SYNTHETIC ESTIMATES FOR SMALL AREAS IN PRAMS

Christopher H. Johnson and the PRAMS Working Group States
Christopher H. Johnson, MS-K21, 4770 Buford Hwy NE, Atlanta, GA 30341-3724

Key Words: Small Area Estimation, Standardized Rates, Behavioral Surveys

The Pregnancy Risk Assessment Monitoring System, or PRAMS, is an ongoing, population-based surveillance system designed to supplement vital records data on maternal behaviors and to generate state-specific data for planning and assessing perinatal health programs. This Federal-state cooperative program is funded in part by the Centers for Disease Control and Prevention (CDC). Within state health departments, PRAMS program structures cross several existing organizational units, such as those that deal with maternal and child health and vital records.

The survey began with the District of Columbia, Oklahoma, Indiana, Maine, Michigan, and West Virginia in 1988, then added Alaska in 1990 and Alabama, California, Florida, Georgia, New York, South Carolina, and Washington in 1991. Each participating state draws a stratified random sample of birth certificate records for women who have recently delivered a live-born infant. Annual sample sizes range from 1700 to 3400, divided among three to six strata. The questionnaire solicits information about barriers to and content of prenatal care, the mother's experiences and environment during pregnancy, and early infant development. PRAMS staff collect data through statewide mailings that are followed up with telephone calls for nonrespondents.

States participating in PRAMS have long needed data for substate areas to plan and evaluate programs. Sample sizes are often inadequate for producing sufficiently precise risk factor estimates by counties and other small areas. Typically, the annual sample is large enough for estimating statewide risk factor proportions within 3.5% with a 95% confidence interval, and stratum estimates are less precise. States need data on maternal and infant health to evaluate their statewide public health programs, but they also need data for targeting specific areas within states.

The four states whose data we used in this study vary in the way their constituent substate areas divide the state. South Carolina and Indiana group counties into 13 health regions based on geography. Alaska has 8 family health plan regions that are aggregations of census areas (Alaska does not have counties). Maine, a small state to begin with, has 16 relatively large counties that combine disparate rural and urban areas, making counties internally heterogeneous. Township codes, which were unfortunately not available, would further subdivide counties. Because there is thought to be little variability among Maine counties compared with their within-county variability, the relationship between demographic predictors and outcomes is attenuated, and none of the estimators was expected to do well for Maine.

BACKGROUND

The modeling approaches developed for generating small area estimates are most appropriate in areas that embody substantial demographic variability among the areas for which estimates are wanted. The idea behind synthetic estimation is that most differences among areas arise from their differing from the whole state in the composition of mothers by such demographic characteristics as age, race, education, and marital status. Women possessing these characteristics in common, but living in different areas, are assumed to be more like each other than like women with different characteristics in the same area.

Schaible, Brock, Casady, and Schnack (1979) described an empirical study of several ways to make estimates for small areas in the National Health Interview Survey. Their work undoubtedly popularized the use of synthetic estimation in health and demographic surveys, and it stated the potential difficulty in synthetic estimation: "the squared error of the synthetic estimator is subject only to a small sampling variance inherent in the estimated large area mean but is usually dominated by a bias component which is independent of sample size." Trying to get a handle on this bias component has been the focus of many subsequent investigations of the synthetic estimator and its competitors.

Holt, Smith, and Tomberlin (1979) phrased the problem in a predictive framework and explicitly laid out the model assumptions. They talked about "the constancy of the assumed relationship that forms the basis for borrowing information from one small area to another" and stressed the need for empirical investigation of the relationship, upon which the usefulness of any of these estimators depends.

Särndal (1984) developed a generalized regression estimator that borrows information in the same way that the synthetic estimator does but is design-consistent (unbiased for the sample design). He explained the tradeoff in mean squared error for estimators and the interplay between properties of the sample (sample size and fractions) and of the population (domain size and the adequacy of the model on which synthetic estimation is based). Design-based variances are available for this estimator, although not pursued here. Evaluating its performance and ease of use was of particular interest in these comparisons.

