.0498_{0.0086}$	$0.0939_{0.0078}$	$0.3227_{0.0203}$	$0.9883_{0.0094}$
	24	IS	$0.1010_{0.0232}$	$0.1212_{0.0198}$	$0.2592_{0.0104}$	$0.7025_{0.1267}$
		HoS	$0.0502_{0.0092}$	$0.0861_{0.0110}$	$0.2812_{0.0295}$	$0.9775_{0.0159}$
		HeS	$0.0521_{0.0080}$	$0.0964_{0.0093}$	$0.3279_{0.0281}$	$0.9904_{0.0127}$
High	12	IS	$0.1653_{0.0281}$	$0.1771_{0.0263}$	$0.2568_{0.0078}$	$0.2875_{0.1399}$
		HoS	$0.0510_{0.0064}$	$0.0857_{0.0088}$	$0.2776_{0.0270}$	$0.9783_{0.0159}$
		HeS	$0.0562_{0.0062}$	$0.1014_{0.0075}$	$0.3389_{0.0237}$	$0.9825_{0.0114}$
	24	IS	$0.1582_{0.0316}$	$0.1704_{0.0300}$	$0.2540_{0.0170}$	$0.3242_{0.1618}$
		HoS	$0.0522_{0.0046}$	$0.0886_{0.0090}$	$0.2882_{0.0335}$	$0.9788_{0.0099}$
		HeS	$0.0570_{0.0039}$	$0.1025_{0.0069}$	$0.3412_{0.0302}$	$0.9875_{0.0126}$

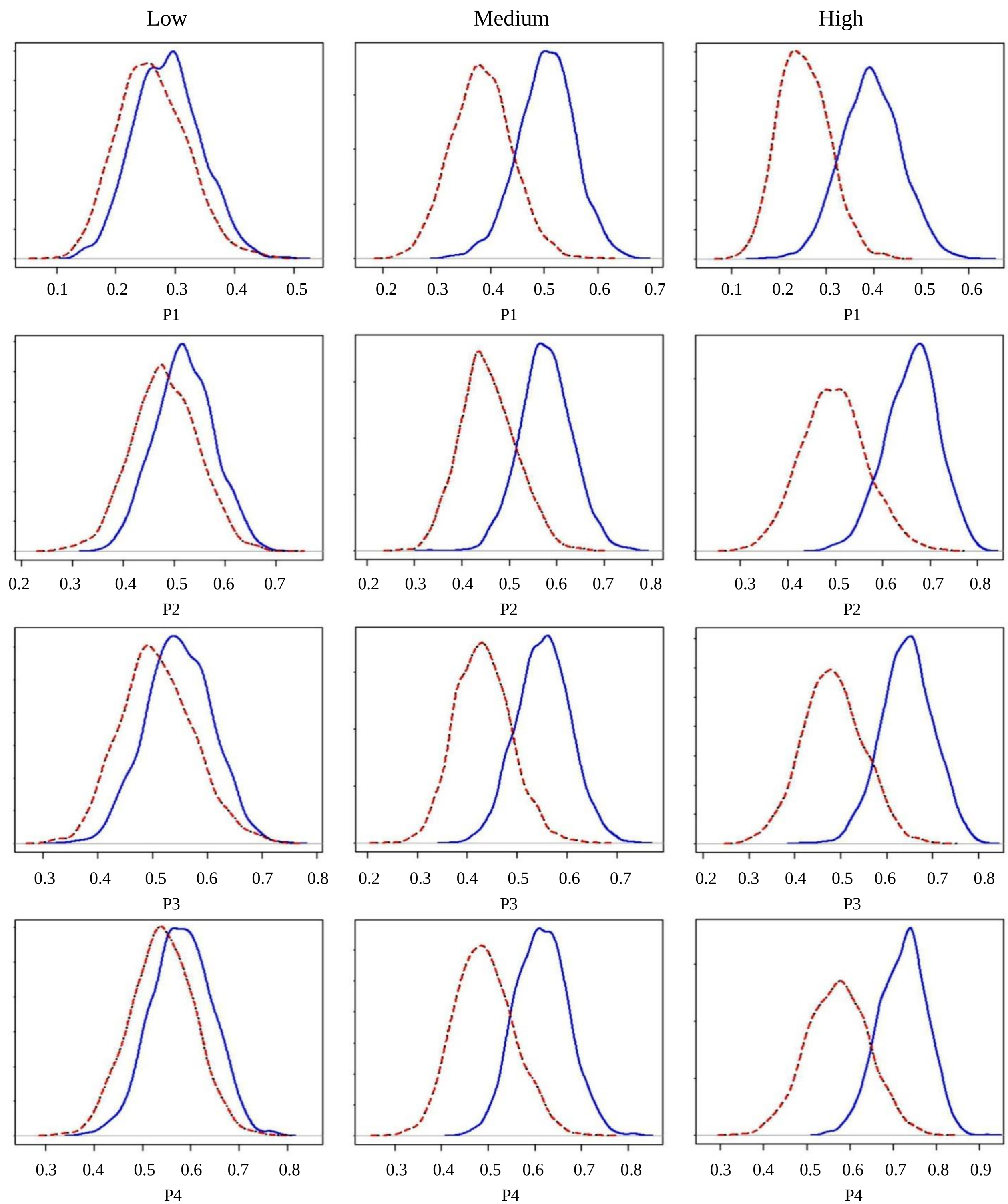
NOTE: The notation a_b means a is an average of areas and b is the standard error.

