Post-Stratification: A Modeler's Perspective

R. J. A. LITTLE*

Post-stratification is a common technique in survey analysis for incorporating population distributions of variables into survey estimates The basic technique divides the sample into post-strata, and computes a post-stratification weight $w_h = rP_h/r_h$ for each sample case in post-stratum h, where r_h is the number of survey respondents in post-stratum h, P_h is the population proportion from a census, and r is the respondent sample size Survey estimates, such as functions of means and totals, then weight cases by w_h. Variants and extensions of the method include truncation of the weights to avoid excessive variability and raking to a set of two or more univariate marginal distributions Literature on post-stratification is limited and has mainly taken the randomization (or design-based) perspective, where inference is based on the sampling distribution with population values held fixed. This article develops Bayesian model-based theory for the method A basic normal post-stratification model is introduced which yields the post-stratified mean as the posterior mean, and a posterior variance that incorporates adjustments for estimating variances Modifications are then proposed for small sample inference, based on (a) changing the Jeffreys prior for the post-stratum parameters to borrow strength across post-strata, and (b) ignoring partial information about the post-strata In particular, practical rules for collapsing post-strata to reduce posterior variance are developed and compared with frequentist approaches Methods for two post-stratifying variables are also considered Raking sample counts and respondent counts is shown to provide approximate Bayesian inferences when the margins of the two post-stratifiers are available but their joint distribution is not When the joint distribution is available, raking effectively ignores the information it contains, and hence can be compared with other techniques that ignore information such as collapsing For inference about means, it is suggested that raking is most appropriate when post-stratum means have an additive or near-additive structure, whereas collapsing is indicated when interactions are present

KEY WORDS: Collapsing strata, Raking, Stratification, Superpopulation models, Survey inference, Weighting

1. INTRODUCTION

Probability sampling is one of the major contributions of statistics to science; however, many dislike the lack of control inherent in simple random sampling Stratified sampling maintains the probability nature of the sample while controlling its composition with respect to important characteristics. Much has been written about the method since Neyman's (1934) landmark paper.

Stratified sampling is limited to variables that are known for survey units prior to data collection. Post-stratification combines data collected in the survey with aggregate data on the population from other sources For example, a demographic survey generally cannot stratify on age, because the ages of individuals are not available until the interview is conducted. But the population age distribution may be available in aggregate form, from census data. Post-stratification (in its basic form) classifies the sample by age group and then weights individuals in each group, or post-stratum, up to the population total in that group. Specifically, the weight $w_h = rP_h/r_h$ is computed for each sample case in post-stratum h, where r_h is the number of respondents in post-stratum h, P_h is the population proportion from a census, and r is the respondent sample size; weights are scaled so that they sum to the respondent sample size. (Post-stratification has a different meaning in clinical trials, where it refers to stratified analysis of data from unstratified randomized designs; this article concerns the survey technique only.)

Post-stratification can improve the accuracy of survey estimates, both by reducing bias and by increasing precision. In the preceding example, the mean age of the population might be estimated by the sample mean, given an equal probability sample But the mean age weighted by the $\{w_h\}$ is much more precise, because it essentially reproduces the population mean aside from effects of grouping. Furthermore, if the unweighted mean is biased by differential nonresponse by age, then the post-stratified mean corrects for this bias. These properties are of academic interest given the availability of data on age from the census, but they also apply in diluted form to other survey variables that are correlated with age.

Post-stratification is very common in practice, playing an important role in many government surveys (see, for example, Hanson 1978, Harte 1982; Waterton and Lievesley 1987). In a seminal paper, however, Holt and Smith (1979) noted that its statistical properties have received relatively little attention. The literature by and large studies the method from the randomization perspective, where the bias and mean squared error of estimators are assessed over the sampling distribution, with population values treated as fixed. Post-stratification is considered here from the predictive modeling perspective, where population values are treated as random variables under a model and inference about finite population quantities is based on their predictive distribution under the model. I apply the Bayesian version of the modeling approach, where unknown parameters in the model are assigned prior distributions, rather than the superpopulation model formulation, where such parameters are treated as fixed With noninformative priors, these approaches often yield equivalent or very similar answers.

* R J A Little is Professor in the Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109-2029 This article was the 1991 *Journal of the American Statistical Association* Theory and Methods Invited Paper at the Joint Statistical Meetings, Atlanta, GA The research was supported by National Institute of Mental Health Grant MH 37188 The author thanks Ed Wegman for his encouragement and Robert Fay, Laura Lazzeroni, Nat Schenker, Robert Weiss, an associate editor and two referees for constructive comments

Journal of the American Statistical Association
September 1993, Vol. 88, No. 423, Theory and Methods

From the randomization perspective it is natural to view post-stratification as a method of weighting adjustment, which is the form in which the method typically appears to the survey analyst. A weight is attached to each sample unit, that is proportional to the number of population units the unit "represents" (Fig. 1a). From a predictive modeling perspective, the data are more accurately depicted as in Figure 1b, where the post-stratifier Z is known for all N units in the population and the survey variables are measured only for the n sampled units. Analysis effectively fills in the missing data in the full rectangular array. Figure 1a has a convenient rectangular form for analysis, but Figure 1b is the basic form of the data.

Section 2 discusses post-stratification on a single categorical margin, the basic version of the method. Bayesian analyses are considered when a simple random sample of the population is selected and the role of post-stratification is to reduce variance and when selection is biased by nonresponse or frame errors, where post-stratification also has a role in reducing bias. Methods for collapsing small cells are provided in the two cases, and other approaches to the small-cell problem are outlined. A Bayesian perspective is provided for the intriguing issues of conditioning that arise under randomization inference.

Section 3 considers methods of post-stratification to margins from two or more variables. The method of raking is commonly used in such situations; Bayesian justifications of raking sample counts and respondent counts are presented, and the added posterior variance from raking is discussed. When the joint population counts are known, raking can be viewed as ignoring information to deal with the problem of sparse cells and as such can be compared with alternative methods, such as collapsing. Section 4 suggests some topics for future research.

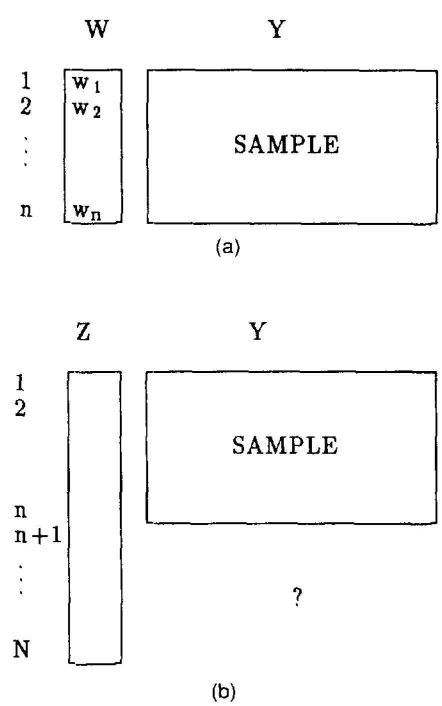

Figure 1. Weighted Data from a Post-Stratified Survey (a) and Post-stratification as a Prediction Problem (b)

2. POST-STRATIFICATION TO A SINGLE MARGIN

2.1 The Post-Stratified Mean

Let Z denote the post-stratifying variable, let Y denote a survey variable, and consider inference for the finite population mean $\bar{Y} = \sum_h P_h \bar{Y}_h$, where $\bar{Y}_h$ is the mean in post-stratum h. Suppose that a simple random sample of size n is selected, r of which respond to the survey; additional complexities from clustering and stratification of the sample design are not considered here. Let n_h and $r_h \leq n_h$ denote the number sampled and the number responding in post-stratum h. Write $\mathbf{P} = \{P_1, \ldots, P_H\}$, $\mathbf{n} = \{n_1, \ldots, n_H\}$, and $\mathbf{r} = \{r_1, \ldots, r_H\}$. Then $\mathbf{P}$ and $\mathbf{r}$ are assumed known. The sample counts $\mathbf{n}$ may or may not be known, depending on whether Z is known for sample nonrespondents. For example, if Z denotes age group and if nonrespondents are individuals who refused to participate, then their ages may be available from a household listing, in which case $\mathbf{n}$ is known. On the other hand, if nonrespondents are from noncontacted households, then their ages may not be known, and hence $\mathbf{n}$ is also unknown. Frame undercoverage can be included by treating as nonrespondents cases excluded from the frame who should have been sampled. In that case $\mathbf{n}$ is also unknown. We shall see that for a single post-stratifier, the question of whether $\{\mathbf{n}\}$ is known or not is irrelevant, because it plays no role in inference. In Section 3.2 it is shown that when there is partial information on two or more post-stratifiers, $\{\mathbf{n}\}$ can play a role in inference, so the distinction matters.

