A Model-Based Approach to Weight Trimming

Michael R. Elliott, Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics and Epidemiology, University of Pennsylvania Medical Center, 612 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104.

1.0 Introduction

Analysis of data from samples with differential probabilities of inclusion typically use case weights equal to the inverse of the probability of inclusion to reduce or remove bias in the estimators of population quantities of interest. An example is the Horvitz-Thompson estimator (Horvitz and Thompson 1952) of a population mean $\overline{Y} = N^{-1}\sum_{i=1}^{N} y_i$ given by $\hat{\overline{Y}} = N^{-1}\sum_{i\in s} w_i y_i$, or more generally by

$$\hat{\overline{Y}} = \hat{N}^{-1}\sum_{i\in s} w_i y_i \tag{1}$$

where $w_i = 1/\pi_i, \pi_i$ is the probability of selection, s is the subset of the population units sampled, and $\hat{N} = \sum_{i\in s} w_i$. These fully-weighted estimators are unbiased for their population values for linear estimators such as means and totals. Bias can be reduced and consistent estimates of population values obtained for a wide class of non-linear estimators such as linear or generalized linear regression slopes that are implicit functions of linear statistics by replacing, e.g., implicit means or totals with their weighted equivalents (Binder 1983).

There is little debate that sampling weights be utilized when considering descriptive statistics such as means and totals obtained from disproportional probability-of-selection designs. However, when estimating "analytical" models (Cochran 1977, p.4) that focus on associations between, e.g., risk factors and health outcomes estimated via linear and generalized linear models, the decision to use sampling weights is less definitive (c.f. Korn and Gaubard, 1999, p.180-182). To make the issues concrete, consider a population generated from

$$Y_i \mid X_i \sim N(\mu(X_i), \sigma^2) \tag{2}$$

where $\mu(X_i) = \alpha + \beta X_i + \gamma X_i^2$, but the superpopulation model of interest assumes $\mu(X_i) = \alpha + \beta X_i$; hence, the superpopulation model is correctly specified when $\gamma = 0$ and misspecified when $\gamma \neq 0$. We assume that the goal of the modeler is to describe the association between Y and X using the regression slope β from the superpopulation model. If the superpopulation model is correctly specified, the target quantity of interest could be either the superpopulation slope β, or the population slope defined by $B = \sum_{i=1}^{N} A_i(Y_i - \overline{Y})$, where $A_i = (X_i - \overline{X})/\sum_{i=1}^{N}(X_i - \overline{X})^2$, $\overline{Y} = N^{-1}\sum_{i=1}^{N} Y_i$, and $\overline{X} = N^{-1}\sum_{i=1}^{N} X_i$ (the "corresponding descriptive population quantity" in Pfeffermann [1993]). If the superpopulation model is misspecified, then only the population slope makes sense as a target quantity, in the sense of being a linear approximation that summarizes the relationship between Y and X. The unweighted OLS estimator and case-weighted WLS estimator of the population slope B are given by

$$\hat{B} = \frac{n\sum_{i\in s} x_i y_i - \sum_{i\in s} x_i \sum_{i\in s} y_i}{n\sum_{i\in s} x_i^2 - (\sum_{i\in s} x_i)^2} \tag{3}$$

$$\hat{B}^w = \frac{\sum_{i\in s} w_i \sum_{i\in s} w_i x_i y_i - \sum_{i\in s} w_i x_i \sum_{i\in s} w_i y_i}{\sum_{i\in s} w_i \sum_{i\in s} w_i x_i^2 - (\sum_{i\in s} w_i x_i)^2}.$$

The unweighted estimator $\hat{B}$ is consistent for B if and only if the superpopulation model is correctly specified ($\gamma = 0$) and sampling ignorable. The weighted estimator $\hat{B}^w$ is consistent for B even if the superpopulation model misspecified ($\gamma \neq 0$) and sampling is non-ignorable (sampling indicator S independent of Y given X); however, this bias reduction is sometimes purchased at the cost of high variance.