Table 4.12
Comparison of the three models using absolute bias (AB), root posterior mean squared error (RMSE), coverage (C) and width (W) of 95% HPD intervals under the heterogeneity assumption

ρ	ℓ	Model	AB	RMSE	W	C
Low	12	IS	$0.0501_{0.0050}$	$0.0842_{0.0051}$	$0.2703_{0.0109}$	$0.9733_{0.0156}$
		HoS	$0.0509_{0.0049}$	$0.0862_{0.0049}$	$0.2800_{0.0112}$	$0.9758_{0.0168}$
		HeS	$0.0556_{0.0083}$	$0.0959_{0.0075}$	$0.3132_{0.0127}$	$0.9767_{0.0227}$
	24	IS	$0.0457_{0.0054}$	$0.0768_{0.0054}$	$0.2478_{0.0135}$	$0.9675_{0.0205}$
		HoS	$0.0470_{0.0055}$	$0.0818_{0.0053}$	$0.2703_{0.0146}$	$0.9779_{0.0169}$
		HeS	$0.0537_{0.0093}$	$0.0922_{0.0082}$	$0.3018_{0.0167}$	$0.9792_{0.0193}$
Medium	12	IS	$0.0850_{0.0206}$	$0.1085_{0.0179}$	$0.2586_{0.0147}$	$0.8033_{0.1026}$
		HoS	$0.0837_{0.0196}$	$0.1109_{0.0170}$	$0.2816_{0.0187}$	$0.8517_{0.0720}$
		HeS	$0.0534_{0.0060}$	$0.1054_{0.0077}$	$0.3672_{0.0248}$	$0.9917_{0.0111}$
	24	IS	$0.0824_{0.0218}$	$0.1056_{0.0190}$	$0.2531_{0.0179}$	$0.7992_{0.1040}$
		HoS	$0.0810_{0.0205}$	$0.1074_{0.0179}$	$0.2736_{0.0218}$	$0.8546_{0.0825}$
		HeS	$0.0533_{0.0077}$	$0.1046_{0.0090}$	$0.3637_{0.0289}$	$0.9954_{0.0093}$
High	12	IS	$0.1401_{0.0289}$	$0.1547_{0.0278}$	$0.2537_{0.0232}$	$0.4417_{0.1333}$
		HoS	$0.1372_{0.0263}$	$0.1543_{0.0253}$	$0.2741_{0.0319}$	$0.5375_{0.1231}$
		HeS	$0.0653_{0.0110}$	$0.1249_{0.0113}$	$0.4215_{0.0257}$	$0.9833_{0.0287}$
	24	IS	$0.1336_{0.0302}$	$0.1481_{0.0289}$	$0.2487_{0.0263}$	$0.4746_{0.1316}$
		HoS	$0.1302_{0.0270}$	$0.1470_{0.0261}$	$0.2663_{0.0348}$	$0.5567_{0.1396}$
		HeS	$0.0601_{0.0090}$	$0.1203_{0.0108}$	$0.4141_{0.0360}$	$0.9900_{0.0153}$