DATA USED

The present work is partly motivated by our wish to take advantage of frame data in estimation. We routinely use frame data in poststratifying to come up with nonresponse weights. We would like to exploit the accurate counts of the number of births and the corresponding auxiliary information available to us, in particular the distributions by the demographic characteristics mentioned previously.

Vital records, taken from a census of birth certificate records, were the source of the population values for the four states studied. These known totals were the benchmarks used to judge the performance of the estimators.

The two variables of primary interest are the behavioral risk factors of smoking during pregnancy and late initiation of prenatal care (late PNC). Prenatal care begun after the first trimester (13 weeks) of pregnancy is considered to have been initiated late. In these four states, smoking during pregnancy is common, and failing to initiate PNC during the first trimester is less common. Logistic regression shows that the demographic characteristics have significant predictive power (which is hardly surprising, given the huge sample size). Many studies have established the association of mother's young age, non-white race, less than high school education, and unmarried status with late PNC, low birth weight (LBW), and other negative outcomes affecting the newborn or mother.

The two variables of secondary interest are the negative and rare outcomes low birthweight (less than 2500g) and prematurity (gestation less than 37 weeks). Again, the demographic classifiers are highly significant predictors in logistic regression.

We used the most recent weighted data available: 1992 births for Alaska, Indiana, and Maine, 1993 births for South Carolina. As part of the weighting process, we calculated nonresponse weights for respondents. Multiplying the nonresponse weight by the basic sampling weight yields the analysis weight. For each of these states, we also carried out a frame omission study, but we did not include that component of the weight because frame deficiencies are another issue entirely. From the results of the frame omission studies, however, we know that no serious errors in frame construction occurred for these states.

Table 1: States and Years Included and Their Stratification Schemes, Sampling Fractions, and Sample Sizes

Alaska (race x adequacy of prenatal care), 1992		
AK Native, inadequate PNC	f = 3/5	n = 575
AK Native, adequate PNC	f = 2/5	n = 598
Nonnative, inadequate PNC	f = 1/3	n = 662
Nonnative, adequate PNC	f = 1/12	n = 568
Unknown race or PNC	f = 1/1	n = 140
Indiana (race X birthweight), 1992		
Black, LBW	f = 1/2	n = 495
Black, NBW	f = 1/9	n = 868
Nonblack, LBW	f = 5/47	n = 378
Nonblack, NBW	f = 1/139	n = 479
Unknown race or BW	f = 1/1	n = 130
Maine (birthweight), 1992		
LBW	f = 3/4	n = 519
NBW	f = 1/17	n = 835
Unknown BW	f = 1/1	n = 77
South Carolina (birthweight X region), 1993		
VLBW*, Healthy Start region	f = 1/1	n = 48
VLBW, rest of state	f = 1/1	n = 453
LBW, Healthy Start region	f = 1/1	n = 83
LBW, rest of state	f = 2/15	n = 249
NBW, Healthy Start region	f = 2/15	n = 276
NBW, rest of state	f = 1/77	n = 650

* VLBW: < 1500g; LBW: < 2499g; NBW: $\geq$ 2500g

The demographic groups used in estimation are derived from the cross-classification of records by mother's age (less than 20 years, 20 to 29 years, and 30 years or over), mother's education (some high school, high school graduate, and women with some college), mother's marital status (married and other), and mother's race (black or nonblack for Indiana and South Carolina; Alaska Native or nonnative for Alaska). What would be a 3X3X2X2 classification with 36 cells is reduced to 32 cells by combining the few under-20 mothers having some college with their same-age compatriots having only a high school diploma. The stratification scheme for Maine is further collapsed to 16 cells because nearly 99% of births were to white women in 1992.

In choosing these demographic factors for groups, we checked their predictive value through logistic regressions of frame data by state. The tremendous sample size made it possible for us to carry out extensive multivariate modeling. All the demographic factors are highly significant predictors in first-order models for the outcome variables considered.