The fully-weighted estimator of the population slope $\hat{B}^w$ and the fully-weighted counterpart estimator of the intercept $\hat{A}^w = \frac{\sum_{i\in s} w_i y_i}{\sum_{i\in s} w_i} - \hat{B}^w \frac{\sum_{i\in s} w_i x_i}{\sum_{i\in s} w_i}$ are sometimes termed "pseudo-maximum likelihood" estimators (PMLEs) (Binder 1983, Pfeffermann 1993) because they are "design consistent" for the MLEs that would solve the score equations under the superpopulation model that assumes $\mu(X_i) = \alpha + \beta X_i$, if we had observed data for the entire population:

$$\boldsymbol{U}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{d}{d\boldsymbol{\theta}} \log f(y_i \mid \boldsymbol{\theta}) = \sum_{i=1}^{N} \boldsymbol{X}_i(\boldsymbol{X}_i^T\boldsymbol{\theta} - y_i) = 0 \tag{4}$$

for $\boldsymbol{\theta} = (\alpha\ \beta)^T$. Use of PMLEs in samples with unequal probabilities of selection underlies use of sampling weights in a wide variety of settings where independence across observations can be assumed, including generalized linear regression (Binder 1983) and latent class analysis (Patterson et al. 2002).

1.1 Weight Trimming

While PMLEs are popular in practice for the reasons discussed above, the resulting increase in variance can overwhelm the reduction in bias, so that the MSE actually increases under a weighted analysis. This is particularly likely if a) the sample size is small, b) the difference in the probability of inclusion is large, or c) the association between the probability of inclusion and the data (which drives the bias) is weak. Perhaps the most common approach to dealing with this problem is *weight trimming* or winsorization (Potter 1990, Kish 1992, Alexander et al. 1997), in which weights larger than some value w_0 are fixed as w_0. Typically w_0 is chosen in an ad hoc manner – say 3 or 6 times the mean weight – without regard to whether the chosen cutpoint is optimal with respect to MSE. Thus bias is introduced to reduce variance, with the goal of an overall reduction in MSE. Less commonly used but more systematic are weight distribution techniques (Potter 1990) which assume weights follow an inverse beta distribution; w_0 is chosen as an upper tail value, say $1 - F(w_0) < .01$. The MSE trimming procedure (Elliott and Little 2000) determines the empirical MSE under the assumption that the true population mean is given by the fully weighted estimate, and a trimming level is chosen as the value that minimized the empirical MSE.

In contrast, we consider a Bayesian modeling approach that allows us to develop "weight-smoothing" models where the disproportional probability of selection is accounted for by considering the case weights as stratifying variables within strata defined by the probability of inclusion (Little 1983, 1991). Standard weighted estimates are then obtained when the stratum means of survey outcomes are treated as fixed effects, and smoothing of the weights is achieved by treating the underlying stratum means as random effects (Elliott and Little 2000, Holt and Smith 1979, Ghosh and Meeden 1986, Little 1991, 1993, Lazzeroni and Little 1998, Rizzo 1992). We term these random-effects models "weight smoothing models." Our goal is to extend these "weight smoothing" models that provide general model-based weight trimming estimators of population statistics, with particular application to estimation of the posterior predictive distribution of population parameters in linear and generalized linear models.

2.0 Bayesian Population Inference

As in design-based population inference, Bayesian population inference focuses on population quantities of interest $Q(Y)$, such as population means or population least-squares regression slopes. Inference is made about $Q(Y)$ by considering the marginal posterior predictive distribution (Ericson 1969, Holt and Smith 1979, Little 1993):

$$p(Q(Y) \mid y) = \int f(Q(Y) \mid \theta)p(\theta \mid y)d\theta = \quad (5)$$

$$\frac{\int f(Q(Y) \mid \theta)f(y \mid \theta)p(\theta)d\theta}{\int f(y \mid \theta)p(\theta)d\theta}.$$

If the sampling indicator I is independent of Y, as is the case in probability sampling design, then the sampling mechanism is said to be unconfounded or non-informative (Rubin 1987, Little 2004), and inference about $Q(Y)$ can be made using $p(Q(Y) \mid y)$ alone.

However, sensible models in (5) still need to account for the sample design in both the likelihood and prior model structure. In the case of the disproportional probability-of-inclusion sample designs, this can be done by developing an index $h = 1, \ldots, H$ of the probability of inclusion; this could either be a one-to-one mapping of the case weight order statistics to their rankings, or a preliminary "pooling" of the case weights using, e.g., the $100/H$ percentiles of the case weights. The data are then modeled by $y_{hi} \mid \theta_h \sim f(y_{hi}; \theta_h)$ where θ_h allows for an interaction between the model parameter(s) θ and the inclusion stratum h.