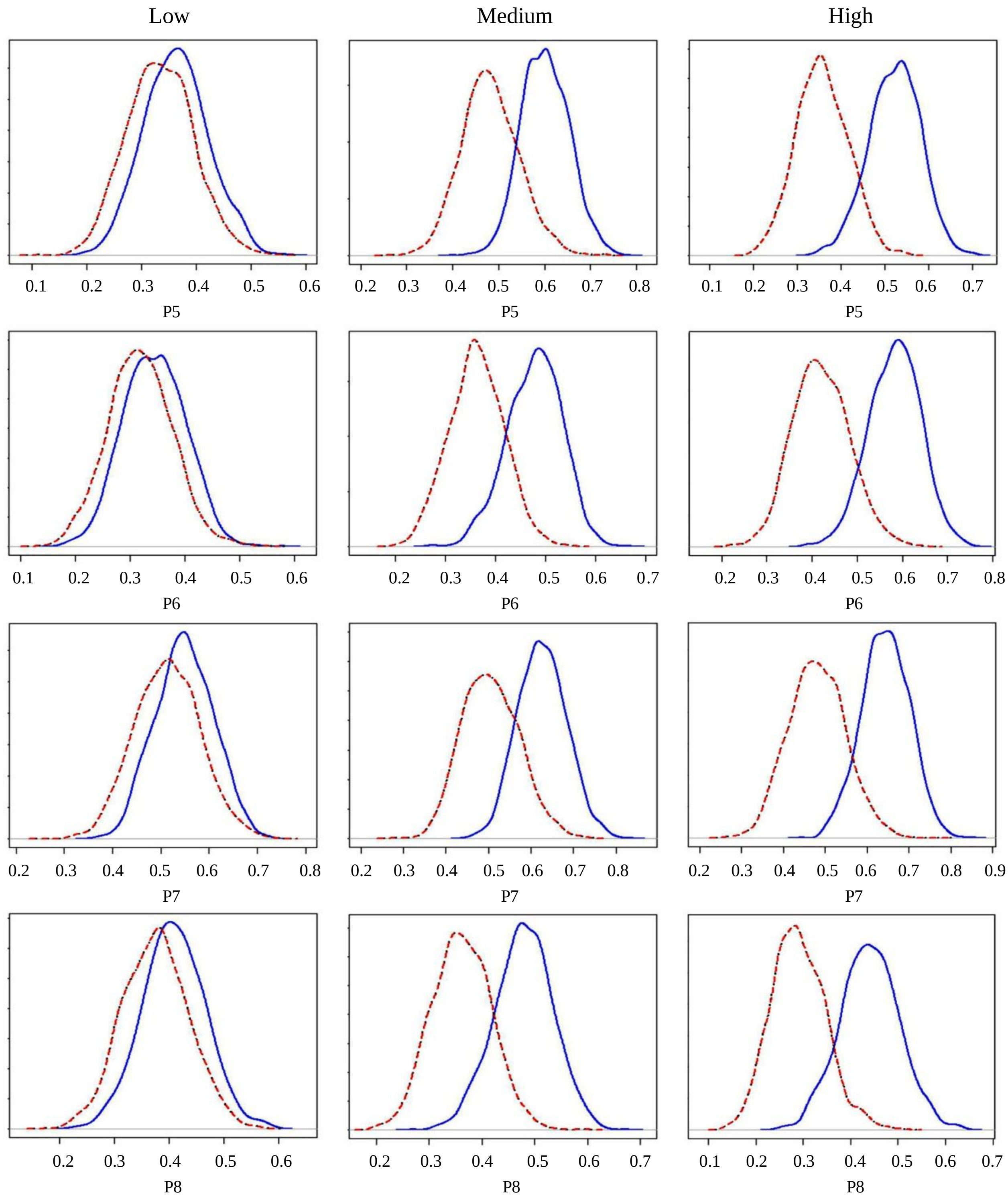
NOTE: The notation a_b means a is an average of areas and b is the standard error.

Figure 4.3 The posterior densities of the finite population proportions for low, medium, high correlation cases using a simulated data under the homogeneity assumption.