I assume that nonresponse is ignorable (Little 1982; Rubin 1976), in the sense that respondents within post-stratum h can be treated as a random subsample of sampled cases in post-stratum h, formally, r_h has a binomial distribution with index n_h and response probability φ_h:

$$r_h | \mathbf{n} \sim \mathrm{bin}(n_h, \varphi_h).$$

A stronger assumption is that the nonresponse rate is the same across post-strata. Then $\varphi_h = \varphi$ for all h, and respondents are a random subsample of sampled cases overall. I call this the *missing completely at random* (MCAR) assumption, using Rubin's (1976) terminology; it applies trivially with complete response (i.e., when $\mathbf{r} = \mathbf{n}$). Under MCAR, post-stratification is not needed for reducing or eliminating nonresponse bias, but it can reduce variance. If data are not MCAR, then post-stratification can reduce bias from frame errors or nonresponse.

The usual estimator of $\bar{Y}$ is the post-stratified or weighted mean

$$\bar{y}_{\mathrm{ps}} = \sum_{h=1}^{H} P_h \bar{y}_h = \frac{1}{r} \sum_{i=1}^{r} w_i y_i, \tag{1}$$

where y_i is the value of Y for respondent i, $\bar{y}_h$ is the respondent sample mean in post-stratum h, and $w_i = rP_h/r_h$ if $z_i = h$; that is, case i belongs to stratum h.

2.2 The Randomization Variance of the Post-Stratified Mean

The appropriate randomization variance for $\bar{y}_{ps}$ is controversial. The unconditional sampling variance is

$$\begin{aligned}\text{var}(\bar{y}_{ps}) &= E\{\text{var}(\bar{y}_{ps}|\mathbf{r})\} + \text{var}\{E(\bar{y}_{ps}|\mathbf{r})\} \\ &= E\{\text{var}(\bar{y}_{ps}|\mathbf{r})\}, \end{aligned} \quad (2)$$

the second term being 0 because $E(\bar{y}_{ps}|\mathbf{r}) = \bar{Y}$, a constant. Most sampling texts offer estimates of approximations to (2) (see, for example, Cochran 1977) But Holt and Smith (1979) argued that $\{\mathbf{r}\}$ is ancillary and (2) yields a poor estimate of precision when $\{\mathbf{r}\}$ deviates markedly from its expectation They instead proposed estimates of the conditional variance

$$\text{var}(\bar{y}_{ps}|\mathbf{r}) = \sum P_h^2(1 - f_h)S_h^2/r_h, \quad (3)$$

where S_h^2 is the variance of Y in post-stratum h and $1 - f_h = 1 - r_h/N_h$ is a finite population correction. This is the usual expression for variance of the stratified mean. Some survey statisticians (Kalton and Maligalig 1991, Oh and Scheuren 1983) have followed this conditional approach The difference between (2) and (3) is of order r^{-2} and hence not a major issue given large samples; however, with many post-strata and estimates for subdomains of the population, the difference can be nonnegligible.

The basic underlying problem is that the randomization approach is ambiguous about whether to condition on ancillaries that do not provide direct information about the quantity of interest but do index precision. Indeed, although $\{\mathbf{r}\}$ is intuitively ancillary, I am not aware of a formal theory of ancillarity in the randomization theory of surveys

2.3 Bayesian Inference for $\bar{Y}$ in the Presence of Post-Strata

Bayesian inference for $\bar{Y} = \sum P_h \bar{Y}_h$ requires a model for Y given Z, yielding predictions of $\{\bar{Y}_h\}$; the posterior distribution of $\bar{Y}$ is summarized by the posterior mean

$$E(\bar{Y}|\text{data}) = \sum_h P_h E(\bar{Y}_h|\text{data}), \quad (4)$$

which is analogous to the estimator in randomization inference and the posterior variance

$$\text{var}(\bar{Y}|\text{data}) = \sum_h P_h^2 \text{var}(\bar{Y}_h|\text{data}) + \sum\sum_{h \neq k} P_h P_k \text{cov}(\bar{Y}_h, \bar{Y}_k|\text{data}), \quad (5)$$

which plays the role of the estimate of precision. Now consider inference under the basic normal post-stratification model (BNPM)

$$(y_i|z_i = h, \mu_h, \sigma_h^2) \sim_{\text{ind}} G(\mu_h, \sigma_h^2);$$
$$p(\mu_h, \log \sigma_h) = \text{const}, \quad (6)$$

where y_i is the value of Y for unit i, z_i identifies post-stratum, $G(a, b)$ is a normal (Gaussian) distribution with mean a and variance b, and $p(\mu_h, \log \sigma_h)$ is the noninformative Jeffreys prior for the post-stratum means and variances. The Gaussian assumption may be inappropriate and profitably refined, but it is of secondary importance; the main features of (6) are the inclusion of a distinct mean and variance in each post-stratum and the iid assumption within post-strata, which requires modification for designs with clustering and differential selection rates within post-strata.

A standard Bayesian analysis under (6) yields the posterior distribution of $\bar{Y}$ as a mixture of t distributions with mean and variance

$$E(\bar{Y}|\mathbf{Z}, \mathbf{Y}_s) = \bar{y}_{ps},$$
$$\text{var}(\bar{Y}|\mathbf{Z}, \mathbf{Y}_s) = v_{ps} = \sum_h P_h^2(1 - f_h)\delta_h s_h^2/r_h, \quad (7)$$

where in post-stratum h, s_h^2 is the sample variance of Y and $\delta_h = (r_h - 1)/(r_h - 3)$ is a small-sample correction for estimating the variance; at least four respondents in each post-strata are required to apply this correction. The data are denoted $(\mathbf{Z}, \mathbf{Y}_s)$, where $\mathbf{Z}$ represents the sample and census data on Z and $\mathbf{Y}_s = \{y_i: i = 1, \ldots, r\}$.

As noted by Holt and Smith (1979), this model analysis supports the randomization variance (3) that conditions on $\{\mathbf{r}\}$; $\text{var}(\bar{Y}|\mathbf{Z}, \mathbf{Y}_s)$ in (7) differs from (3) only in the substitution of sample estimates for the (unknown) population variances and in the attendant small-sample correction.

2.4 Comparisons With the Unweighted Mean

The post-stratified mean achieves gains in efficiency in large samples, where it has similar properties to the stratified mean It is unstable, in small samples, and is undefined if any post-strata have no respondents. One view of the problem is that the weight $w_h = rP_h/r_h$ is unstable when r_h is small and needs trimming or smoothing to control variance.

The most extreme form of smoothing the weights is to set them all equal to 1, yielding the unweighted mean $\bar{y} = 1/r \sum_{i=1}^r y_i = \sum_{h=1}^H p_h \bar{y}_h$, where $p_h = r_h/r$; this corresponds to ignoring information in the post-strata. Holt and Smith (1979) provided numerical comparisons of $\bar{y}$ and $\bar{y}_{ps}$ assuming MCAR and found $\bar{y}_{ps}$ to be superior unless the sample size is small and the ratio of between-stratum to within-stratum variance is small.

Questions about whether to condition on $\{\mathbf{r}\}$ also arise when comparing $\bar{y}_{ps}$ with $\bar{y}$. The unconditional sampling variance of $\bar{y}$ is

$$\text{var}(\bar{y}) = (1 - f)S^2/r, \quad (8)$$

where S^2 is the population variance of Y and $f = r/N$. But it is not clear whether (3), which treats $\{\mathbf{r}\}$ as fixed, can be compared with (8), which treats $\{\mathbf{r}\}$ as random. Holt and Smith (1979) compared (3) with the conditional mean squared error (MSE) of $\bar{y}$,

$$\text{mse}(\bar{y}|\mathbf{r}) = \text{var}(\bar{y}|\mathbf{r}) + b^2, \quad (9)$$

where $\text{var}(\bar{y}|\mathbf{r}) = \sum_h p_h^2(1 - f_h)S_h^2/r_h$ and $b = \sum_h (p_h - P_h)\bar{Y}_h$ is the conditional bias of $\bar{y}$. But (9) is a peculiar measure of precision for the unweighted mean and is rarely, if ever, used in practice.