2.1 Weight Smoothing in Linear and Generalized Linear Regression Models

Generalized linear regression models postulate a likelihood for y_i of the form

$$f(y_i; \theta_i, \phi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right] \quad (6)$$

where $a_i(\phi)$ involves a known constant and a (nuisance) scale parameter ϕ, and the mean of y_i is related to a linear combination of covariates $\boldsymbol{x}_i$ through a link function $g(\cdot)$: $E(y_i \mid \theta_i) = \mu_i$, where $g(\mu_i) = g(b'(\theta_i)) = \eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta}$. We also have $\text{Var}(y_i \mid \theta_i) = a_i(\phi)V(\mu_i)$, where $V(\mu_i) = b''(\theta_i)$. The link is canonical if $\theta_i = \eta_i$, in which case $g'(\mu_i) = V^{-1}(\mu_i)$.

Indexing the inclusion stratum by h, we have $g(E[y_{hi} \mid \boldsymbol{\beta}_h]) = \boldsymbol{x}_{hi}^T\boldsymbol{\beta}_h$. We assume a hierarchical model of the form

$$(\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_H^T)^T \mid \boldsymbol{\beta}^*, G \sim N_{Hp}(\boldsymbol{\beta}^*, G). \quad (7)$$

We consider the target population parameter of interest $\boldsymbol{B} = (B_1, \ldots, B_p)^T$ to be the slope that solves the population score equation $U_N(\boldsymbol{B}) = 0$ where

$$U_N(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{\partial}{\partial \beta} \log f(y_i; \boldsymbol{\beta}) = \quad (8)$$

$$\sum_{h=1}^{H} \sum_{i=1}^{N_h} \frac{(y_{hi} - g^{-1}(\mu_i(\boldsymbol{\beta})))\boldsymbol{x}_{hi}}{V(\mu_{hi}(\boldsymbol{\beta}))g'(\mu_{hi}(\boldsymbol{\beta}))}.$$

Note that the quantity $\boldsymbol{B}$ such that $U(\boldsymbol{B}) = 0$ is always defined as the linear approximation of $\boldsymbol{x}_i$ to $\eta_i = g(\mu_i)$ even if the model is misspecificed (i.e., η_i is not exactly linear with respect to the covariates). Under the model given by (6) and (7), a first-order approximation (assuming a negligible sampling fraction) to $E(\boldsymbol{B} \mid y, X)$ is given by $\hat{\boldsymbol{B}}$ where

$$\sum_{h=1}^{H} W_h \sum_{i=1}^{n_h} \frac{(\hat{y}_{hi} - g^{-1}(\mu_i(\hat{\boldsymbol{B}})))\boldsymbol{x}_{hi}}{V(\mu_{hi}(\hat{\boldsymbol{B}}))g'(\mu_{hi}(\hat{\boldsymbol{B}}))} = 0 \quad (9)$$

where $W_h = N_h/n_h$ $\hat{y}_{hi} = g^{-1}(\boldsymbol{x}_{hi}^T \hat{\boldsymbol{\beta}}_h)$ and $\hat{\boldsymbol{\beta}}_h = E(\boldsymbol{\beta}_h \mid y, X)$, as determined by the form of (7). (If N_h is unknown, it can be replaced with $\sum_{iinh} w_{hi}$.) Thus, in the example of linear regression, where $V(\mu_i) = \sigma^2$ and $g'(\mu_i) = 1$, (9) resolves to

$$\hat{\boldsymbol{B}} = E(\boldsymbol{B} \mid y, X) = \left[\sum_h W_h \sum_{i=1}^{n_h} \boldsymbol{x}_{hi}\boldsymbol{x}_{hi}^T\right]^{-1} \times \quad (10)$$

$$\left[\sum_h W_h \left(\sum_{i=1}^{n_h} \boldsymbol{x}_{hi}\boldsymbol{x}_{hi}^T\right) \hat{\boldsymbol{\beta}}_h\right].$$

In the example of logistic regression, where $V(\mu_i) = \mu_i(1-\mu_i)$ and $g'(\mu_i) = \mu_i^{-1}(1-\mu_i)^{-1}$, $E(\boldsymbol{B} \mid y, X)$ is given by solving for B_j, $j = 1, \ldots, p$

$$\sum_{h=1}^{H} W_h \sum_{i=1}^{n_h} x_{hij} \frac{\exp(x_{hij}B_j)}{1 + \exp(x_{hij}B_j)} = \quad (11)$$

$$\sum_{h=1}^{H} W_h \sum_{i=1}^{n_h} x_{hij} \frac{\exp(x_{hij}\hat{\beta}_{hj})}{1 + \exp(x_{hij}\hat{\beta}_{hj})}$$

This can be accomplished via simple root-finding numerical methods such as Newton's Method.