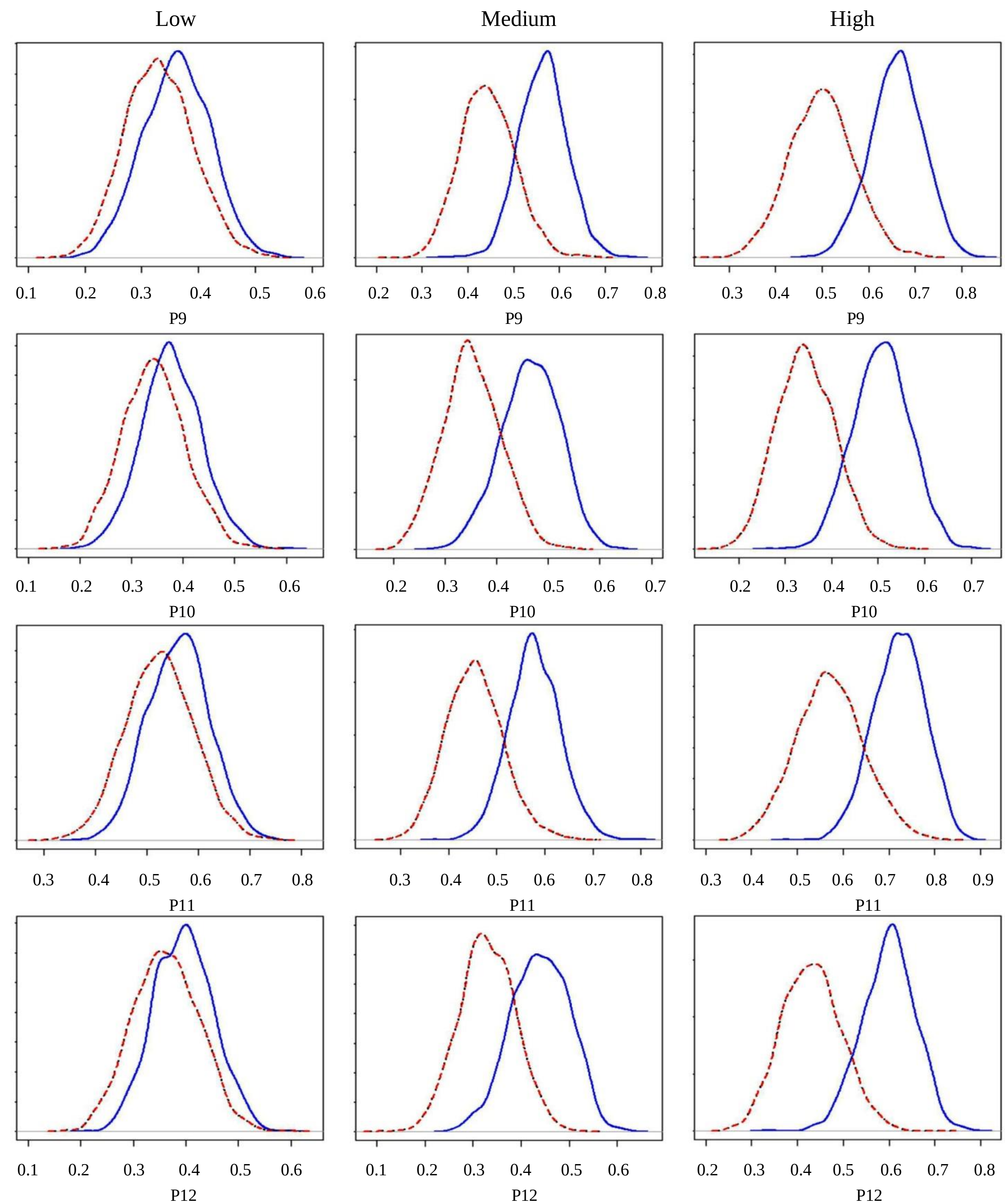
NOTE: Here the solid line represents the densities of P under the IS model, the dotted line represents that under the HoS model, and the dashed line represents that under the HeS model. (Dashed line and dotted line coincide.)

Figure 4.4 The posterior densities of the finite population proportions for low, medium, high correlation cases using a simulated data under the homogeneity assumption (continued).



