WEIGHTED HOT-DECK IMPUTATION OF MEDICAL EXPENDITURES BASED ON A RECORD CHECK SUBSAMPLE

Rick L. Williams, Research Triangle Institute
Ralph E. Folsom, Research Triangle Institute

## 1. INTRODUCTION

The National Medical Care Expenditure Survey (NMCES), a survey of the civilian, noninstitutionalized population of the United States, was designed to collect medical care data for calendar year 1977. The data collected cover use of medical services and the associated costs and sources of payment, health insurance coverage, and access to medical care.

Two interrelated survey components of NMCES were the Household Survey (HHS), a survey of 13,500 randomly selected households interviewed six times over a 15-month period during 1977-78, and the Medical Provider Survey (MPS), a survey of the physicians and facilities that provided medical care to a selected subsample of household respondents during 1977.

Recognizing that Household Survey respondents have difficulty providing accurate data for certain types of health care services, the Medical Provider Survey was implemented to correct for this. MPS was an administrative record check carried out on a probability subsample of the HHS respondents. Through MPS, verifying data were collected on the occurrence of patient visits, diagnoses, charges, and sources of payments for approximately 12,200 HHS respondents.

The primary objective of this task was to project the MPS expenditure reporting experience of those HHS respondents selected for provider record checking to those not selected. This was to result in a data file of all HHS reported visits which could be analyzed as if medical provider reported verifying data had been obtained for all of the HHS respondents. This was accomplished by imputing a medical provider response to each HHS reported visit for which provider data were not available.

A second objective was to evaluate the extent to which the primary goal was satisfied. To this end an evaluation study was implemented to elucidate any possible defects in the implemented imputation procedure and to suggest methods by which the process could be improved.

## 2. IMPUTATION ALGORITHM

The expenditure imputation strategy used for NMCES is often referred to as hot-deck imputation. Under this form of imputation, those survey respondents with complete data are used to donate responses to those with incomplete data. For each survey respondent with incomplete data, called a recipient, a donor with complete data is selected so that the donor is similar to its recipient with respect to known characteristics which are related to the data being imputed. The donor's data are then used to replace the missing recipient data. The term hot-deck refers to the fact that "hot" data obtained from respondents to the current survey are being used for the donor group. This is in contrast to a "cold" procedure which would use donor data obtained from a past survey.

A weighted version of, hot-deck imputation discussed by Cox (1980) was used for NMCES. The weighted sequential hot-deck approach proceeds by first placing donors and recipients on separate files. Assume that there are d donors and r recipients with sampling weights s(i) (i = 1,2,...,d) and w(j) (j = 1,2,...,r), respectively. Also, define s(+) and w(+) to be

$$s(+) = \sum_{i=1}^{d} s(i)$$

and

$$w(+) = \sum_{j=1}^{r} w(j)$$

The algorithm can be conceived of as arraying the donors along a line segment of total length s(+), with donor-i contributing a continuous section of length s(i). The complete line segment is then divided into r mutually exclusive and exhaustive zones, one for each recipient, with the j-th zone being of length w(j)s(+)/w(+). The length of the j-th zone reflects the effect of the weight of recipient-j, w(j), in relation to the sum of weights for all recipients, w(+). From among the set of donors that overlap the j-th zone, one is probabilistically selected to donate data to recipient-j. This introduces the effect of the donor weights. The details of the probabilistic selection process are given by Cox (1980)

Stratification and file ordering can be taken advantage of. For stratification, the sample populations of both donors and recipients are first subdivided and then the imputation procedure is applied independently within each imputation stratum. This guarantees that donors and recipients take on the exact same values with respect to the stratification variables. The advantages of file ordering are obtained by sorting the donors and the recipients by the same variables. In this way, the donor that is as similar as possible to a paricular recipient in the ordering variables will be selected to donate data to the recipient.

Several benefits accrue as a result of using the weighted sequential hot-deck approach. Most important of these is that the weighted hot-deck procedure is constructed so as to insure that within each stratum the expectation, over repeated imputations, of the weighted distribution of imputed recipient values is the same as that of the observed weighted donor distribution. In other words, in expectation, the donor based weighted estimate of the target population distribution is reproduced in the weighted recipient data. This is in contrast to the usual unweighted hot-deck approach which would impute the unweighted sample distribution to the recipients. In the face of unequal weighting that is not desirable.

Another benefit is that each and every donor has a chance of actually being selected to donate data. Under the usual nearest neighbor hot-deck approach, the file ordering may preclude some donors from ever being used and force others to be used an inordinate number of times.

