INVESTIGATION OF THE IMPACT OF IMPUTATION ON VARIANCE ESTIMATION IN THE MEDICAL EXPENDITURE PANEL SURVEY (MEPS)[1]

Robert M. Baskin, Lap-Ming Wun, John Sommers, Marc Zodet, Steven R. Machlin, Trena M. Ezzati-Rice, and Sourish Saha
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850

KEY WORDS: balanced repeated replication (BRR); Rao-Shao adjustment; Taylor method; collapsing; generalized variance function (GVF)

Introduction

The Medical Expenditure Panel Survey (MEPS) is a complex national probability sample survey sponsored by the Agency for Healthcare Research and Quality (AHRQ). MEPS is designed to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. MEPS consists of a family of three interrelated surveys with the Household Component (HC) as the core survey. Healthcare expenditures is considered one of the primary analysis variables in MEPS. The MEPS-HC, like most sample surveys, experiences item nonresponse despite efforts to collect complete information. There is a substantial amount of item nonresponse on expenditures in MEPS. To compensate for missing data, a weighted sequential hotdeck imputation approach is used to reduce the potential bias in estimating expenditures. The standard variance estimators do not account for any impact on variance due to imputation. The purpose of this study is to investigate methods for adjusting the variance estimates for variance due to imputation. Specifically, for imputing 2001 outpatient facility expenditures, two methods of adjusting the variance for imputation are studied and compared to the usual estimator of variance. Also, an approach for providing users with the ability to adjust variances is discussed.

Background: MEPS Sample

The sample of households for the MEPS-HC is a subsample of households that responded to the prior year's National Health Interview Survey, conducted by the National Center for Health Statistics. The MEPS-HC uses an overlapping panel design in which data are collected through a series of five rounds of interviews over a two and one-half year period. Analytic weights, accounting for survey nonresponse, are calculated for MEPS and the details on the weights as well as detailed information on the MEPS sample design can be found in (Cohen, 1997; Cohen, 2000).

Medical expenditures is one of the primary analysis variables collected in the MEPS-HC. It has a high percent of missingness. To compensate for the missing data and to improve the accuracy of the survey estimates, data on expenses for household respondents are also collected from a sample of their health care providers in the Medical Provider Component of MEPS. However, frequently expense data are not available from either survey. The method used for imputation is the weighted sequential hotdeck described in Cox (1978). A further description of the methodology for imputing missing expenditure data can be found in Machlin and Dougherty (2004).

Impact of Imputation on Variance

The main advantage of imputation is that full case analysis can be employed on data with missing values and, if the imputation is effective, the correct multivariate structure of the data can be maintained. However, the use of standard software on data with imputed values assumes that all the data points are observed and doesn't take into account any variability in the imputation. Thus, standard variance estimates are downwardly biased. For a recent review of the state of the art of imputation, see Schafer and Graham (2000) and for a detailed source

[1] The views expressed in this paper are those of the authors and no official endorsement by the Department of Health and Human Services or the Agency for Healthcare Research and Quality is intended or should be inferred.

see Little and Rubin (2002).

The current method of imputation implemented by Westat for MEPS expenditure data is a form of weighted hotdeck. Although the survey weights are used to match donors to recipients, the hotdeck approach used includes a random number for selection of which donor to match to a recipient. This adds a level of variability to the data that is not reflected in the usual variance estimators.

Methods of Adjusting Variance for Imputation

The purpose of this study is to attempt to investigate added variance due to imputation. Two types of methods had been previously published for estimating the impact of imputation on variance estimates. Multiple Imputation is a method of creating more than one imputation for each missing item and can be found in Little and Rubin (2002). There are resampling methods to account for variance due to imputation. Two methods for bootstrap and two for jackknife, also mentioned in Little and Rubin (2002), are found in Rao (1996) and Fay (1996). For hotdeck imputation there is an adjustment for balanced repeated replication (BRR) variance estimators that can be found in Rao and Shao (1999). For the current imputation methodology in MEPS there are problems with all of these approaches. For multiple imputation our weighted hotdeck is not a proper imputation in the sense of Little and Rubin (2002). We could try to perform a proper imputation but the increased variance would be for the proper imputation and not the production imputation that we publish in the public use files (PUF). The bootstrap requires an imputation for each missing value in each bootstrap replicate and this a priori appeared too computationally intensive for an investigative study. And finally, replicates are not created for MEPS data.