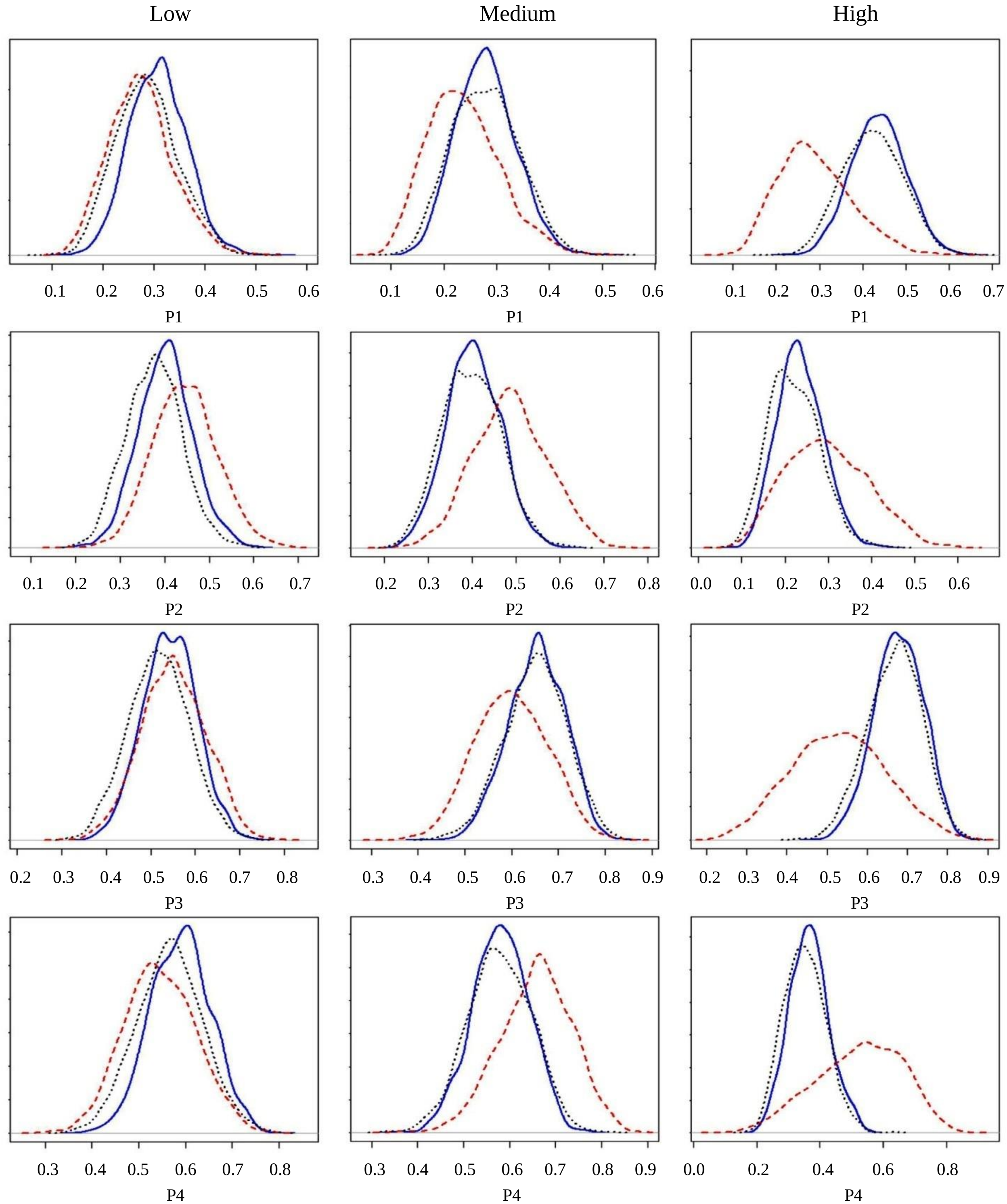
NOTE: Here the solid line represents the densities of P under the IS model, the dotted line represents that under the HoS model, and the dashed line represents that under the HeS model. (Dashed line and dotted line coincide.)

Figure 4.5 The posterior densities of the finite population proportions for low, medium, high correlation cases using a simulated data under the homogeneity assumption (continued).