From a modeling perspective, variable weights are a symptom rather than the cause of the problem of post-stratification with small samples. The key issue is that the underlying model (6) yields poor predictions of the distribution of Y in post-strata where r_h is small; the posterior mean $\bar{y}_h$ is poorly determined, and the posterior variance v_{ps} requires at least four respondents in each post-stratum to be defined at all. Hence the inference (7) must be modified to allow pooling of strength across post-strata. The principled modeling approach is to compute the posterior mean and variance under a modified model for the distribution of Y given Z. A less principled approach is to model ignoring some or all of the information in the Z margin. The former approach is more principled in that it builds a model for all the relevant data; however, ignoring some information may be reasonable if it simplifies the modeling process or limits the effects of model misspecification (Rubin 1984).

Versions of both these approaches yield $\bar{y}$ as the posterior mean and hence provide a Bayesian context for the preceding discussion. The null model that assumes the same distribution of Y across post-strata,

$$(y_i | z_i = h, \mu, \sigma^2) \sim_{\text{ind}} G(\mu, \sigma^2); \qquad p(\mu, \log \sigma) = \text{const}, \quad (10)$$

yields $E(\bar{Y} | \mathbf{Z}, \mathbf{Y}_s) = \bar{y}$ and $\text{var}(\bar{Y} | \mathbf{Z}, \mathbf{Y}_s) = v = \delta(1 - f)s^2/r$, where s^2 is the sample variance of Y and $\delta = (r - 1)/(r - 3)$. Alternatively, ignoring the data on Z we might assume that

$$(y_i | \mu, \sigma^2) \sim_{\text{ind}} G(\mu, \sigma^2); \qquad p(\mu, \log \sigma) = \text{const}, \quad (11)$$

which again yields $E(\bar{Y} | \mathbf{Y}_s) = \bar{y}$ and $\text{var}(\bar{Y} | \mathbf{Y}_s) = v$. Note that (10) implies (11), but not vice versa. Under MCAR the distinction between (10) and (11) is of no practical import, because they yield the same inference. When data are not MCAR, inference under (10) is unaffected, but (11) can yield different answers because it does not assume independence of Y and Z.

Bayesian analysis under (6) leads to posterior mean $\bar{y}_{ps}$ and variance v_{ps}; Bayesian analysis under (10) or (11) leads to posterior mean $\bar{y}$ and variance v. The comparison of precisions thus is between v_{ps} and v, which correspond to the conditional variance of $\bar{y}_{ps}$ given $\{\mathbf{r}\}$ and the unconditional variance of $\bar{y}$ in the randomization framework. Thus the Bayesian approach leads to the two natural measures of precision (v_{ps} and v) and also provides small-sample corrections (δ, δ_h) for estimating the variance.

2.5 Algorithms for Collapsing Post-strata

2.5.1 An Algorithm Assuming Missing Completely at Random. A useful compromise between using and ignoring post-stratum counts is to collapse small post-strata that contribute excessively to the variance. Frequentist strategies for collapsing were considered by Tremblay (1986) and by Kalton and Maligalig (1991). This subsection proposes practical criteria for collapsing under MCAR, aimed at reducing the expected value of the posterior variance v_{ps} associated with the BNPM (6). The next subsection proposes a modification when MCAR is not assumed, and hence bias may be an issue.

Collapsing two post-strata $Z = i$ and $Z = j$ is interpreted as combining the population proportions P_i and P_j and modifying (6) by replacing the means and variance in these post-strata by a single mean and variance for the combined post-strata. The posterior mean under this collapsed model is

$$\bar{y}_{ps}^{(ij)} = \sum_{h \neq i,j} P_h \bar{y}_h + (P_i + P_j)\bar{y}_{(i+j)},$$

where $\bar{y}_{(i+j)} = (r_i \bar{y}_i + r_j \bar{y}_j)/(r_i + r_j)$ is the respondent mean of Y pooled over i and j. The posterior variance can be written as

$$v_{ps}^{(ij)} = v_{ps} - \Delta v_{ij},$$

$$\Delta v_{ij} = \frac{P_i^2 \delta_i (1 - f_i) s_i^2}{r_i} + \frac{P_j^2 \delta_j (1 - f_j) s_j^2}{r_j} - \frac{(P_i + P_j)^2 \delta_{ij} (1 - f_{ij}) s_{ij}^2}{r_i + r_j}, \quad (12)$$

where $\delta_{ij} = (r_i + r_j - 1)/(r_i + r_j - 3)$, $f_{ij} = (r_i + r_j)/(N_i + N_j)$, and s_{ij}^2 is the sample variance after collapsing post-strata i and j. If the correction terms (δ_i and δ_{ij}) for estimating the variance are ignored, then the expected reduction in posterior variance is

$$E(\Delta v_{ij}) = \frac{P_i^2 (1 - f_i) \sigma_i^2}{r_i} + \frac{P_j^2 (1 - f_j) \sigma_j^2}{r_j} - \frac{(P_i + P_j)^2 (1 - f_{ij}) \sigma_{ij}^2}{r_i + r_j}, \quad (13)$$

where σ_i^2 and σ_{ij}^2 are the expected values of s_i^2 and s_{ij}^2. To express (13) in terms of sample weights, note that if Z and Y are independent, then $\sigma_i^2 = \sigma_j^2 = \sigma_{ij}^2 = \sigma^2$, say, and (13) reduces to

$$E\{\Delta v_{ij}/\sigma^2\} \simeq \frac{r_i r_j}{r^2(r_i + r_j)} \{w_i - w_j\}^2.$$

Hence in this case the variance is always reduced by collapsing, provided that the weights w_i and w_j in the collapsed post-strata are unequal. But when Z and Y are associated, this reduction is counteracted by an increase in variance σ_{ij}^2 from collapsing. (The increase in variance is squared bias from the conditional frequentist perspective of Section 2.2.)

The following collapsing algorithm based on (13) is suggested:

1. Order the post-strata so that neighbors are a priori relatively homogeneous. If they are based on an ordered variable (such as grouped age), then this step is not needed.
2. Collapse the post-stratum pair (i, j) that maximizes (13), subject to the restriction that $j = i + 1$, that is, only neighboring pairs are considered. (This restriction is designed to limit the increase in the variance from collapsing, in that neighboring post-strata are more likely to be homogeneous with respect to outcomes.)
3. Proceed sequentially until a reasonable number of pooled post-strata remain, or (13) becomes noticeably negative.

The main task in operationalizing this algorithm is to choose forms for updating the variances terms σ_i^2 and σ_{ij}^2, which depend on the form of association between Z and Y.

Note that

$$\sigma_{ij}^2 = p_{ij}\sigma_i^2 + (1 - p_{ij})\sigma_j^2 + p_{ij}(1 - p_{ij})(\mu_i - \mu_j)^2, \quad (14)$$

where $p_{ij} = r_i/(r_i + r_j)$, r_i is the sample size in collapsed post-stratum i before pooling with j, and μ_i is the expected post-stratum mean. Suppose that prior to any collapsing, Z has K ordered post-strata and Y is linearly related to Z with correlation ρ and a constant residual variance σ^2. Initially, the means of two adjacent post-strata deviate by $4\rho\sigma/(K\sqrt{1-\rho^2})$, where the standard deviation of Z has been approximated as $K/4$. Substituting in (14) with $\sigma_i^2 = \sigma_{i+1}^2 = \sigma^2$, the variance after pooling post-strata i and $i + 1$ is

$$\sigma_{i,i+1}^2 = \sigma^2[1 + 16p_{i,i+1}(1 - p_{i,i+1})\rho^2/\{K^2(1-\rho^2)\}].$$

Later in the algorithm, let m_i be the number of original post-strata and let σ_i^2 be the expected variance in current post-stratum i. The means of collapsed post-strata i and $i + 1$ under the linear model deviate by approximately $2\rho(m_i + m_{i+1})/(K\sqrt{1-\rho^2})$, so applying (14) yields

$$\sigma_{i,i+1}^2 = p_{i,i+1}\sigma_i^2 + (1 - p_{i,i+1})\sigma_{i+1}^2 + 4\sigma^2(m_i + m_{i+1})^2 \times p_{i,i+1}(1 - p_{i,i+1})\sigma^2/\{K^2(1-\rho^2)\}. \quad (15)$$

The value of σ^2 does not affect the collapsing algorithm and can be set to 1. It remains to choose a value of ρ^2. In the example that follows, ρ^2 is set to the middle of its range (.5), but sensitivity to the choice of this parameter is assessed.