In order to carry out this study we considered many methodologies but actually settled on a resampling method that, as far as the authors are aware, has not been previously implemented using a production imputation for a complex survey. For the purpose of estimating the variance including increased variance due to imputation we created 32 BRR replicates and independently reimputed missing data within each replicate as well as independently performing a full sample imputation.

The imputation was carried out using the production software implemented by Westat. Because of the computationally intensive nature of performing the imputation 33 times, the imputation was only run for missing outpatient facility expenditures. The full sample and each set of 32 replicates was run through the outpatient imputation process in a manner mimicking the production process as closely as possible.

There are certain limitations to the results of the study but it does provide important information to evaluate whether this type of study could be performed on a larger scale. First, the number of replicates, 32, was settled on as a compromise because of issues of timing and resource constraints. There are some rules of thumb which indicate that at least 50 BRR replicates are considered the minimum required but for this study only 32 were used. A second limitation is that only outpatient facility imputations were performed. In the 2001 PUF, outpatient facility events account for about nine percent of total expenditures.

Once the imputation runs were performed on the replicates, BRR weights were needed to calculate BRR estimates of variance which accounted for the increased variance due to imputation. These replicates and weights are not part of any PUF but only used internally for evaluation purposes so they cannot be considered production quality.

Evaluation of the Methods

Once the full sample dataset along with the 32 replicated imputed datasets were available, the variances of estimates of mean, median and total outpatient facility expenditures were calculated using four methods for comparison purposes. For the purpose of this paper only the estimates and variances of total are reported. These variances of the total were calculated for the overall sample as well as for subsets of the sample corresponding to sex, race-ethnicity, education status, region and MSA status.

Two naive methods of calculating a variance that ignored the imputation were used. First, using the stratification from the 2001 PUF, a standard SUDAAN weighted estimate of variance of total expenditures was calculated. Second, since BRR weights and replicates were available, a naïve BRR estimate of variance of total expenditures based on the full sample imputation only was calculated.

Two estimates that account for imputation were also calculated. First, the BRR estimate of variance of total expenditures using the replicated imputation was calculated. There is also a method of adjusting the BRR variance estimate for imputation due to Rao and Shao (1999). This method requires the calculation of the full sample mean as well as a mean for each replicate within each imputation adjustment cell. The imputed data, but not the observed data, in each replicate are then adjusted by the difference in the full sample mean and the corresponding replicate mean. This method does not require reimputing the missing data in each replicate but it does require knowledge of the cells used for imputation. Because of the complicated collapsing of imputation cells the final set of imputation cells in each replicate as well as the final collapsed cells in the full sample is not known. This creates a difficulty in applying the Rao-Shao method but an approximation is still available and that was employed. The estimate can be applied to the uncollapsed cells and to cells that are as collapsed as possible and, on average, this gives lower and upper bounds on the adjustment. This actually led to the observation that collapsing of imputation cells can increase the part of the variance *due to imputation*. This was strongly supported by the empirical evidence from the study based on the two versions of the Rao-Shao adjustment.

While in retrospect it seems obvious that collapsing often does increase the variance *due to imputation*, before the study the intuition was that collapsing could decrease the variance *due to imputation*, possibly based on the thinking that collapsing of imputation cells could decrease the variance *ignoring imputation* similar to collapsing of nonresponse adjustment cells. All of this means that we don't know exactly what happens to the total variance as a result of collapsing, but that collapsing does have implications in terms of the variance estimate. It should be clearly stated that collapsing of imputation cells cannot be avoided. We often find imputation cells which contain recipients and no donors and this factor overrides any consideration of variance.

Results

The estimate of standard error accounting for imputation using replication indicated that for the overall estimate of total outpatient facility expenditures the increase was about 30% compared to the naïve SUDAAN estimates of standard error. The following table, Table 1, gives the point estimates and standard errors of total outpatient facility expenditures for the overall sample as well as for the subgroups of males and females. The percent increase for the subgroups of males and females is smaller than the overall increase. For males and females the increase in standard error in accounting for imputation is on the order of twelve percent.

Table 1.