We consider four forms of $\boldsymbol{\beta}^*$ and G in (10) in this manuscript:

Exchangeable Random Slope (XRS): (12)

$$\boldsymbol{\beta}^*{}_h = (\beta_0^*, \ldots, \beta_p^*) \text{ for all } h,\ G = I_H \otimes \Sigma.$$

Autoregressive Random Slope (ARS):

$$\boldsymbol{\beta}^*{}_h = (\beta_0^*, \ldots, \beta_p^*) \text{ for all } h,\ G = A \otimes \Sigma,\ A_{jk} = \rho^{|j-k|},\ j,k = 1, \ldots, H.$$

Linear Random Slope (LRS):

$$\boldsymbol{\beta}^*{}_h = (\beta_{00}^* + \beta_{01}^* h, \ldots, \beta_{p0}^* + \beta_{p1}^* h),\ G = I_H \otimes \Lambda.$$

Nonparametric Random Slope (NPRS):

$$\boldsymbol{\beta}^*{}_h = (f_0(h), \ldots, f_p(h)),\ G = 0.$$

$$\left\{ f_j : f_j^v \text{absolutely continuous}, v = 0, 1, \int (f_j^{(2)}(u))^2 du < \infty, \right.$$

$$\left. \min_{f_j} \sum_h \sum_i (\beta_{hj}^* - f_j(h))^2 + \lambda_j \int (f_j^{(2)}(u))^2 du \right\}$$

where Σ is an unconstrained $p \times p$ covariance matrix, Λ is a $p \times p$ diagonal matrix, and $f_j(h)$ is a twice differentiable smooth function of h that minimizes the residual sum of squares plus a roughness penalty parameterized by λ_j (Wahba 1978, Hastie and Tibshirani 1990). Reformulating the NPRS model as in Wang (1998) and Lin and Zhang (1999), we have

$$y_{hi} \mid \boldsymbol{\beta}_h \overset{ind}{\sim} N(\boldsymbol{x}_{hi}^T \boldsymbol{\beta}_h, \sigma^2)$$

$$\beta_{hj} = \beta_{j0}^* + \beta_{j1}^* h + \boldsymbol{\omega}_h \boldsymbol{u}_j$$

$$\boldsymbol{u}_j \overset{ind}{\sim} N_{H-1}(0, I\tau_j^2), \tau_j^2 = \sigma^2/(H\lambda_j)\ j = 0, \ldots, p$$

where $\boldsymbol{\omega}_h$ is the hth row of Choleski decomposition of the cubic spline basis matrix Ω where $\Omega_{hk} = \int_0^1 ((h-1)/(H-1)-t)_+ ((k-1)/(H-1)-t)_+ dt$, $(x)_+ = x$ if $x \geq 0$ and $(x)_+ = 0$ if $x < 0$, $h, k = 1, \ldots, H$. Note that, in the LRS and NPRS models, we assume *a priori* independence for the regression parameters associated with a given covariate, i.e., $(\beta_{1j}, \ldots \beta_{Hj}) \perp (\beta_{1j'}, \ldots \beta_{Hj'})$, $j \neq j'$. This is because we model trends in these parameters across the inclusion stratum, and do not wish to "link up" these trends across the covariates.

We complete the specification of the model by postulating a hyperprior for the second-stage parameters:

$$p(\phi, \boldsymbol{\beta}^*, G) \propto p(\boldsymbol{\zeta}).$$

Typically the hyperprior $p(\boldsymbol{\zeta})$ is either weakly informative or non-informative. Gibbs sampling (Gelfand and Smith 1990; Gelman and Rubin 1992) can then be utilized to obtained draws from the full joint posterior of $(\boldsymbol{\beta}, \boldsymbol{\beta}^*, \phi, G)^T \mid \boldsymbol{y}, X$.