Example 1 Collapsing Post-Strata in the Los Angeles Epidemiologic Catchment Area (LA ECA) survey The 1979 Los Angeles Epidemiologic Catchment Area (LA ECA) survey of mental health status was based on an equal probability sample of households in two catchment areas: East Los Angeles and West Los Angeles (Eaton and Kessler 1985) Data were post-stratified to Census population counts by Gender (M or F), Race (H = Hispanic, N = Not Hispanic), Catchment Area (E or W), and Age. Table 1 displays raw sample counts and population counts for single year ages 18–99 for the eight subgroups formed by combinations of Gender, Race, and Catchment Area. The collapsing algorithm was applied to the age post-strata, separately for each of the eight subgroups, using (13)–(15) with $K = 99 - 18 + 1 = 82$ and $\rho^2 = .5$ Finite population corrections $(1 - f_i)$, $(1 - f_{ij})$ are set to 1

Figure 2 plots the expected posterior variance as a function of the number of post-strata for each of the eight subgroups. The extreme right point on each plot is the expected variance after collapsing only post-strata with zero counts This variance has been rescaled to the value 100, so other points on the curves represent percentages of that value. As cells are collapsed via the algorithm, the incremental change in variance (13) is used to update the expected posterior variance. Although ρ^2 is set to .5 in the collapsing algorithm, three values of ρ^2—0, .5, and .8—are used when computing the variance increments, yielding the three curves in the figures. Values of ρ^2 other than .5 provide an assessment of how the collapsing algorithm works when ρ^2 is incorrectly specified, as is inevitable if a single collapsing algorithm is applied to more than one outcome variable.

The effect of collapsing the large numbers of post-strata on the right side of the graphs is to reduce variance for all three values of ρ^2. When $\rho^2 = 0$, post-stratification simply adds variance; hence the posterior variance increases monotonically with the number of post-strata, and is a minimum at 1; that is, when Z is effectively ignored. On the other hand, when ρ^2 is non-0, the posterior variance curves upward when the number of post-strata is small, reflecting an excessive degree of collapsing. This effect is particularly noticeable when $\rho^2 = .8$. A sensible strategy is to collapse until the increase in variance from post-stratifying on Z when $\rho^2 = 0$ is modest and the upturn in the curve from excessive collapsing when ρ^2 is non-0 is avoided. Based on these considerations, a choice of about 8 to 10 post-strata seems reasonable for the examples in Figure 2.

Figure 3 displays the post-stratum weights for subgroup FNE, the left panel when collapsing is confined to the empty post-strata, and the right panel after collapsing to the 10 categories indicated by the algorithm; the plotted X-values represent the left ends of the post-stratum intervals. The algorithm reduces the variation in the weights substantially more than the standard method of truncating the weights at a fixed value, such as 3. Plots for the other groups are similar and are omitted.

One might ask why collapsing is based on the *expected* reduction in posterior variance. When data are available, why not simply compute (12) directly for each variable and collapse to reduce the actual posterior variance? Intuitively, this approach seems too sensitive to sampling fluctuations in the within-stratum means and variances, particularly when the number of post-strata is large. Also, it seems too opportunistic to use the data to determine the model in such a direct way; model selection should be based on prior information to the greatest extent possible.

2.5.2 Collapsing Strategies When Bias Is Present. When MCAR does not hold, post-stratification removes the component of bias arising from differential selection across post-strata. If post-strata i and j are collapsed, the posterior mean may be distorted if their response rates φ_i and φ_j differ; hence the strategy of the previous section needs to be modified to avoid collapsing post-strata with significantly differing response rates. The choice of modification is delicate, because such strata are prime candidates for collapsing to reduce variance; that is, there is a tension between bias increase and variance reduction.

To determine bias from collapsing i and j, note that

$$E(P_i\bar{y}_i + P_j\bar{y}_j \mid P_i + P_j, \varphi_i, \varphi_j, \text{data}) \simeq (P_i + P_j)\frac{\varphi_j r_i \bar{y}_i + \varphi_i r_j \bar{y}_j}{\varphi_j r_i + \varphi_i r_j}.$$

Collapsing yields the posterior mean from setting $\varphi_i = \varphi_j$ in this expression. The bias introduced by collapsing is thus

$$(P_i + P_j)\left(\frac{r_i\bar{y}_i + r_j\bar{y}_j}{r_i + r_j} - \frac{\varphi_j r_i \bar{y}_i + \varphi_i r_j \bar{y}_j}{\varphi_j r_i + \varphi_i r_j}\right).$$

Taking expectations and replacing φ_i/φ_j by the estimate

Table 1 Los Angeles ECA (A) Sample and (B) Census Counts by Age in Eight Stratum Groups

	Stratum group															
	1.FNE		2.FNW		3 FHE		4 FHW		5 MNE		6·MNW		7 MHE		8·MHW	
AGE	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)
18	3	193	12	721	12	1,433	4	338	2	231	8	729	6	1,517	1	403
19	1	206	6	830	11	1,581	4	402	4	227	12	819	9	1,584	2	430
20	5	197	11	998	17	1,497	2	389	1	254	7	991	15	1,670	1	468
21	5	180	13	1,058	12	1,514	2	401	1	247	13	1,128	18	1,603	5	479
22	2	218	15	1,361	13	1,606	4	372	2	252	12	1,265	22	1,675	1	499
23	2	214	23	1,582	18	1,597	10	416	2	243	23	1,656	15	1,626	5	459
24	5	226	9	1,663	20	1,460	7	394	2	233	16	1,586	18	1,535	7	473
25	3	197	17	1,675	21	1,388	0	398	4	224	30	1,672	16	1,483	6	483
26	1	183	21	1,723	14	1,288	8	368	3	212	27	1,700	13	1,317	7	401
27	2	180	21	1,778	16	1,327	4	373	3	212	33	1,800	14	1,272	5	397
28	4	181	33	1,609	18	1,176	2	313	2	193	35	1,671	17	1,143	2	364
29	6	184	21	1,760	18	1,170	9	360	3	195	43	1,811	19	1,224	4	373
30	4	196	32	1,668	15	1,153	5	312	4	201	20	1,726	12	1,028	2	324
31	4	212	23	1,570	20	1,038	4	277	1	213	22	1,668	15	991	5	325
32	2	179	28	1,611	18	993	7	269	0	203	28	1,548	12	1,005	2	274
33	4	187	31	1,469	14	954	5	316	3	198	23	1,635	12	934	2	285
34	4	147	25	1,135	11	901	3	234	5	166	23	1,286	14	830	4	252
35	1	143	26	1,280	17	817	4	210	2	148	25	1,287	16	874	6	254
36	2	144	22	1,084	12	817	3	232	2	181	21	1,148	11	729	5	263
37	7	143	20	1,059	10	789	0	216	0	137	15	1,139	18	732	3	185
38	1	127	17	836	12	693	1	178	2	143	12	941	12	676	5	197
39	1	150	20	887	9	709	2	201	2	151	15	870	11	646	3	202
40	4	161	18	759	10	688	2	178	0	146	14	812	9	633	1	182
41	4	149	16	716	2	673	2	164	1	151	8	760	15	623	2	177
42	2	138	10	691	9	634	6	170	2	157	6	814	11	598	4	220
43	2	137	14	652	9	577	3	157	4	147	12	691	5	522	3	150
44	2	162	7	648	6	655	3	139	2	137	9	731	6	555	1	179
45	1	118	6	644	5	646	3	162	0	143	13	718	5	557	1	144
46	0	145	9	618	5	609	1	153	3	148	12	657	1	490	2	125
47	1	161	8	680	10	597	1	140	0	151	6	718	4	505	1	127
48	2	161	6	625	4	595	3	147	0	140	5	616	9	563	0	133
49	2	199	7	710	5	716	3	151	5	141	5	764	4	553	1	158
50	0	175	10	758	3	731	1	150	4	159	10	768	5	612	1	151
51	1	172	6	763	5	641	1	153	0	150	4	697	4	595	1	141
52	3	213	5	846	3	723	2	122	1	173	5	719	7	574	2	112
53	2	200	8	763	4	684	2	135	2	184	11	756	6	566	0	118
54	1	229	6	825	6	645	3	145	1	211	10	778	2	572	0	119
55	3	257	18	841	8	649	1	167	6	189	6	777	5	483	1	112
56	2	275	7	829	5	609	0	105	2	188	7	824	7	560	0	92
57	2	220	9	876	7	573	3	153	4	188	4	793	4	515	0	94
58	4	241	5	843	4	550	2	93	2	224	8	814	8	465	1	82
59	2	237	8	763	7	566	1	96	1	213	2	769	6	423	0	70
60	1	243	12	765	2	508	1	95	0	196	9	760	4	416	1	81
61	3	221	10	721	6	442	1	102	3	220	8	657	7	349	2	67
62	4	237	11	736	3	383	1	115	3	189	8	652	7	344	0	66
63	4	240	7	643	8	399	0	72	2	199	8	566	0	255	0	46
64	5	244	9	588	8	387	1	67	2	212	5	647	3	309	0	63
65	2	259	5	704	7	385	0	66	4	194	7	574	3	270	0	45
66	3	269	11	569	6	408	1	70	4	184	8	480	3	232	0	51
67	7	258	6	622	3	396	0	66	1	183	4	456	3	275	0	45
68	5	211	7	507	1	370	0	54	5	200	7	445	6	272	1	42
69	5	190	5	446	7	393	3	56	0	164	4	376	1	200	0	39
70	3	198	7	422	5	323	1	48	0	165	2	342	1	191	0	35
71	2	175	7	400	3	284	0	65	1	158	6	316	0	201	1	30
72	3	211	5	389	6	295	0	58	0	141	5	296	5	221	0	80
73	2	166	3	346	3	294	1	42	1	106	3	246	1	204	0	35
74	3	182	1	313	3	281	0	28	1	105	3	263	1	176	2	28
75	5	160	4	306	1	266	1	35	4	103	5	186	1	170	0	24
76	3	134	0	270	5	255	0	51	4	123	1	227	2	160	0	22
77	2	120	3	276	1	191	0	26	3	85	1	139	3	157	0	39
78	2	114	7	235	6	179	1	23	3	72	2	136	1	149	0	14
79	4	145	3	328	4	243	0	32	1	76	3	115	3	157	0	20
80	1	102	0	219	0	138	0	16	1	61	1	102	2	112	0	10
81	3	97	4	203	2	141	0	14	1	46	1	93	0	79	0	5
82	1	91	2	177	7	94	0	17	0	50	1	72	2	59	0	5
83	2	92	0	165	3	100	0	12	1	46	1	64	0	63	0	11
84	1	73	0	141	1	81	0	15	0	42	1	60	2	55	0	6