ESTIMATORS CONSIDERED

Let y_{ijk} denote the observed value of the k^{th} unit in the i^{th} area and the j^{th} demographic group, with w_{ijk} its associated weight. N_{ij} is the known total number of units in cell ij. With a period used to denote aggregation over groups, $p_{i\cdot}$ is the proportion of units in area i that have the characteristic of interest, and similarly for $p_{\cdot j}$ in group j. The estimated proportions for direct, synthetic, and generalized regression (GREG) estimators are given by:

Direct:

$$\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}} w_{ijk}y_{ijk}/N_{i.}$$

Synthetic :

$$\sum_{j=1}^{J} (N_{ij})\,\hat{p}_{.j}/N_{i.}$$

GREG:

$$[\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}} y_{ijk} + \sum_{j=1}^{J} (N_{ij}-n_{ij})\,\hat{p}_{.j}]\,/N_{i.}$$

The GREG estimator adds to the known sample total an estimate for the nonsampled population, and it produces this estimate just as the synthetic estimator does. The synthetic estimator can be viewed as a limiting case of the GREG estimator: When one has no sample in area i, or perhaps some reason not to use the sample data directly, the double sum drops out and the size of the nonsampled population, $N_{ij} - n_{ij}$, reduces to N_{ij}. This situation represents complete reliance on the model-based component of the estimator. On the other hand, $N_{ij} - n_{ij}$ decreases as the sample size in area i increases, and the estimator assigns more importance to the sample data.

We estimate proportions for the variable of interest for each small area, then compare these to the known proportions from vital records. Note that we are using only vital records (VR) data, not questionnaire data, though all the files are linked and questionnaire variables are available. However, we purposely included variables that are also on the questionnaire (smoking during pregnancy, timing of first PNC visit). A separate study not involving small area analysis is in progress comparing PRAMS data with VR data when items are available from both sources. Previous work suggests that use of both tobacco and alcohol (about which PRAMS also asks) during pregnancy is underreported by VR. That study, which includes several other data items common to both VR and the PRAMS questionnaire, will enrich our understanding of how responses to PRAMS differ from VR data and how best to incorporate auxiliary information in estimation.

RESULTS

In evaluating estimators, one is concerned with bias. The direct estimator is unbiased. As the sample size increases, the estimate converges to the true value of the parameter for the area. The synthetic estimator has a bias that does not diminish as the sample size in the area increases; its bias depends, rather, on model adequacy. The regression estimator is unbiased, so long as its application properly accounts for the sample design. Table 2 (next page) presents bias results for all characteristics and all states.

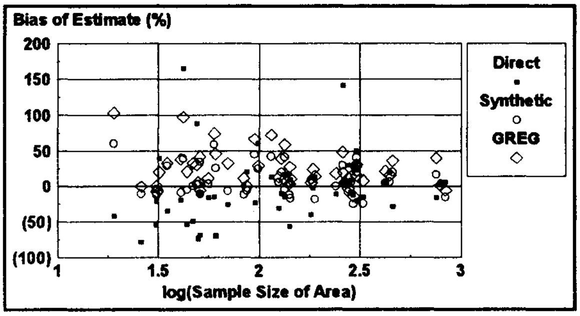

Figure 1: Relative Bias of Estimates by log(Sample Size), Premature Births

Table 2: Relative Bias of Estimates (Averaged Across Areas)

	Direct	Synthetic	GREG
Late PNC			
AK	1.16%	4.20%	20.10%
ME	3.80%	2.75%	3.84%
IN	-6.07%	-4.31%	-4.25%
SC	-4.18%	-5.19%	-4.40%
Prematurity			
AK	5.70%	6.98%	8.07%
ME	-5.51%	-7.63%	22.40%
IN	-2.91%	-2.14%	5.74%
SC	11.18%	11.77%	29.25%
Smoking			
AK	5.04%	6.61%	7.27%
ME	1.82%	3.59%	4.79%
IN	-16.17%	-15.47%	-15.20%
SC	-0.24%	-0.96%	-0.08%
LBW			
AK	10.53%	13.16%	11.25%
ME	2.50%	2.81%	62.45%
IN	2.37%	2.97%	18.03%
SC	0.61%	0.74%	26.94%