In the XRS model, we consider $p(\sigma, \boldsymbol{\beta}^*, \Sigma) \propto \sigma^{-2}|\Sigma|^{-(p+1/2)} \exp(-1/2 \text{ tr } \{r\Sigma^{-1}\})$, that is, non-informative priors on the scale and prior mean parameters and an independent inverse-Wishart hyperprior on the prior variance G centered at the identity matrix scaled by r with p degree of freedom. The same prior is used for the ARS model, with the additional assumption that $\rho \overset{ind}{\sim} U(0,1)$ (non-negative autocorrelation between inclusion strata). In the LRS and NPRS models, $p(\sigma, \boldsymbol{\beta}^*, \Lambda)$ and $p(\sigma, \boldsymbol{\beta}^*, \tau) \propto \sigma^{-2}$ (standard non-informative scale prior and hyperprior).

3.0 Simulation Results

3.1 Linear Regression

For the linear regression model, we generated population data as follows:

$$Y_i \mid X_i, \sigma^2 \sim N(X_i + X_i^2, \sigma^2), \tag{13}$$

$$X_i \sim UNI(0,10),\ i = 1, \ldots, N = 20000.$$

An ignorable, disproportionally stratified sampling scheme sampled elements as a function of X_i (I_i equals 1 if sampled and 0 otherwise):

$$H_i = \lceil 2X_i \rceil$$

$$P(I_i = 1 \mid H_i) = \pi_h = 500H_i/210000.$$

This created 20 strata, defined by $H_i = 1$ if $0 \le X_i \le .5$, $H_i = 2$ if $.5 \le X_i \le 1, \ldots, H_i = 20$ if $9.5 \le X_i \le 10$. Elements (Y_i, X_i) had 1/20th the selection probability when $0 \le X_i \le .5$ as when $9.5 \le X_i \le 10$. A total of $n = 500$ elements were sampled without replacement for each simulation. The object of the analysis is to obtain the population slope $B_1 = \frac{\sum_{i=1}^N (Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1}^N (X_i - \overline{X})^2}$. The linear model is misspecified; the effect of this misspecification will increase as $\sigma^2 \to 0$ and decrease as $\sigma^2 \to \infty$. We considered values of $\sigma^2 = 10^{l/2}$, $l = 1, \ldots, 10$; 50 simulations were generated for each value of σ^2. The inverse-Wishart hyperprior on the prior variance G was centered at the identity matrix (r=1).

In addition to the exchangeable random slope (XRS), autoregressive random slope (ARS), linear random slope (LRS), and nonparametric random slope (NPRS) models discussed in Section 2.2, we consider the standard designed-based unweighted (UNWT), fully weighted (FTW), and trimmed weight (TWT) estimators. We obtained inference about the design-based estimators via the standard Taylor Series approximation (Binder 1983).

Tables 1 and 2 shows the square root mean square error (RMSE) for the three design-based and four model-based estimators of the population slope (second component of $\hat{\boldsymbol{B}}$) under consideration relative to the fully-weighted estimator, as a function of the variance σ^2. The unweighted and trimmed weight estimators perform poorly for small values of σ^2, where the bias due to model misspecification is critical, and well for larger values of σ^2, where the instability of the fully-weighted estimator is more important than bias reduction. The exchangeable model-based estimator has good RMSE properties for small and large values of σ^2, but tends to oversmooth for intermediate values of σ^2. The autoregressive model performance equals that of the exchangeable model for small and large values of σ^2, but is largely protected against the oversmoothing of the exchangeable models for intermediate values of σ^2. The linear and non-parametric model-based estimators essentially dominate the fully-weighted estimator with respect to RMSE, with reductions of RMSE on the order of 5-20%, but do not perform as well as the exchangeable and autoregressive models for large values of σ^2.

Tables 3 and 4 shows the true coverage of nominal 95% confidence intervals (CIs) or nominal 95% posterior predictive intervals (PPIs) for the three design-based and four model-based estimators of the population slope (second component of $\hat{\boldsymbol{B}}$) under consideration, as a function of the variance σ^2. The unweighted and winsorized estimators have poor coverage except for the largest values of σ^2. The nominal coverage of fully-weighted estimator is good for small values of σ^2, but drops for large values of σ^2. The failure of the bias-variance tradeoff for the exchangeable estimator is evident in the poor coverage of the estimator for intermediate values of σ^2; this effect is ameliorated, but not completely removed, for the autoregressive estimator. The linear estimator has generally good coverage for all but the largest values of σ^2; the nominal coverage of the nonparametric estimator is essentially correct throughout the range of σ^2 considered.