(continued)

Table 1 (continued)

	Stratum group															
	1 FNE		2 FNW		3 FHE		4 FHW		5 MNE		6.MNW		7 MHE		8 MHW	
AGE	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)
85	2	74	1	140	0	66	0	12	1	20	0	44	0	55	0	8
86	0	57	0	112	1	55	1	12	2	33	0	48	0	25	0	4
87	0	55	0	96	1	47	0	5	1	21	0	37	1	34	0	22
88	1	48	0	89	0	34	0	5	0	17	1	32	0	14	0	7
89	1	29	0	74	0	48	0	10	0	15	1	28	1	20	0	2
90	0	30	0	75	0	29	0	3	2	12	0	24	0	13	0	5
91	0	30	0	53	0	31	0	5	0	17	0	21	0	9	0	2
92	1	21	0	45	0	24	0	5	0	10	0	11	0	5	0	2
93	0	15	0	32	0	18	0	1	0	8	0	6	0	10	0	1
94	0	21	0	33	0	10	0	2	0	2	0	7	0	5	0	1
95	0	11	0	23	0	7	0	2	0	1	0	5	0	3	0	1
96	0	9	0	12	1	8	0	4	0	3	0	4	0	2	0	2
97	0	6	0	8	0	6	0	0	0	0	0	4	0	3	0	1
98	0	2	0	6	0	4	0	0	0	0	0	0	0	5	0	1
99	0	4	0	3	0	16	0	3	0	4	0	0	0	12	0	4

NOTE F = Female, M = Male, N = Non-Hispanic, H = Hispanic, E = East Los Angeles, W = West Los Angeles

$(r_i/P_i)/(r_j/P_j)$ in this expression yields the following bias:

$$E\{\Delta m_{ij}\} = (P_i + P_j)\left(\frac{r_j}{(r_i + r_j)} - \frac{P_j}{(P_i + P_j)}\right)(\bar{\mu}_i - \bar{\mu}_j). \tag{16}$$

The reduction in expected MSE from collapsing is thus

$$E(\text{MSE}_{ij}) = E(\Delta v_{ij}) - E\{\Delta m_{ij}\}^2, \tag{17}$$

where $E(\Delta v_{ij})$ is given by (13) and $E(\Delta m_{ij})^2$ is given by the square of (16). This expression can be shown to be a refinement of Kalton and Maligalig's (1991) MSE collapsing criterion.

Example 2 (Example 1 continued) The analysis of Example 1 was repeated with (13) replaced by (17). Plots of gains against number of collapsed post-strata looked similar to Figure 2, except that losses for small number of post-strata were a bit larger, suggesting a final choice with one or two additional post-strata. Weights for the collapsed post-strata were slightly more variable than those in Figure 3, reflecting more resistance to smoothing. The only marked change in results was for subgroup 2, where the collapsed weight for post-stratum aged 83–99 was nearly 6, compared to 2 in the MCAR analysis. For these data the refinement for non-MCAR seems unnecessary, but more study of this question for other data sets seem needed.

2.6 Alternative Strategies to Collapsing

As noted earlier, the basic problem with inference based on (6) is that the non-informative prior on (μ_h, σ_h^2) provides poor predictive inferences in small post-strata. Collapsing post-strata i and j effectively assumes homogeneity of the distribution of Y in i and j, that is, $\mu_i = \mu_j$ and $\sigma_i^2 = \sigma_j^2$ Model-based alternatives to collapsing can be based on other modifications of the prior in (6).

If the stratum means can be regarded as exchangeable, they might be modeled as iid from a common distribution, yielding a random effects model With normal specifications, the prior

$$p(\mu_h \mid \sigma_h) \sim_{\text{ind}} G(\mu, \tau^2); \qquad \sigma_h^2 = \sigma^2 \quad \text{for all} \quad h;$$
$$p(\mu, \log \tau^2, \log \sigma^2) = \text{const.} \tag{18}$$

yields estimates of $\bar{Y}$ that smooth $\bar{y}_{ps}$ toward $\bar{y}$ (Scott and Smith 1969; Holt and Smith 1979; Little 1986). If variances are not assumed constant:

$$p(\mu_h \mid \sigma_h) \sim_{\text{ind}} G(\mu, \tau^2); \quad p(\mu, \log \tau^2, \log \sigma_h^2) = \text{const.},$$

the resulting posterior mean shrinks $\bar{y}_{ps}$ towards a mean where cases in post-stratum h are weighted by s_h^{-2}. These estimators are appealing compromises between probability-weighting and variance-weighting. In large samples the posterior means converge to $\bar{y}_{ps}$ and hence are design consistent. In small samples the shrinkage is greatest when the ratio of between to within variance is small, as should be the case. But simulations (Little 1991a) suggest that the inferential properties of estimates based on these models can be adversely affected by lack of exchangeability of the $\{\mu_h\}$. Thus gains in efficiency have a cost in additional model assumptions. If Z is a grouped continuous or ordered categorical variable, the exchangeable model does not seem appropriate, because one might expect the mean of an associated outcome to vary systematically with Z. In such cases a model that shrinks towards a regression line might yield better inferences. A referee notes that in other cases it may be possible to find groups of post-strata within which exchangeability is plausible. Studies comparing these approaches with the collapsing methods of the previous section may be of interest.