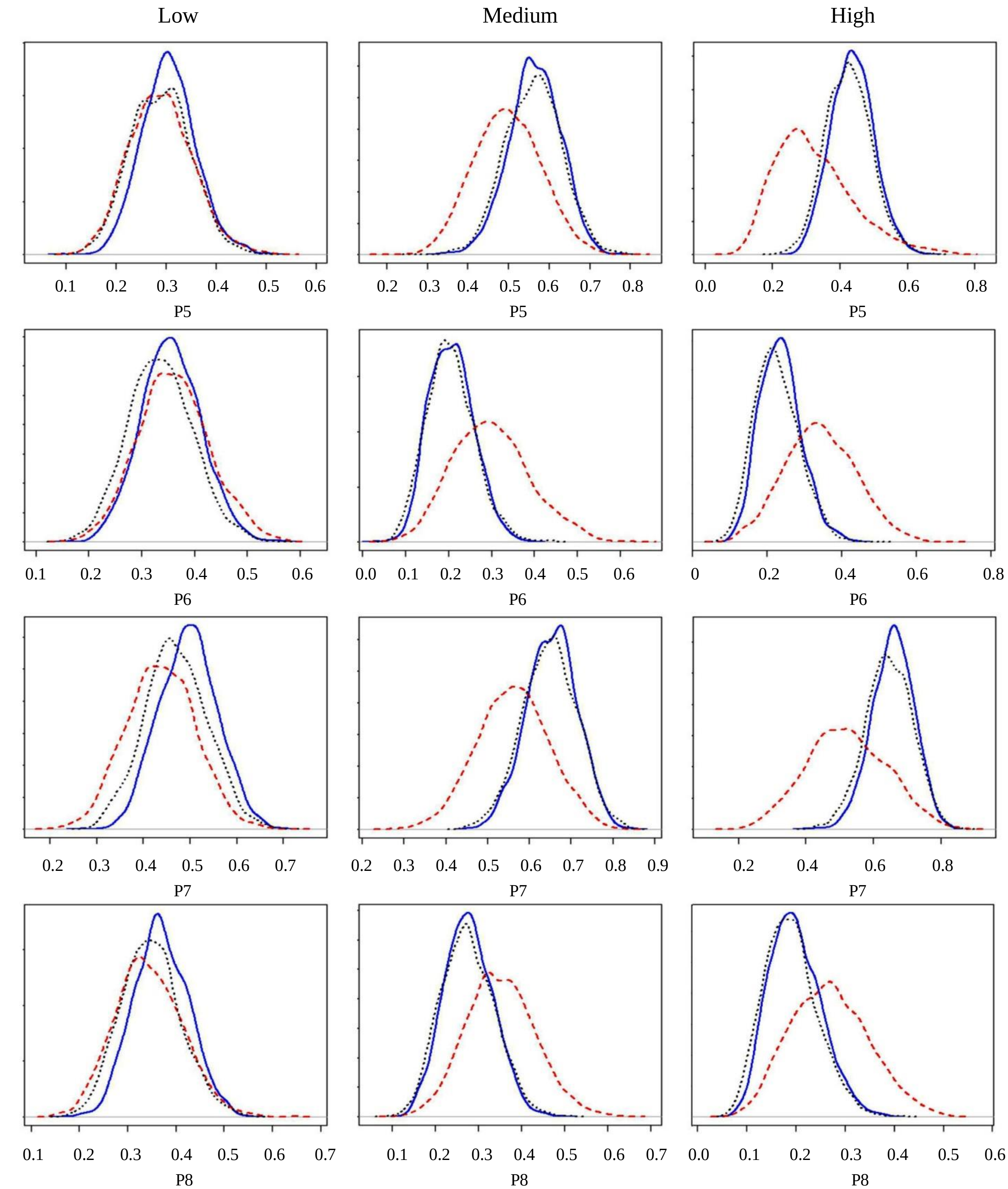
NOTE: Here the solid line represents the densities of P under the IS model, the dotted line represents that under the HoS model, and the dashed line represents that under the HeS model. (Dashed line and dotted line coincide.)

Figure 4.6 The posterior densities of the finite population proportions for low, medium, high correlation cases using a simulated data under the heterogeneity assumption.