Group	Events for Males	Events for Females	Overall Events
Cell size	6,719	9,179	15,898
Total Expenses (in millions)	$22,787	$30,007	$52,795
Taylor SE (in millions)	$1,736	$1,632,	$2,537
BRR Ignoring Imputation (in millions)	$1,584	$1,603	$2,566
BRR for replicated Imputation (in millions)	$1,970	$1,813	$3,313
Rao-Shao ignoring collapsing (in millions)	$1,896	$1,920	$3,304
Rao-Shao fully collapsed for an upper bound (in millions)	$2,075	$2,106	$3,756

The following graph shows the comparison of standard errors computed by the four methods for the overall sample as well as all fifteen subgroups formed by the variables sex, race-ethnicity, education status, region and MSA status. Note that the points are plotted with the x-axis corresponding to the sum of the weights associated with the subgroup.

(insert graph 1)

Accounting for imputation in the variance estimates has an effect on confidence intervals. Since imputation increases the standard errors, then the width of confidence intervals will increase by the same percent. Since the overall estimate of standard error increased by about 30% when comparing the naïve estimate to the BRR estimate accounting for imputation, then a confidence interval for the overall estimate of total outpatient facility expenditures would also increase by 30%. The increase in the width of confidence intervals can be seen in the following graph.

(insert graph 2)

Discussion

Because replicate definitions, replicate weights, and replicate imputations are not available to users, a user would not be able to account for the increase in variance due to imputation as was done in this study. In order to provide users with the information from this study, one option to consider is the feasibility of providing users with generalized variance functions (GVF) that account for variance due to imputation. Variances accounting for imputation could be calculated for many subgroups of the data and GVFs could be fit to data points which are pairs of sums of weights and the estimated variances. These GVFs are assumed to have a functional form of $\mathrm{var} = a(weight)^2 + b(weight)$, i.e., the intercept term is assumed to be zero. Graph 1 above was a precursor of this idea and the general shape of the square root of the GVF would follow the curve for the plotted standard error. Using the fifteen subgroups cited previously, GVFs for all four types of standard errors were fitted using the lm (linear model) package in R. These GVFs are plotted in graph 3.

(insert graph 3).

The GVF curves appear to fit the points well, with the exception of the points corresponding to Region. The GVFs are easy to calculate once the variance estimates have been made. So it would appear that GVFs are an option to provide users with estimates of variances that account for imputation. Note that any method of estimating the variance accounting for imputation could be used in conjunction with the idea of the GVF. Thus multiple imputation, the Rao-Shao adjustment using BRR or replicating the imputation could be used and the results could be provided through a GVF.

There is one further issue to be dealt with in terms of GVFs. For MEPS, a typical rule of thumb is that if a coefficient of variation (CV) of a standard error (the relative standard error) exceeds 30 percent the result should be flagged for publication purposes. For each of the methods of estimating a variance and for each of the subgroups, a CV of the point estimate was calculated. For subdomains with over 50 million estimated events the CVs were nearly always smaller than 10% but as the subdomain size decreased the CVs approached 40% within the smallest subgroups. These results are shown in graph 4.

(insert graph 4)

Thus, for the GVF option, a discussion would need to be included on how to warn users of the GVFs on the size of the relative standard errors for small domain sizes.

Conclusion

This study attempted to determine if it was possible to measure the impact of imputation on variance estimation for MEPS expenditure data and if so what was the size of the variance due to imputation. A narrow study of one type of imputation, outpatient facility events, was conducted and with limited resources it appears that the current methodology might be a viable approach to providing an estimate of the impact of imputation on variance estimates. In this study the increase in variance of the overall estimate of total due to imputing outpatient facility events is estimated to be 30%.

The next step in this project for future research would be to estimate the impact of imputation for inpatient events since this is a more complicated imputation and inpatient events represent a larger share of total expenditures.

References

Cohen S.B. (1997). "Sample Design of the Medical Expenditure Panel Survey Household Component". Agency for Health Care Policy and Research, MEPS Methodology Report, No. 2, Rockville, MD., AHCPR Pub. No. 97-0027.

Cohen S. B. (2000), "Sample Design of the 1997 Medical Expenditure Panel Survey Household Component". Agency for Healthcare Research and Quality, MEPS Methodology Report, No. 11, Rockville, MD., AHRQ Pub. No.01-0001.

Cox, Brenda (1980), "The Weighted Sequential Hot Deck Imputation Procedure", *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 271-276.

Fay, R. E.(1996), ``Alternative paradigms for the analysis of imputed survey data (Pkg: p473-520)'', *Journal of the American Statistical Association*, **91** , 490-498

Little, R.J.A. and Rubin, D.B. (2002) Statistical Analysis With Missing Data (2nd Ed.). New York: Wiley.