These procedures require a separate model for each Y. Other methods smooth the weights in each post-stratum directly. A common practical device is to place an upper limit on the weight—a value of 3 was used in Hanson (1978)—and trim higher weights to this value. Potter (1990) considers procedures for determining a good upper limit, based on iterative MSE computations. Another approach to smooth-

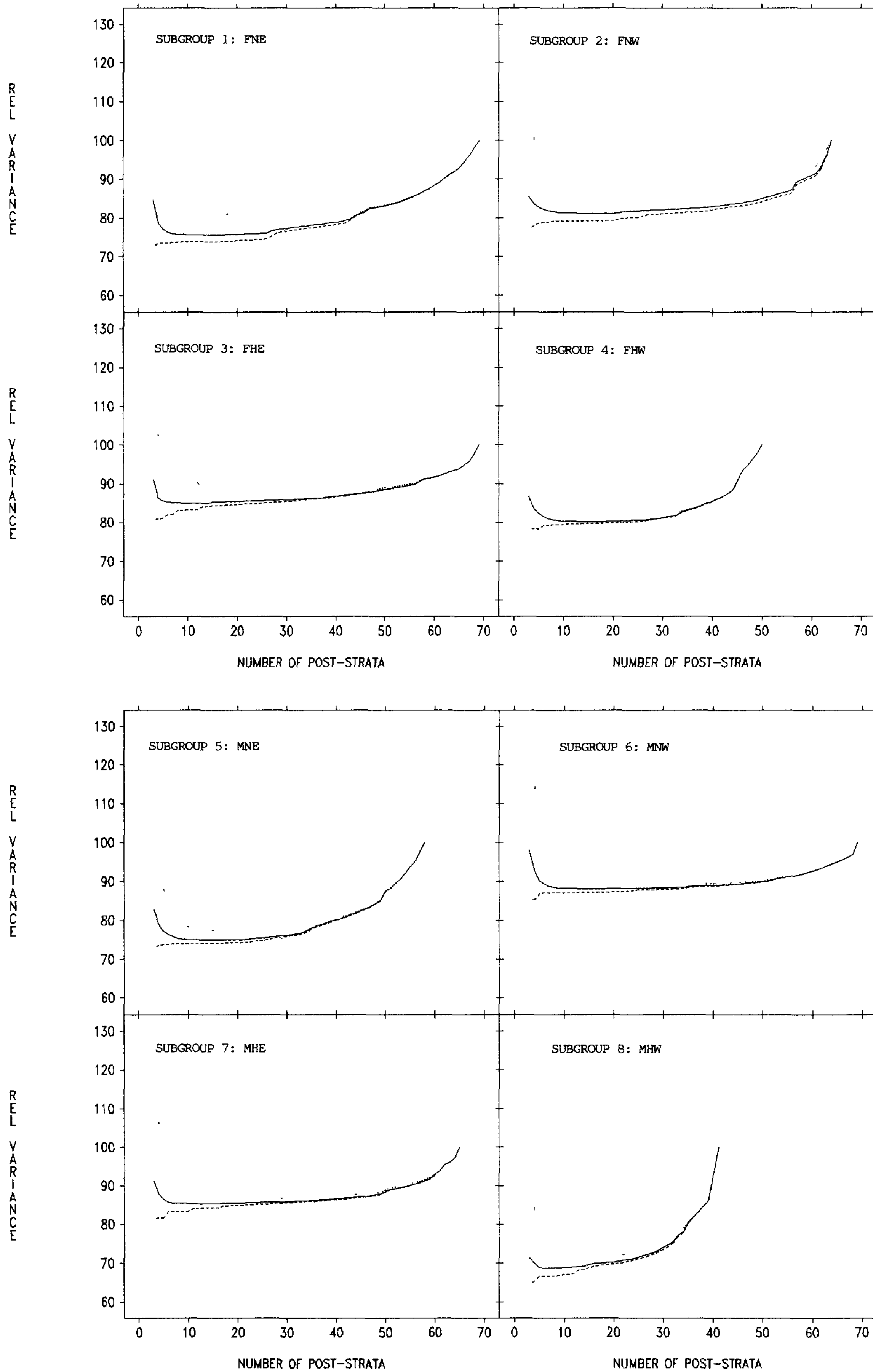

Figure 2 Relative Variance vs Number of Post-Strata, for ρ^2 = 0.0 (---), 0.5 (——), and 0.8 (· · ·)

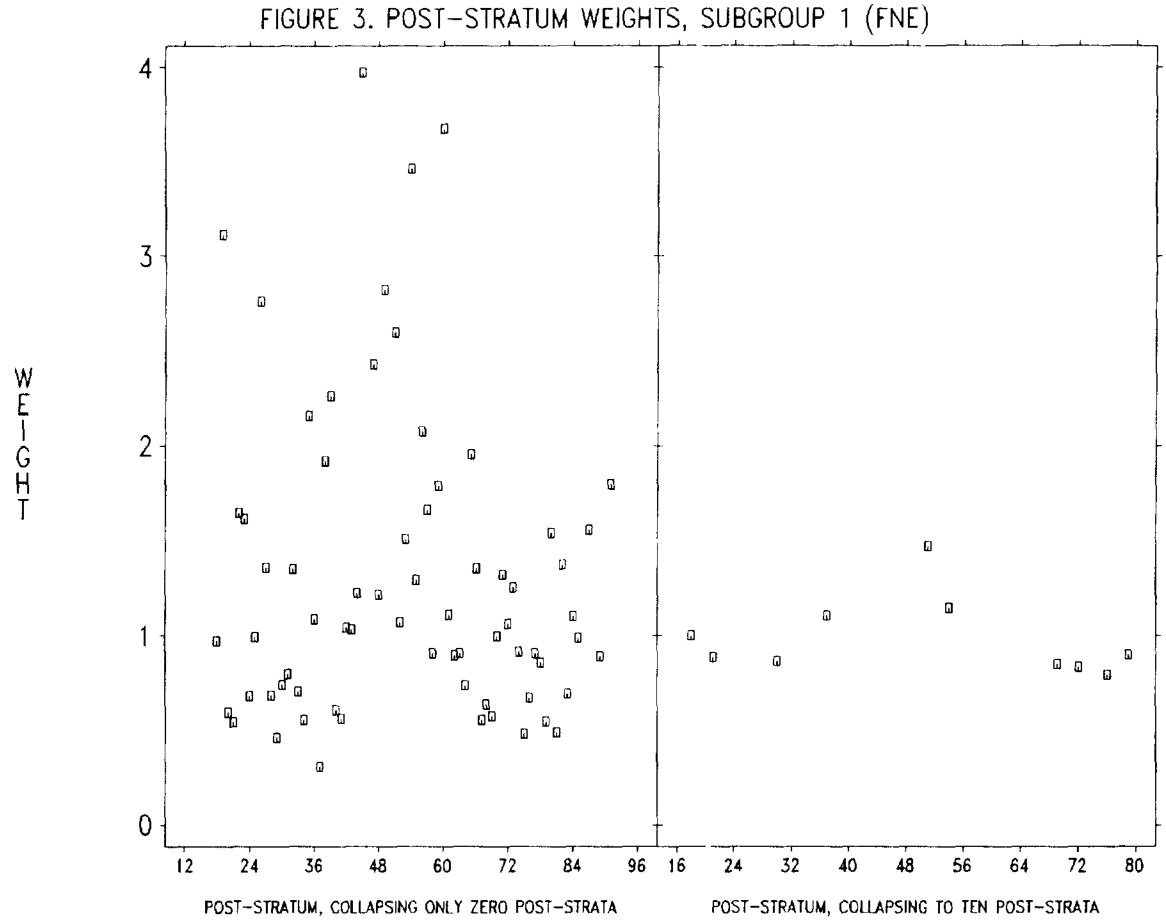

Figure 3 Post-Stratum Weights, Subgroup 1 (FNE)

ing is to model the response rates. Let δ be the overall probability of selection and let $\pi_h = \delta\varphi_h$ be the probability of selection and response in post-stratum h. Then model $r_h | N_h$, π_h as binomial with index N_h and probability π_h, and assume further that the selection rates π_h have a beta distribution with mean π and variance $\kappa\pi(1-\pi)$; in terms of the usual parameterization of the beta, $\pi = \alpha/(\alpha+\beta)$ and $\kappa = 1/(\alpha+\beta+1)$. Then the posterior distribution of π_h is beta with mean

$$E(\pi_h|\text{data}) = \tilde{\pi}_h = \lambda_h\pi + (1-\lambda_h)\frac{r_h}{N_h},$$

where $\lambda_h = (1-\kappa)/(1-\kappa+\kappa N_h)$, which smooths the observed selection rates toward π. Estimating π by r/N, smoothed weights from this model have the form

$$\tilde{w}_h \propto w_h/\{1-\hat{\lambda}_h+\hat{\lambda}_h w_h\},$$

where $\hat{\lambda}_h$ is obtained from λ_h by replacing κ by an estimate.

These approaches to weight trimming or smoothing, even if based on models for the response rates, are *not* forms of predictive modeling. As noted in Section 2.2, for predictive inference models for the nonresponse rate are irrelevant, because model uncertainty lies in the distribution of Y given Z. To illustrate the distinction, note that under MCAR, the correct model for the selection rates set $\pi_h = \pi$ and hence $\hat{\pi}_h$ equal to a constant. The resulting estimator of $\bar{Y}$ is $\bar{y}$, which ignores information in the post-strata completely; but we have seen that this is not an efficient estimator if the post-strata are predictive of the outcome.