The bias presented in Figure 1 is the difference between the estimate of premature births and the actual value, divided by the actual value, for areas in all four states. This relative bias is plotted against the base-ten logarithm of the sample size of the area. Because stratification is not by area, the sample size of an area is effectively proportional to the number of births in the area. The direct estimator does best when the sample size is large; i.e. 500 or more. The points representing the relative bias of direct estimates tend to lie within a band that narrows with increasing sample size (see Figure 1).

Bias points for the synthetic estimator, on the other hand, tend to fall in a band of constant width. Although the synthetic estimator does comparatively well for the smallest areas, it does not exploit the large samples available in the larger areas. Data are used only indirectly -- through their contribution to statewide estimates.

Bias points for the GREG estimator are almost all positive. The nonsampled population contains a lower proportion of LBW infants than does the general population because of the sampling designs in three states, and premature births are responsible for a large share of low weight births. When the term $(N_{ij} - n_{ij})$ is multiplied by the estimate of the proportion of premature births in demographic cell ij, it overestimates the total. Although n_{ij} may be small in comparison to N_{ij}, it represents a disproportionately high proportion of units possessing the characteristic estimated because stratification is not by the demographic classifiers.

Estimating LBW in a situation where stratification is by LBW is something one would never do in practice: stratification requires auxiliary information known for all frame units and need not be estimated. The GREG estimator has the same shortcoming in estimating characteristics that are highly correlated with LBW, such as prematurity. In such a situation, some type of ratio estimation without the model structure might better exploit the correlation.

As another measure of the quality of estimation, we checked the correlation of estimates with actual values. A biased estimator may be acceptable if its variance is low. If the bias is constant, Schaible et al suggest that direct estimates may be better for looking at differences between areas. If the bias subtracts out, one ends up with smaller MSE for the comparison.

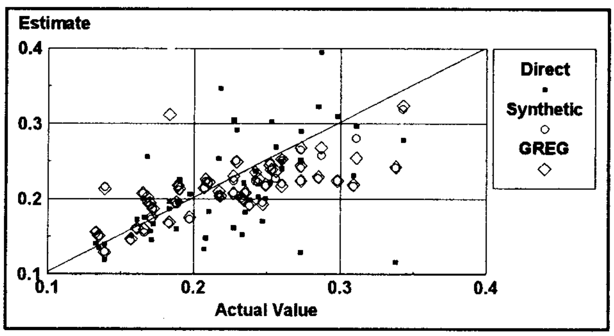

Figure 2: Estimated vs. Actual Proportion, Premature Births

In all instances but one, the GREG estimator performs better than (or at least as well as) the other estimators (see Table 3). The exception was in estimating LBW, where GREG overestimation was the worst, and even there it displays better correlation than the direct estimator.

Table 3: Correlation of Estimates with Actual Proportions

	Direct	Synthetic	GREG
Late PNC	0.81	0.82	0.84
Smoking	0.70	0.71	0.84
LBW	0.78	0.88	0.84
Prematurity	0.65	0.79	0.79

As a measure of robustness, we looked at the ratios of estimates to actual values. If neither over- nor underestimation is severe, both extremes will be close to one. The graph below (Figure 3) shows how the estimators performed in this respect for the prevalence of late PNC. The low end of the range represents the worst underestimation for some area of the state, and the high end corresponds to the area with the worst overestimation.