3.2 Logistic Regression

For the logistic regression model, we generated population data as follows:

$$P(Y_i = 1 \mid X_i) \sim BER(\text{expit}(1.5 - .75 * X_i + C * X_i^2)), \tag{14}$$

$$X_i \sim UNI(0,10),\ i = 1, \ldots, N = 20000.$$

where $\text{expit}(\cdot) = exp(\cdot)/(1 + exp(\cdot))$. The sampling scheme is as described in the linear regression simulations. The object of the analysis is to obtain the logistic population regression slope, defined as the value B_1 in the equation

$$\sum_{i}^{N}(y_i - \text{expit}(B_0 + B_1 x_i)) \begin{pmatrix} 1 \\ x_i \end{pmatrix} = 0. \tag{15}$$

The linear model is misspecified unless $C = 0$, with the degree of misspecification increasing as C increases. We consider values of $C = 0, .1, .2, \ldots, .9$; 50 simulations were generated for each value of C. The inverse-Wishart hyperprior on the prior variance G was again centered at the identity matrix (r=1).

Tables 5 and 6 shows the RMSE relative to the RMSE of the fully-weighted estimator for each of the five other estimators under consideration. The unweighted estimator had better MSE behavior than the fully-weighted estimator when the linear slope model was correctly specified, or when the linear slope model was badly misspecified. In the latter case, the fully-weighted estimator's attempts to correctly capture the effect of the misspecification on the linear approximation makes it extemely unstable. The XRS estimator does not fail as spectacularly at intermediate values of misspecification as the unweighted estimator, but does not have substantially improved MSE performance over the fully-weighted estimator except under the same conditions under which the unweighted estimator also bests the fully-weighted estimator. The ARS estimator does improve over the fully-weighted estimator for small amounts of misspecification, but fails again at the intermediate regions. The TWT estimator generally improves upon the fully-weighted estimator, while the LRS and NPRS estimators improve substantially on the fully-weighted estimator over the entire range of misspecification considered, with MSE reductions of up to 25%.

Tables 7 and 8 shows the true coverage of the nominal 95% CIs for the three design-based estimators and the nominal 95% PPIs for the three model-based estimators. The unweighted estimator had poor coverage except when the linear slope model was correctly specified, or when the linear slope model was badly misspecified. Interestingly, the fully-weighted estimator's true coverage was as low as 70% for large degrees of misspecification. However, the trimmed estimator's true coverage was only somewhat below its nominal value, even for large degrees of misspecification. The XRS and ARS estimators generally had correct or nearly correct coverage levels for all but the largest degree of misspecification. The LRS estimator had coverage similar to that of the fully-weighted estimator, while the NPRS estimator had marginally better coverage.

4.0 Discussion

The models discussed in this manuscript generalize the work of Lazzaroni and Little (1998) and Elliott and Little (2000), where population inference was restricted to population means under Gaussian distributional assumptions. Viewing weighting as an interaction between inclusion probability and model parameters opens up an alternative paradigm for winsorization as a random effects model that smoothes model parameters of interest across inclusion classes. Models with exchangeable mean structures offer the largest degree of shrinkage or trimming but the most sensitivity to model misspecification; models with highly structured means are potentially less efficient but are more robust to model misspecification.

Disproportional sample designs are more likely to be ignorable when population regression parameters are of interest, given independence of the sampling indicator I and the outcome Y is required only conditional on the predictors X, not marginally. Consequently we might expect that the weight smoothing models will have greater application than in the analyses where population means and totals are of interest. This is generally true in the Gaussian linear model setting, where the estimators obtained from the linear and nonparametric smoothing models dominated the fully-weighted estimators with respect to squared error loss in the simulations considered. Matters are somewhat less clear-cut in the logistic model setting; although the LRS model again dominated the fully-weighted estimator with respect to squared error loss in the simulations considered, both estimators had lower nominal coverage than the XRS and ARS estimators, which in turn did not perform as well with respect to MSE as either the LRS or the fully-weighted estimator under certain model misspecification conditions. Indeed, the simple trimming estimator had reasonable good performance from both an MSE and nominal coverage perspective.