3. TWO OR MORE POST-STRATIFIERS

3.1 Introduction

With more than one post-stratifier, the likelihood of sparse or empty post-strata increases, so the need to modify the BNPM is greater. Also, the range of possible modifications is considerably greater. Attention is restricted to the case of two post-stratifiers. Let Z_1 and Z_2 be two categorical post-stratifying variables and let P_{hk} be the population proportion with $Z_1 = h$, $Z_2 = k$. Then

$$\bar{Y} = \sum_{h=1}^{H}\sum_{k=1}^{K} P_{hk}\bar{Y}_{hk},$$

where $\bar{Y}_{hk}$ is the population mean in cell (h, k). Suppose that a random sample is taken, resulting in n_{hk} individuals in cell (h, k), r_{hk} of whom respond to Y. Model-based inference about $\bar{Y}$ involves two distinct components: inference about $\{P_{hk}\}$ based on a model for the joint distribution of Z_1 and Z_2 and inference about $\{\bar{Y}_{hk}\}$ based on a model for the distribution of Y given Z_1 and Z_2. A direct extension of

(6) for Y given Z_1, Z_2 is the basic normal two-way post-stratification model

$$(y_i | z_{1i} = h, z_{2i} = k, \mu_{hk}, \sigma^2_{hk}) \sim_{\text{ind}} G(\mu_{hk}, \sigma^2_{hk});$$
$$p(\mu_{hk}, \log \sigma_{hk}) = \text{const.} \quad (19)$$

The next section discusses inferences under this model. Section 3.3 considers modifications for small cells.

3.2 Inferences Under the Basic Normal Two-Way Post-Stratification Model

3.2.1 The Posterior Mean of $\bar{Y}$. The BNPM model (19) with $\{P_{hk}\}$ known yields

$$E(\bar{Y} | \mathbf{Z}, \mathbf{Y}_s) = \bar{y}_{\text{ps}} = \sum_h \sum_k P_{hk}\bar{y}_{hk}, \quad (20)$$

the post-stratified mean for the combined post-strata. In some circumstances $\{P_{hk}\}$ is unknown but the marginal distributions $\{P_{h+}\}$, $\{P_{+k}\}$ of Z_1 and Z_2 are known; then a model is needed to predict $\{P_{hk}\}$. Model inference for $\bar{Y}$ then depends on what is known about $\mathbf{n} = \{n_{hk}\}$ (Little 1991b).

Suppose first that Z_1 and Z_2 are recorded for nonrespondents, so $\mathbf{n}$ is observed. Under random sampling, a natural model for $\mathbf{n}$ is

$$\{n_{hk}\} | n \sim \text{MNOM}[\{P_{hk}\}, n], \quad (21)$$

the multinomial distribution with index n and probabilities $\{P_{hk}\}$ With the Jeffreys's prior on $\{P_{hk}\}$, the resulting posterior distribution of $\{P_{hk}\}$ given $\{n_{hk}\}$ is Dirichlet with parameters $\{n_{hk} + 1/2\}$. The posterior distribution of $\{P_{hk}\}$ given $\{P_{h\cdot}\}$, $\{P_{\cdot k}\}$, and $\{n_{hk}\}$ does not have a simple form But its posterior mean can be approximated as

$$\{\hat{P}^{(1)}_{hk}\} = \text{Rake}[\{n_{hk}\}; \{P_{h\cdot}\}, \{P_{\cdot k}\}], \quad (22)$$

which denotes the result of *raking* the sample counts $\{n_{hk}\}$ to the known margins. Raking means iterative proportional fitting (Deming and Stephan 1940) of $\{n_{hk}\}$ to match successively the row and column marginal distributions $\{P_{h\cdot}\}$ and $\{P_{\cdot k}\}$; (22) approximates the posterior mean of $\{P_{hk}\}$ because it is asymptotically equivalent to the maximum likelihood (ML) estimate (Brackstone and Rao 1979; Ireland and Kullback 1968), which is in turn asymptotically equivalent to the posterior mean. Note that (22) involves the sample counts, $\mathbf{n}$, which do not enter into inferences about $\bar{Y}$ for the case of a single post-stratifier.

If either Z_1 or Z_2 is subject to nonresponse as well as Y, then the sample counts $\{n_{hk}\}$ are not known. A model is then required to relate the respondent counts $\{r_{hk}\}$ to the target population. Suppose that

$$\{r_{hk}\} | r \sim \text{MNOM}[\{P_{hk}\varphi_{hk}/\bar{\varphi}\}, r], \quad (23)$$

where φ_{hk} is the probability of response in cell (h, k) and $\bar{\varphi} = \sum_h \sum_k P_{hk}\varphi_{hk}$. This model is underidentified given the available data. But Little and Wu (1991) showed that the estimates based on raking $\{r_{hk}\}$ to the margins, namely

$$\{\hat{P}^{(2)}_{hk}\} = \text{Rake}[\{r_{hk}\}; \{P_{h\cdot}, P_{\cdot k}\}], \quad (24)$$

are ML under the assumption that

$$\varphi_{hk} = \alpha_h\beta_k; \quad (25)$$

that is, the response rate is a product of row and column effects (see also Binder and Theberge 1988, sec. 3) Little and Wu showed that the assumptions about the φ_{hk} are untestable from the data, and that other choices lead to other estimation methods.

Combining these raking estimates with estimates of $\{\bar{Y}_{jk}\}$ from (19) yields the posterior mean of $\bar{Y}$ as

$$\hat{y} = \sum_{h=1}^{H} \sum_{k=1}^{K} \hat{P}_{hk}\bar{y}_{hk}, \quad (26)$$

where $\{\hat{P}_{hk}\}$ is given by (22) or (24), depending on the data at hand.

3.2.2 The Posterior Variance of $\bar{Y}$. If $\{P_{hk}\}$ is known, then the posterior variance of $\bar{Y}$ under (19) is

$$\text{var}(\bar{Y} | \mathbf{Z}, \mathbf{Y}_s) = v_{\text{ps}} = \sum_h P^2_{hk}(1 - f_{hk})\delta_{hk}s^2_{hk}/r_{hk}, \quad (27)$$

where in post-stratum hk, $f_{hk} = r_{hk}/N_{hk}$, s^2_{hk} is the sample variance of Y, and $\delta_{hk} = (r_{hk} - 1)/(r_{hk} - 3)$. If $\{P_{hk}\}$ is unknown, then the posterior variance can be written as

$$\text{var}(\bar{Y}/d) = E[v_{\text{ps}} | d] + \text{var}[\bar{y}_{\text{ps}} | d], \quad (28)$$

where d stands for the data. The first term on the right side of (28) is approximated by replacing $\{P_{hk}\}$ in (27) by estimates under the model. The second component reflects the added variance from uncertainty in the $\{P_{hk}\}$ and can be written as

$$\text{var}(\bar{y}_{\text{ps}} | d) = \sum_h \sum_k \sum_l \sum_m \bar{y}_{hk}\bar{y}_{lm}\text{cov}(P_{hk}, P_{lm} | d). \quad (29)$$

For the raking estimates, the covariance matrix $\text{cov}(P_{hk}, P_{lm} | d)$ can be approximated by the asymptotic expressions derived in Binder and Theberge (1988). Another approach, which does not rely on asymptotics, is to simulate the posterior covariance matrix by computing T draws $\{P^{(t)}_{hk}\}$ ($t = 1, \ldots, T$) from the posterior distribution of $\{P_{hk}\}$, computing post-stratified means $\bar{y}^{(t)}_{\text{ps}}$ by substituting $\{P^{(t)}_{hk}\}$ for $\{P_{hk}\}$ in $\bar{y}_{\text{ps}}$, and then estimating (29) by the sample variance of $\bar{y}^{(t)}_{\text{ps}}$ over $t = 1, \ldots, T$. For the model (23) that leads to raking the respondent counts, the appropriate draws are computed as

$$P^{(t)}_{hk} = \text{Rake}[\{\rho^{(t)}_{hk}\}; \{P_{h\cdot}, P_{\cdot k}\}], \quad (30)$$

where $\{\rho^{(t)}_{hk}\}$ are drawn from a Dirichlet distribution with parameters $\{r_{hk} + 1/2\}$. For the model (21) that leads to raking the sample counts $\{n_{hk}\}$, an approximate procedure is to apply (24) where $\{\rho^{(t)}_{hk}\}$ are drawn from a Dirichlet distribution with parameters $\{n_{hk} + 1/2\}$. (Unfortunately, in this case the method does not yield exact draws from the posterior distribution.) These simulation methods avoid computing and inverting the covariance matrix of $\{P_{hk}\}$ and hence may be useful when this matrix has high dimension, as in large tables.