## 3. IMPUTATION STRATIFICATION

Stratification prior to imputation provides a means of controlling and improving the imputation process. This basically occurs in one of several ways. First, stratifying by characteristics which are correlated with the items to be imputed will limit the variability of the imputed values and, hence, make the imputation more precise. In addition, stratifying by characteristics correlated with the imputation items will make the imputation more accurate. In other words, if imputation strata can be formed that group donors and recipients together which share common distributions of imputation items in the target population, then the weighted hot-deck imputation process will be unbiased since an estimate of the appropriate distribution will be imposed on the recipients.

A further consideration, which will be empirically reinforced when dicusssing the evaluation study, is the control of key analysis groups. If study groups can be identified prior to imputation for which it is important to be as accurate and precise as possible, then these groups should be incorporated into the imputation stratification scheme. Recall that the weighted sequential hot-deck algorithm is only unbiased within strata. Thus, if an analysis group is also a stratum, it will receive an unbiased imputation. This is not the case for an uncontrolled (i.e., not a stratum) analysis group. The donor sets for an uncontrolled analysis group will contain donors which are not also members of the analysis group. The analysis group estimates will be tainted if the distribution of imputation items for the non-analysis group donors differ substantivally from those of the analysis group members.

A final consideration for selecting imputation strata is that when using the weighted hot-deck approach to adjust for nonresponse the strata should explain the patterns of differential nonresponse. This again goes back to a basic principle of poststratification or weighting class adjustments for nonresponse as elucidated by Chapman (1976). Since nonresponse bias is a function of the stratum to stratum variation in response rates as well as the stratum to stratum differences between respondent and nonrespondent data, the selection of strata should also be sensitive to response rate variations.

From the preceeding discussion, it can be concluded that stratification prior to imputation should be carried as deeply as possible in order to derive the maximum benefit. However, stratifying too deeply may actually be injurious. The extreme case would be stratifying to the point where each imputation stratum contains only one donor. This scenario implicitly assumes that the totality of strata explains all the variation in imputation items in the target population. In other words, this assumes that all the units in the target population from a particular stratum take on the same values for the imputation items as the donor that represents them in the sample. To see this important point. recall that the weighted hot-deck procedure imputes the estimated donor distribution to the recipients within each stratum. Hence, with only one donor, the estimated donor distribution is a single point which will be imputed to every recipient in the imputation stratum. The entire purpose of going to the additional difficulty to implement the weighted hot-deck procedure is defeated since the within imputation stratum target population distribution cannot be properly estimated from the single donor in each stratum. This problem is analogous to over specifying a regression model, that is, trying to estimate as many parameters as there are data points. These considerations argue for forming strata which contain sufficient donors to allow the donor data distribution to be accurately estimated.

The first step in selecting an imputation stratification scheme is to assemble a collection of candidate stratification variables. The collection should include all variables which are potentially correlated with the imputation items, as well as any key analysis groups. The candidate variables must be available for both donors and recipients alike. Preliminary analyses, conducted during the design phase of MPS by Folsom and Williams (1977), indicated that Household Survey reported medical charges bore the strongest association with provider reported charges, while the effects of the other variables investigated were minimal. Thus, the HHS reported charges were initially considered as prime candidates for stratification variables. Additional candidate stratification variables were supplied by NCHSR and represented the basic reporting groups that will be used in their analyses. Also, many of these variables were thought to have a potential relationship with the imputation items.

In addition, since visits vary substantially by the utilization (physician, outpatient, inpatient, hospital) and billing (flat fee, non-flat fee) types the imputation was stratified by the cross-classification of these two factors a priori. The evaluation of additional imputation stratification variables was conducted separately for each of these eight classes.

As noted earlier, a strong linear association between household survey reported medical expenditures and medical provider reported expenditures had been observed in preliminary analyses conducted by Folsom and Williams (1977). These results were obtained from previously conducted surveys and were for per person annual expenditures, rather than the per visit expenditures specified important for NMCES imputations. Notwithstanding these drawbacks, it was still felt that the HHS reported expenditure should be the primary imputation stratification variable. This conclusion was further supported by consideration of various graphs and regression results from NMCES. For example, consider the four bi-variate histograms in Figures 1 through 4. These figures present the number of matched visits by total non-flat fee expenditures as reported by the households and providers. A distinct relationship between these two reporting sources is clearly evident. Further evidence of this relationship is