These results show the promise of adapting model-based methods to attack problems in survey data analysis. Although computationally intensive, they are applications or extensions of the existing random-effect model "toolbox," and can either be implemented in existing statistical packages or executed with relatively simple MCMC methods.

References

Alexander, C.H., Dahl, S., Weidman, L. (1997). Making estimates from the American Community Survey. Proceedings of the Social Statistics Section, American Statistical Association, 2000, 88-97.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, **51**, 279-292.

Cochran, W.G. (1977). *Sampling Techniques, 3rd Edition.* Wiley: New York.

Elliott, M.R., Little, R.J.A. (2000). Model-based approaches to weight trimming. *Journal of Official Statistics*, **16**, 191-210.

Ericson, W.A. (1969). Subjective Bayesian Modeling in Sampling Finite Populations, *Journal of the Royal Statistical Society,* **B31**, 195-234.

Gelfand, A.E., and Smith, A.M.F. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 389-409.

Gelman, A., and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457-472.

Ghosh, M., and Meeden, G. (1986). Empirical Bayes estimation of means from stratified samples. *Journal of the American Statistical Association*, **81**, 1058-1062.

Holt, D., Smith, T.M.F. (1979). Poststratification. *Journal of the Royal Statistical Society*, **A142**, 33-46.

Horvitz, D.G., Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663-685.

Kish, L. (1992). Weighting for unequal P_i. *Journal of Official Statistics*, **8**, 183-200.

Korn, E.L., Graubard, B.I. (1999). *Analysis of Health Surveys.* Wiley: New York.

Lazzaroni, L.C. and Little, R.J.A. (1998). Random-effects models for smoothing post-stratification weights. *Journal of Official Statistics*, **14**, 61-78.

Lin, X., Zhang, D. (1999). Inference in generalized additve mixed models by using smoothing splines. *Journal of the Royal Statistical Society,* **B61**, 381-400.

Little, R.J.A. (1983). Estimating a finite population mean from unequal probability samples. *Journal of the American Statistical Association*, **78**, 596-604.

Little, R.J.A. (1991). Inference with survey weights. *Journal of Official Statistics*, **7**, 405-424.

Little, R.J.A. (1993). Poststratification: A modeler's perspective. *Journal of the American Statistical Association*, **88**, 1001-1012.

Little, R.J.A. (2004). To model or not model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association*, **99**, 546-556.

Patterson, B.H., Dayton, C.M., Graubard, B.I. (2002). Latent class analysis of complex sample survey data: application to dietary data (with discussion). *Journal of the American Statistical Association*, **97**, 721-741.

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

Potter, F. (1990). A study of procedures to identify and trim extreme sample weights. Proceedings of the Survey Research Methods Section, American Statistical Association, 1990, 225-230.

Rizzo, L. (1992). Conditionally consistent estimators using only probabilities of selection in complex sample surveys. *Journal of the American Statistical Association* **87** , 1166-1173

Rubin, D.B. (1987). *Multiple Imputation for Non-Response in Surveys.* New York: Wiley.

Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, **93**, 341-348.

Variance $\log_{10}(\sigma^2)$	Estimator UNWT	FWT	TWT
.5	21.2	1	5.52
1	14.8	1	4.00
1.5	8.14	1	2.19
2	5.25	1	1.67
2.5	2.84	1	1.14
3	1.77	1	0.95
3.5	1.11	1	0.93
4	0.86	1	0.85
4.5	0.80	1	0.83
5	0.74	1	0.85

Table 1: Square root of mean square error (RMSE) relative to RMSE of fully-weighted estimator of population linear regression slope estimator where population slope and intercept are estimated via design-based unweighted (UNWT), fully-weighted (FWT), and weight-trimmed estimators (TWT).

Variance $\log_{10}(\sigma^2)$	Estimator UNWT	FWT	TWT
.5	0	94	32
1	0	96	24
1.5	0	94	46
2	0	94	64
2.5	0	94	80
3	12	94	88
3.5	58	92	96
4	86	90	92
4.5	86	92	98
5	92	86	96

Table 3: True coverage of the 95% CI or PPI of population linear regression slope estimator where population slope and intercept are estimated via design-based unweighted (UNWT), fully-weighted (FWT), and weight-trimmed estimators (TWT). 90% of CI/PPI coverage estimates should lie in 90-100% due to sampling variability if nominal coverage is correct.