3.3 Treatment of Small Cells

In this section I outline modifications of the inferences under the BNPM for two post-stratifiers. A principled Bayesian approach is to change the flat prior in (19) for Y

given Z_1 and Z_2 to borrow strength across post-strata. Modifications should depend on context. One approach of interest is to model high-order interactions of Y on Z as random effects. For example, write

$$\mu_{hk} = \mu + \alpha_h + \beta_k + \gamma_{hk}$$

and replace the prior for $\{\mu_{hk}\}$ in (19) by a prior that is flat for $\{\mu, \alpha_h, \beta_k\}$ but models the interactions γ_{hk} as normal with mean 0 and variance σ_γ^2. The result is a mixed ANOVA model with fixed main effects and random interactions. The resulting prediction for $\bar{Y}_{hk}$ (ignoring finite population corrections) has the form

$$\tilde{y}_{hk} = u_{hk}\bar{y}_{hk} + (1 - u_{hk})\hat{y}_{hk},$$

where $\hat{y}_{hk} = \hat{\mu} + \hat{\alpha}_h + \hat{\beta}_k$ is an additive fit and $u_{hk} = n_{hk}\hat{\sigma}_\gamma^2/(n_{hk}\hat{\sigma}_\gamma^2 + \hat{\sigma}_{hk}^2)$ is a shrinkage factor, with estimates $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_{hk}^2$ of the random effects σ_γ^2 and σ_{hk}^2 computed by ML or the method of moments.

Modifications of the prior in (19) should be tuned to each outcome, and hence they may be impractical in large surveys with many variables. Simpler strategies are based on ignoring partial information in the post-strata. A common strategy is to rake sample or respondent counts to the margins, effectively ignoring information on the joint distribution of Z_1 and Z_2 when this is available; raking the sample counts seems preferable, because no model is implied for the response probabilities Raking is well justified when the means of outcome variables have an approximately additive structure, because if $\bar{Y}_{hk} = \mu + \alpha_h + \beta_k$, then $E(\bar{Y}|\text{data}) = \hat{\mu} + \sum_h P_{h\cdot}\hat{\alpha}_h + \sum_k P_{\cdot k}\hat{\beta}_k$ involves only the marginal distributions of Z_1 and Z_2, so raking does not involve a loss of information. Another way of seeing this is to note that Equation (29) for the added uncertainty from raking remains valid with $\{\bar{y}_{hk}\}$ replaced by $\{\bar{y}_{hk} - \hat{y}_{hk}\}$, where $\{\hat{y}_{hk}\}$ are predicted values from any additive fit. If the additive fit is good, then the added variance (29) must be small

If interactions are expected, then a better approach may be to apply collapsing strategies as discussed in Section 2. Let Z_1 denote a "primary" stratifier most closely related to survey outcomes, and let Z_2 denote a secondary stratifier. Then the collapsing methods of Section 2 could be applied to Z_2 classes separately for each value of Z_1 (as in Example 1) or to the single classification of Z_1 and Z_2 obtained by laying out the cells in serpentine order, with the Z_2 index changing fastest.

If the MCAR assumption is violated, as in the case of nonresponse, then collapsing strategies should focus on limiting nonresponse bias. One way of achieving this is to regress response rates on the post-stratifying variables and then form collapsed post-strata that are homogeneous with respect to the estimated response rate (Little 1986; Rosenbaum and Rubin 1983). This strategy seems particularly useful when the number of post-stratifiers is large. For a recent application, see Goksel, Judkins, and Mosher (1991)

4. CONCLUSION

The main aims of this article are to lay out the Bayesian predictive modeling framework for post-stratification, to provide a Bayesian interpretation of the problem of how to assess precision of the post-stratified mean, to suggest some approaches for the problem of sparse cells, and to provide a Bayesian perspective on raking. Many important topics are left untouched Inference about other quantities, such as population slopes, has not been considered. The survey practitioner can rightly complain that deviations from simple random sampling are restricted to nonresponse and coverage error—models need to be adapted to complex survey designs. Another topic of real practical interest concerns post-stratification to margins that are subject to error, as in an income classification where the census and survey definitions differ modestly I believe that the Bayesian approach provides a framework for addressing this problem, because progress is hard without some working model for the pattern of errors. Another interesting topic untouched here is raking to totals of an auxiliary variable (such as dollar amounts) rather than to sample counts. The algebra for this form of raking is closely related to that for raking to sample counts (see, for example, Binder and Theberge 1988), but the underpinning models appear quite different Even statisticians unpersuaded by the modeling view of surveys can benefit from a study of these underlying models, because they case light on how to choose between alternative analysis approaches, which in the absence of models can appear capricious For modelers, I suggest that post-stratification in particular, and survey methods in general, remain rich and somewhat neglected areas with many problems still to be resolved.

[Received October 1991 Revised September 1992]

REFERENCES

Binder, D A, and Theberge, A (1988) "Estimating the Variance of Raking Ratio Estimators Under Simple Random Sampling," *Canadian Journal of Statistics*, 16, 47–55

Brackstone, G J, and Rao, J N K (1979), "An Investigation of Raking-Ratio Estimators," *Sankhya, C*, 41, 97–114

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

Deming, W E, and Stephan, F F (1940), "On a Least Squares Adjustment of a Sample Frequency Table When the Expected Marginal Totals Are Known" *Annals of Mathematical Statistics*, 11 427–444

Eaton, W W, and Kessler, L G, eds (1985), *Epidemiologic Field Methods in Psychiatry The NIMH Epidemiologic Catchment Area Program*, New York Academic Press

Ericson, W A (1969), "Subjective Bayesian Models in Sampling Finite Populations, I," *Journal of the Royal Statistical Society, Ser B*, 31, 195–234

Goksel, H, Judkins, D R, and Mosher, W D (1991), "Nonresponse Adjustments for a Telephone Follow-up to a National In-Person Survey," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp 581–586

Harte, J M (1982), "Post-Stratification Approaches in the Corporate Statistics of Income Program," in *Proceedings of the Survey Research Methods Section, American Statistical Association 1982*, pp 250–253

Hanson, R H (1978), *The Current Population Survey Design and Methodology* Technical Paper No 40, U S Bureau of the Census

Holt, D, and Smith, T M F (1979), "Post Stratification," *Journal of the Royal Statistical Society, Ser A*, 142, 33–46

Ireland, C T, and Kullback S (1968), "Contingency Tables With Given Marginals" *Biometrika*, 55, 179–188

Kalton, G, and Maligalig, D S (1991), "A Comparison of Methods of Weighting Adjustment for Nonresponse," *Proceedings of the 1991 Annual Research Conference U S Bureau of the Census*, pp 409–428

Little, R J A (1982), "Models for Nonresponse in Sample Surveys," *Journal of the American Statistical Association* 77, 237–250

——— (1986), "Survey Nonresponse Adjustments for Estimates of Means," *International Statistical Review*, 54, 139–157

——— (1991a) "Inference With Survey Weights" *Journal of Official Statistics*, 7, 405–424

——— (1991b), Discussion of Session "Estimation Techniques With Survey Data," *Proceedings of the 1991 Annual Research Conference, U S Bureau of the Census*, 441–446

Little, R J A, and Wu, M M (1991), "Models for Contingency Tables With Known Margins When Target and Sampled Populations Differ," *Journal of the American Statistical Association*, 86, 87–95.

Oh, H L, and Scheuren, F J. (1983), "Weighting Adjustment for Unit Nonresponse," in *Incomplete Data in Sample Surveys, Volume 2—Theory and Bibliographies*, eds W. G Madow, I. Olkin, and D. B Rubin, New York Academic Press, pp 435–483.

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

Potter, F (1990), "A Study of Procedures to Identify and Trim Extreme Sampling Weights," *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp 225–230

Rosenbaum, P R, and Rubin, D B (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55

Rubin, D. B (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592.

——— (1984), "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician," *The Annals of Statistics*, 12, 1151–1172

Scott, A., and Smith, T M F (1969), "Estimation in Multistage Sampling," *Journal of the American Statistical Association*, 64, 830–840.

Tremblay, V (1986), "Practical Criteria for Definition of Weighting Classes," *Survey Methodology*, 12, 85–97.

Waterton, J., and Lievesley, D. (1987) "Attrition in a Panel Study of Attitudes," *Journal of Official Statistics*, 3, 267–282