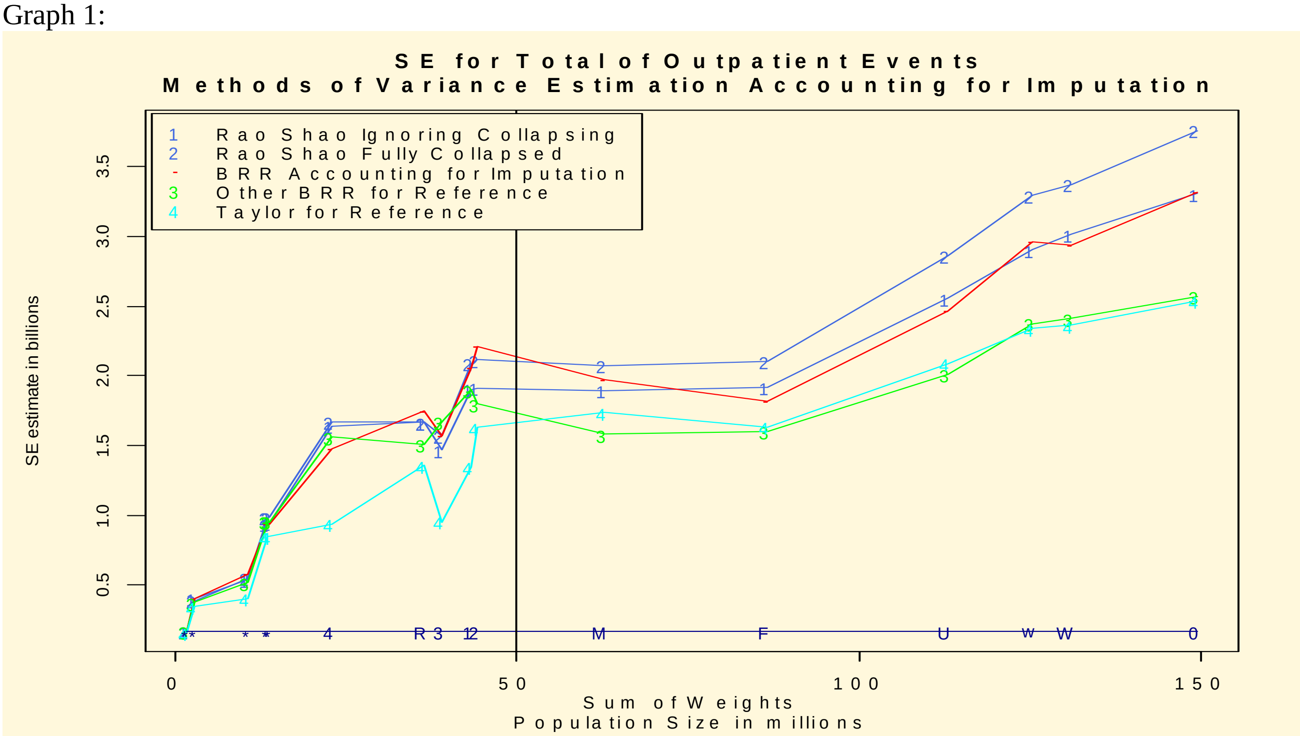
Machlin SR and Dougherty D (2004), "Overview of Methodology for Imputing Missing Expenditure Data in MEPS", *Proceedings of the Survey Research Methods Section*, American Statistical Association, to appear.

Rao, J.N.K (1996), ``On variance estimation with imputed survey data (Pkg: p473-520)'', *Journal of the American Statistical Association*, **91** , 499-506