Variance	Estimator XRS	ARS	LRS	NPRS
.5	0.89	0.88	0.76	0.86
1	0.99	0.97	0.89	0.87
1.5	1.00	0.92	0.84	0.90
2	1.22	1.05	0.99	0.93
2.5	1.27	1.01	0.87	0.92
3	1.22	1.09	0.99	0.93
3.5	0.95	0.90	0.85	0.96
4	0.97	0.97	1.03	0.99
4.5	0.83	0.82	0.91	0.93
5	0.84	0.84	0.96	0.96

Table 2: Square root of mean square error (RMSE) relative to RMSE of fully-weighted estimator of population linear regression slope estimator where population slope and intercept are estimated as the posterior mean in (12) under an exchangeable (XRS), autoregressive (ARS), linear (LRS), and non-parametric (NPRS) prior for the regression parameters.

Variance $\log_{10}(\sigma^2)$	Estimator XRS	ARS	LRS	NPRS
.5	96	94	96	90
1	96	96	94	98
1.5	96	96	98	94
2	88	98	94	100
2.5	78	90	96	96
3	78	86	92	96
3.5	78	92	100	94
4	86	88	90	96
4.5	92	96	94	96
5	88	86	96	92

Table 4: True coverage of the 95% CI or PPI of population linear regression slope estimator where population slope and intercept are estimated as the posterior mean in (12) under an exchangeable (XRS), autoregressive (ARS), linear (LRS), and non-parametric (NPRS) prior for the regression parameters. 90% of CI/PPI coverage estimates should lie in 90-100% due to sampling variability if nominal coverage is correct.

Curvature C	Estimator UNWT	FWT	TWT
0	0.72	1	0.85
.1	2.22	1	1.00
.2	1.76	1	0.98
.3	1.25	1	0.88
.4	1.13	1	0.98
.5	1.01	1	0.93
.6	0.97	1	0.93
.7	0.66	1	0.91
.8	0.45	1	0.72
.9	0.65	1	0.86

Table 5: Square root of mean square error (RMSE) relative to RMSE of fully-weighted estimator of population logistic regression slope estimator where population slope and intercept are estimated via design-based unweighted (UNWT), fully-weighted (FWT), and weight-trimmed estimators (TWT).

Variance C	Estimator UNWT	FWT	TWT
.5	98	94	96
1	2	96	90
1.5	16	90	98
2	58	82	98
2.5	76	78	90
3	84	72	90
3.5	88	82	90
4	94	82	92
4.5	92	78	86
5	76	62	72

Table 7: True coverage of the 95% CI or PPI of population logistic regression slope estimator where population slope and intercept are estimated via design-based unweighted (UNWT), fully-weighted (FWT), and weight-trimmed estimators (TWT). 90% of CI/PPI coverage estimates should lie in 90-100% due to sampling variability if nominal coverage is correct.

Curvature C	Estimator XRS	ARS	LRS	NPRS
0	0.86	0.87	0.81	0.92
.1	0.97	0.92	0.80	0.87
.2	1.13	0.99	0.85	0.94
.3	1.06	1.00	0.88	0.89
.4	1.18	1.12	0.86	0.94
.5	1.12	1.09	0.89	0.79
.6	1.26	1.25	0.86	0.95
.7	0.97	1.04	0.95	0.76
.8	0.85	0.86	0.76	1.01
.9	1.47	1.60	1.25	0.97

Table 6: Square root of mean square error (RMSE) relative to RMSE of fully-weighted estimator of population logistic regression slope estimator where population slope and intercept are estimated as the posterior mean in (12) under an exchangeable (XRS), autoregressive (ARS), linear (LRS), and non-parametric (NPRS) prior for the regression parameters.

Variance C	Estimator XRS	ARS	LRS	NPRS
.5	90	90	88	86
1	90	98	94	90
1.5	88	94	84	92
2	98	94	82	100
2.5	90	88	76	94
3	90	86	70	76
3.5	90	96	82	76
4	92	88	74	78
4.5	88	88	82	82
5	78	78	58	76

Table 8: True coverage of the 95% CI or PPI of population logistic regression slope estimator where population slope and intercept are estimated as the posterior mean in (12) under an exchangeable (XRS), autoregressive (ARS), linear (LRS), and non-parametric (NPRS) prior for the regression parameters. 90% of CI/PPI coverage estimates should lie in 90-100% due to sampling variability if nominal coverage is correct.