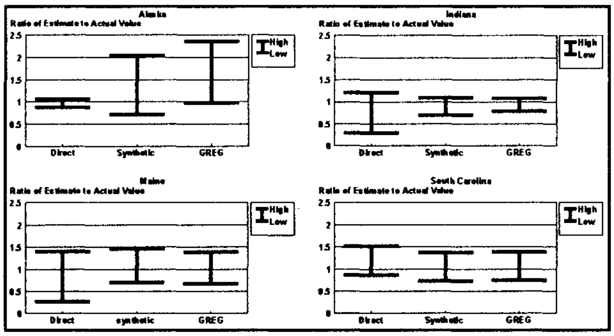

Figure 3: Range of Estimates, Late Prenatal Care

In Alaska, the high and low values of the range for the direct estimator are both close to one (see Figure 3 and Table 4). These values are more spread out for the synthetic estimator, reflecting poor fit of the model in some areas (at least one). Values for the GREG estimator are spread out over the same range as the values for the synthetic estimator, but shifted upward. A look back at Table 1 illustrates that Alaska oversamples women who received inadequate prenatal care, among whom women who began their prenatal care late make up a majority. This oversampling accounts for the superior performance of the direct estimator and for the upward bias of the GREG estimator, as discussed earlier.

The situation in the other three states is typical of the performance of the estimators: the direct estimator displays the greatest range of estimates, usually because there are a few areas with small samples that do not accurately represent the actual situation. Although they may not be better on average than direct estimates in terms of bias, synthetic estimates are less likely to deviate wildly from the actual values. The GREG estimates can be thought of as a compromise -- the model component can moderate an extreme sample value, or the sample data can exert influence to pull the model-based value toward the actual value.

Table 4: Range in Ratio of Estimated to Actual Value

	Direct	Synthetic	GREG
Late PNC			
AK	(0.88 - 1.06)	(0.72 - 2.04)	(0.98 - 2.36)
ME	(0.29 - 1.21)	(0.71 - 1.09)	(0.79 - 1.08)
IN	(0.27 - 1.40)	(0.70 - 1.46)	(0.68 - 0.68)
SC	(0.86 - 1.52)	(0.74 - 1.37)	(0.75 - 1.38)
Smoking			
AK	(0.85 - 1.17)	(0.38 - 1.83)	(0.79 - 1.54)
ME	(0.34 - 1.59)	(0.70 - 1.03)	(0.71 - 1.04)
IN	(0.71 - 1.52)	(0.85 - 1.55)	(0.84 - 1.53)
SC	(0.84 - 1.19)	(0.80 - 1.18)	(0.81 - 1.17)
LBW			
AK	(0.81 - 1.44)	(0.81 - 1.35)	(0.88 - 1.29)
ME	(0.72 - 1.45)	(0.95 - 1.49)	(1.08 - 1.62)
IN	(0.65 - 1.37)	(0.71 - 1.50)	(1.00 - 1.97)
SC	(0.75 - 1.09)	(0.95 - 1.13)	(1.16 - 1.33)
Prematurity			
AK	(0.80 - 1.60)	(0.75 - 1.41)	(0.86 - 1.31)
ME	(0.27 - 1.30)	(0.85 - 1.34)	(0.93 - 1.42)
IN	(0.21 - 2.65)	(0.76 - 1.60)	(1.00 - 2.03)
SC	(0.72 - 2.41)	(0.88 - 1.59)	(1.03 - 1.74)

CONCLUSIONS / DISCUSSION

We were particularly interested in evaluating synthetic estimates because they can be produced as poststratified estimates in SUDAAN. Since PRAMS states are already using this software for data analysis, small area estimation via SUDAAN should be the simplest way for them to proceed.

In comparing estimators, our concern about ease of application led us to impose on ourselves the same limitation to which PRAMS states are subject; namely, we worked with the samples they drew. This was an empirical study, and we did not engage in repeated sampling from the frame. Future work might include the same comparisons for all years available to see whether trends (in how the estimators compare) hold up over time.

For the larger areas, we can recommend direct estimates without hesitation. The direct estimator uses available information in the optimal way. With adequate samples, variances will be small and one need not even consider a biased estimator. For the small areas, however, direct estimates have the potential for wildly misleading results that do occasionally occur with inadequate samples. One loses the potential for discerning an area where something dramatically novel is happening, but guards against false conclusions about such differences inferred from an aberrant sample.