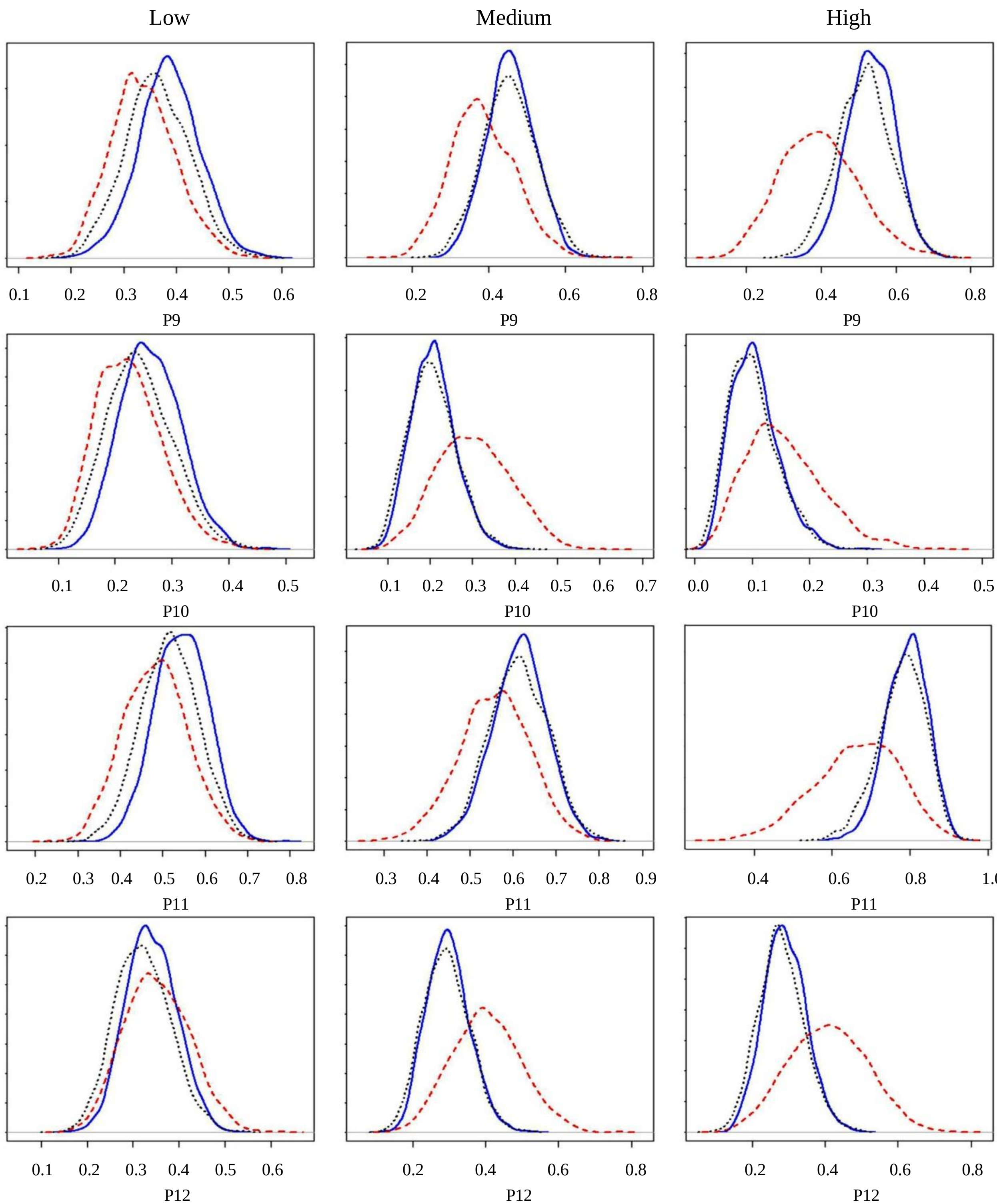
NOTE: Here the solid line represents the densities of P under the IS model, the dotted line represents that under the HoS model, and the dashed line represents that under the HeS model.

Figure 4.7 The posterior densities of the finite population proportions for low, medium, high correlation cases using a simulated data under the heterogeneity assumption (continued).



NOTE: Here the solid line represents the densities of P under the IS model, the dotted line represents that under the HoS model, and the dashed line represents that under the HeS model.

Figure 4.8 The posterior densities of the finite population proportions for low, medium, high correlation cases using a simulated data under the heterogeneity assumption (continued).



NOTE: Here the solid line represents the densities of P under the IS model, the dotted line represents that under the HoS model, and the dashed line represents that under the HeS model.

5. Concluding remarks

We have extended the homogeneous nonignorable selection model of CNK to accommodate selection probabilities that have different distributions in different areas. This is an improvement to handle the relationship between the binary variables and the selection probabilities. As there are numerous additional parameters, the computation has become a lot more difficult and we have used the Metropolis-Hastings sampler to overcome this difficulty.

We have used an example on severe activity limitation in the National Health Interview Survey in which the small areas are formed by crossing education (pre-college, college, post-college), sex (male, female) and race (white, nonwhite). The heterogeneous nonignorable selection model appears to perform better than the homogeneous selection model and the baseline ignorable selection model.

We have used a simulation study to assess the performance of our heterogeneous nonignorable selection model. We have drawn data from the homogeneous nonignorable selection model and fitted the ignorable selection model and the heterogeneous nonignorable selection model. We found little difference between the two nonignorable selection models but substantial difference from the ignorable selection model. However, when we have drawn data from the heterogeneous nonignorable selection model and fitted the ignorable selection model and the homogeneous nonignorable selection model, we found that the heterogeneous nonignorable selection model is a lot better especially when there is medium to strong selection bias using bias, mean square error and coverage. This is evident in Tables 4.9-4.12 and Figures 4.3-4.8.