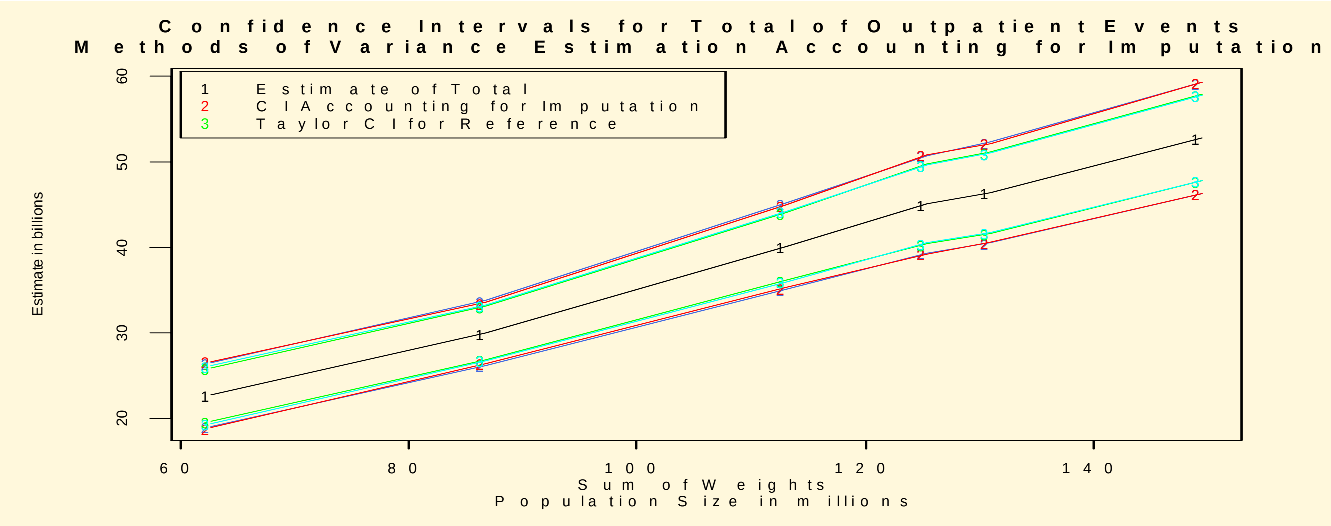
Rao, J.N.K, and Shao, Jun, (1999), "Modified balanced repeated replication for complex survey data", *Biometrika*, vol. 86 no. 2, pp 403-415.

Schaeffer, Joseph and Graham, John (2002), "Missing Data: Our View of the State of the Art", *Psychological Methods*, vol. 7 no. 2, 147-177.

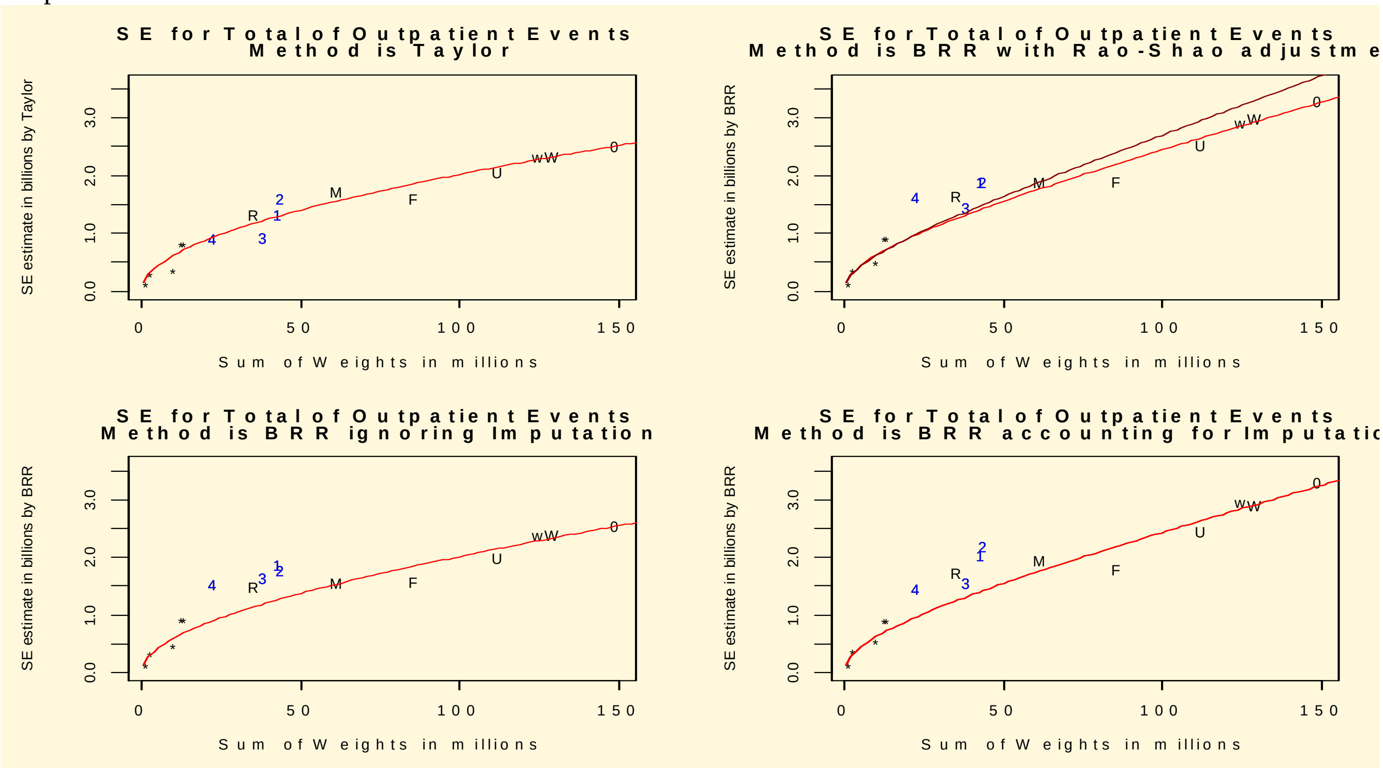
Graph 1:



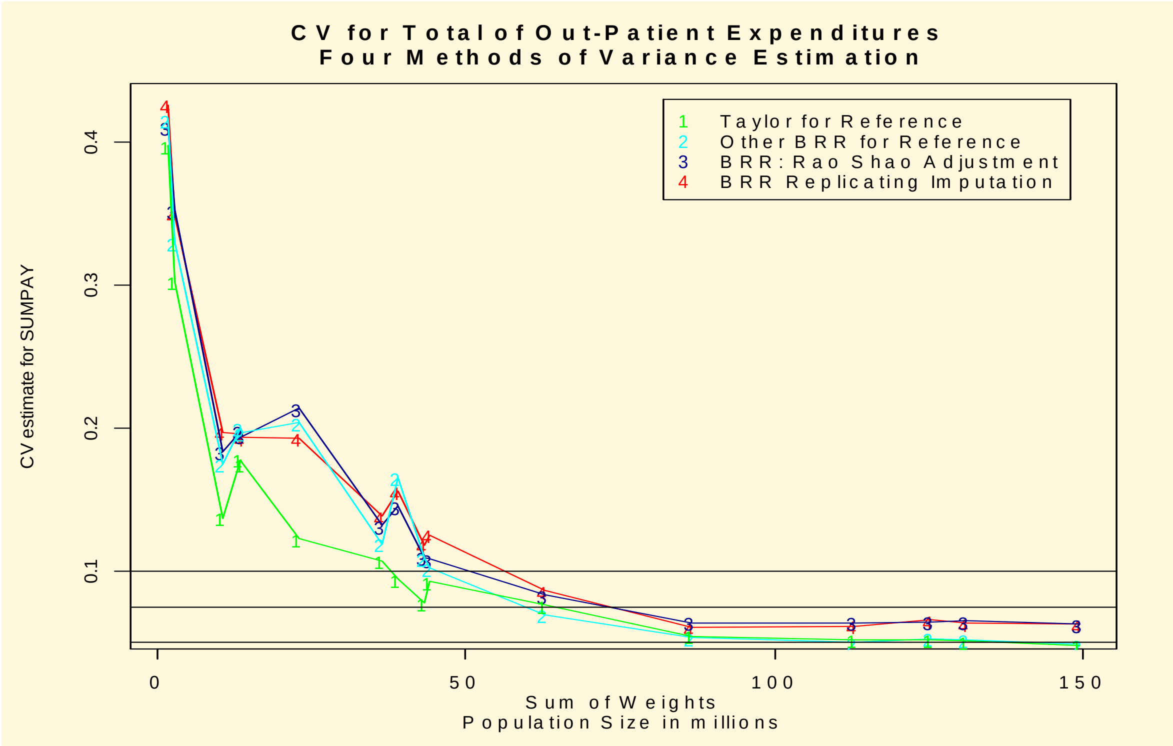
Graph 2



Graph 3



Graph 4


CV for Total of Out-Patient Expenditures
Four Methods of Variance Estimation
1 Taylor for Reference
2 Other BRR for Reference
3 BRR: Rao Shao Adjustment
4 BRR Replicating Imputation
CV estimate for SUMPAY
0.1
0.2
0.3
0.4
0
50
100
150
Sum of Weights
Population Size in millions