LIMITATIONS

Results from repeated sampling might have strengthened the comparisons of the estimators. Having studied a handful of characteristics in a few states, we find it difficult to generalize results to other situations. We used data for only four states, and only the samples they drew -- which is what they have to work with.

We have not presented any variance calculations here. The standard errors for direct estimates varied tremendously with the size of the area. Comparing them to the standard errors of synthetic estimates is problematic. Synthetic estimation has been criticized as an ad hoc method, and some object to forming confidence intervals about biased estimates. SUDAAN, for example, will report the variance of a poststratified estimate, and it will be small because the estimate is based on such a large sample. We spent some time programming a variance estimator for the GREG estimator, but discontinued the work when other aspects of the comparisons proved more interesting and immediate.

These results do not provide a fair comparison of the GREG esimator with the other estimators. We included GREG estimates in part because Alaska does not oversample LBW births. We wanted to see how the estimator performed for this important characteristic, a proxy for other negative birth outcomes, including neonatal mortality, in a state that was not oversampling the characteristic. This study was a preliminary investigation of the estimators, so ways of adapting GREG were left for later work. One idea is to treat the stratifier as simply another variable in the demographic cross-classification. To keep the number of cells manageable, one might then need to drop one of the other demographic variables.

FUTURE WORK

We hope to look at GREG estimates for more characteristics not correlated with stratification variables. Characteristics that do not involve the same limitations would give us a better gauge of the performance of the estimator. Among potential variables for study that we initially considered, we rejected alcohol use because the reported prevalence on the birth certificate is so very low. Otherwise, its lack of correlation with states' stratification variables made it attractive for study.

We have given some thought to a mixed strategy for estimation. We alluded in our conclusions to recommending direct estimates for the larger small areas and synthetic for the genuinely small areas, but this suggestion would require considerable elaboration before it could be considered a rule of thumb. Deciding whether an area is small is not necessarily easy.

The PRAMS Working Group consists of these people: AL, Albert Woolbright; AK, Karen Pearson; CA, Tamara Anderson; DC, Patricia Tompkins; FL, Richard Hopkins; GA, Jacqueline Bennett; IN, Judith Ganser; ME, Martha Burns; MI, Janet Eyster; NY, Michael Medvesky; OK, Richard Lorenz; SC, Betsy Barton; WA, Adrienne Dorf; WV, Marcy Heitmeyer.

BIBLIOGRAPHY

Adams MM, Shulman HB, Bruce C, Hogue C, and Brogan D (1991) "The Pregnancy Risk Assessment Monitoring System: Design, Questionnaire, Data Collection, and Response Rates," *Paediatric and Perinatal Epidemiology*, 5: 333-346.

Holt, D., Smith, T.M.F., and Tomberlin, T.J. (1979) "A Model Based Approach to Estimation for Small Subgroups of a Population," *Journal of the American Statistical Association*, 74: 405-410.

Laake, P. (1979) "A Prediction Approach to Subdomain Estimation in Finite Populations," *Journal of the American Statistical Association*, 74: 355-358.

Sarndal, C.E. (1984) "Design-Consistent Versus Model-dependent Estimation for Small Domains," *Journal of the American Statistical Association*, 79: 624-631.

Schaible, W.L., Brock, D.B., Casady, R.J., and Schnack, G.A. (1979) "Small Area Estimation: An Empirical Comparison of Conventional and Synthetic Estimators for States," (PHS Publication No. 80-1356), Washington, DC: U.S. Government Printing Office.

Shah, B.V., Barnwell, B.G., Hunt, P.N., and LaVange, L.M. (1991) *SUDAAN User's Manual, Release 5.50*, Research Triangle Park, NC: Research Triangle Institute.

Wright, R.L. (1983) "Finite Population Sampling with Multivariate Auxiliary Information," *Journal of the American Statistical Association*, 78: 879-884.