Within the framework of our heterogeneous nonignorable selection model, we can think about several additional problems. First, we can accommodate polychotomous data that are numerous in survey problems. Second, although a bit different from our approach, we can accommodate covariates (e.g., age, race, sex in our application on activity limitation). Third, as there are nonrespondents in numerous surveys, we can attempt to accommodate nonignorable nonresponse and selection simultaneously (e.g., Nandram and Choi, 2010). This can be done within our framework. Fourth, as in a two-fold model, we can accommodate clustering or stratification within each area (e.g., Nandram, 2016; Lee, Nandram and Kim, 2017). Our work is potentially useful to solve problems in nonprobability samples as well (e.g., Elliot and Valliant, 2017).

Acknowledgements

This research was supported by a grant from the Simons Foundation (#353953, Balgobin Nandram).

References

Chambers, R., Dorfman, A. and Wang, S. (1998). Limited information likelihood analysis of survey data. *Journal of the Royal Statistical Society, Series B*, 60, 397-411.

Chen, Q., Elliott, M.R. and Little, R.J.A. (2010). Bayesian penalized spline model-based inference for finite population proportion in unequal probability sampling. *Survey Methodology*, 36, 1, 23-34. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2010001/article/11250-eng.pdf.

Choi, S., Nandram, B. and Kim, D. (2017). A hierarchical Bayesian model for binary data incorporating selection bias. *Communications in Statistics - Simulation and Computation*, 46, 6, 4767-4782.

Cox, N.R. (1974). Estimation of the correlation between a continuous and a discrete variable. *Biometrics*, 30, 171-178.

Elliot, M.R., and Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, 32, 249-264.

Gelman, A. (2006). Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515-533.

Lee, D., Nandram, B. and Kim, D. (2017). Bayesian predictive inference of a proportion under a two-fold small area model with heterogeneous correlations. *Survey Methodology*, 43, 1, 69-92. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2017001/article/14822-eng.pdf.

Ma, J., Sedransk, J., Nandram, B. and Chen, L. (2018). Bayesian predictive inference for finite population quantities under informative sampling. *Statistics and Application*, 16, 207-226.

Malec, D., Davis, W. and Cao, X. (1999). Model-based small area estimates of overweight prevalence using sample selection adjustment. *Statistics in Medicine*, 18, 3189-3200.

Nandram, B. (2007). Bayesian predictive inference under informative sampling via surrogate samples. *Bayesian Statistics and its Applications*, (Eds. S.K. Upadhyay, U. Singh and D. Dey), Anamaya, New Delhi, Chapter 25, 356-374.

Nandram, B. (2016). Bayesian predictive inference of a proportion under a two-fold small area model. *Journal of Official Statistics*, 32, 1, 187-208.

Nandram, B., and Choi, J.W. (2010). A Bayesian analysis of body mass index data from small domains under nonignorable nonresponse and selection. *Journal of the American Statistical Association*, 105, 120-135.

Nandram, B., Bai, Y. and Choi, J.W. (2011). Hierarchical Bayesian models for assessing possible changes in prevalence of activity limitation. *Advances and Applications in Statistical Sciences*, 6, 285-311.

Nandram, B., Bhatta, D., Bhadra, D. and Shen G. (2013). Bayesian predictive inference of a finite population proportion under selection bias. *Statistical Methodology*, 11, 1-21.

Nandram, B., Choi, J.W., Shen, G. and Burgos, C. (2006). Bayesian predictive inference under informative sampling and transformation. *Applied Stochastic Models in Business and Industry*, 22, 559-572.

Opsomer, J.D., Glaeskens, G., Ranalli, M.G., Kauermann, G. and Breidt, F.J. (2008). Non-parametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society, Series B*, 70, 265-286.

Pfeffermann, D. (1988). The effect of sampling design and response mechanism on multivariate regression-based predictors. *Journal of the American Statistical Association*, 83, 824-833.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61, 317-337.

Pfeffermann, D., and Sverchkov, M. (2007). Small-area estimation under informative probability sampling of areas and within selected areas. *Journal of the American Statistical Association*, 102, 1427-1439.

Pfeffermann, D., Krieger, A.M. and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, 8, 1087-1114.

Si, Y., Pillai, N.S. and Gelman, A. (2015). Bayesian nonparametric weighted sampling inference. *Bayesian Analysis*, 10, 3, 605-625.

Sverchkov, M., and Pfeffermann, D. (2004). Prediction of finite population totals based on the sample distribution. *Survey Methodology*, 30, 1, 79-92. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2004001/article/6996-eng.pdf.

Zangeneh, S.Z., and Little, R.J.A. (2015). Bayesian inference for the finite population total from a heteroscedastic probability proportional to size sample. *Journal of Survey Statistics and Methodology*, 3, 162-192.