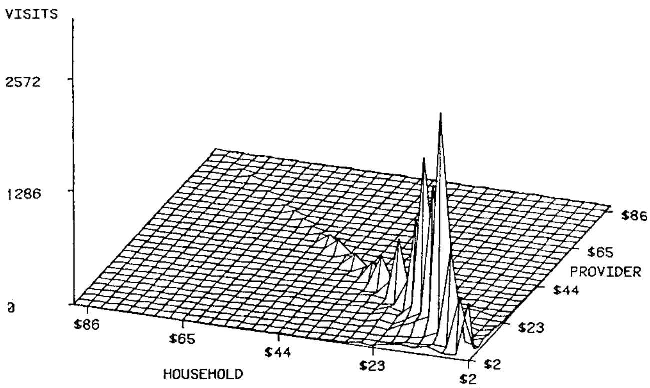


*FIGURE 1.* TOTAL OFFICE BASED PHYSICIAN NON FLATFEE EXPENDITURES
NUMBER OF VISITS BY HOUSEHOLD REPORTED CHARGE & PROVIDER REPORTED CHARGE
18,466 VISITS

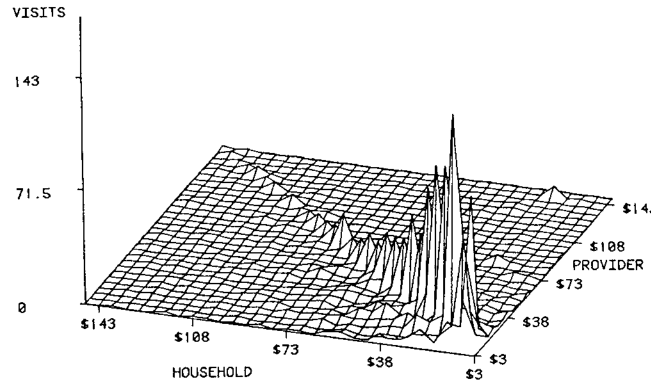


*FIGURE 2.* TOTAL CLINIC & EMER. ROOM NON FLATFEE EXPENDITURES
NUMBER OF VISITS BY HOUSEHOLD REPORTED CHARGE & PROVIDER REPORTED CHARGE
1,861 VISITS

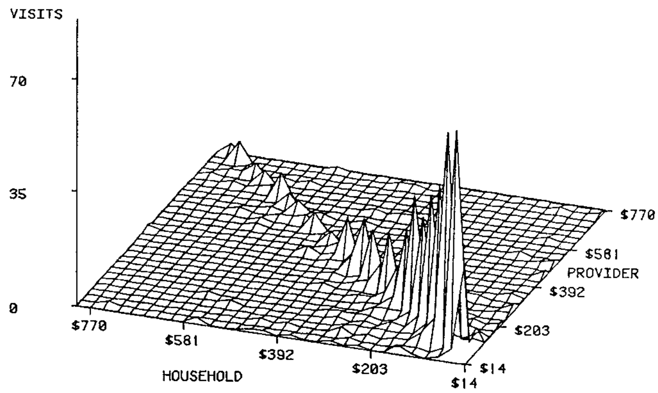


*FIGURE 3.* TOTAL INPATIENT NON FLATFEE EXPENDITURES
NUMBER OF VISITS BY HOUSEHOLD REPORTED CHARGE & PROVIDER REPORTED CHARGE
726 VISITS

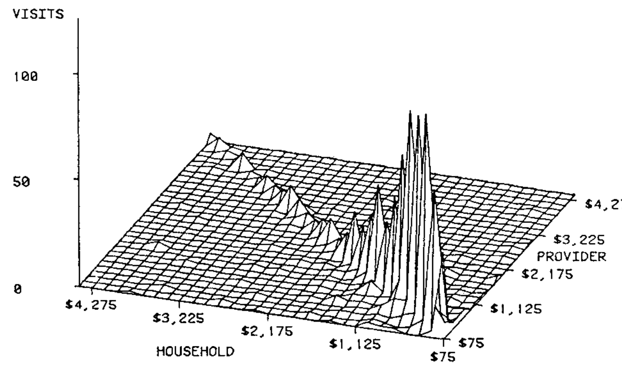


*FIGURE 4.* TOTAL HOSPITAL NON FLATFEE EXPENDITURES
NUMBER OF VISITS BY HOUSEHOLD REPORTED CHARGE & PROVIDER REPORTED CHARGE
1,098 VISITS

presented in Table 3-1 in terms of the percent squared multiple correlation coefficient (R-squared) from regressing MPS reported total charges on HHS reported total charges. Separate means were estimated when the HHS report was missing or zero. While the visit level coefficients of Table 3-1 are not nearly as strong as the person level coefficients of approximately 80 per cent obtained by Folsom and Williams (1977), the relationship is sufficiently strong that gains due to stratification can still be obtained.

Table 3-1

Percent R-squares from Regressing MPS Total Charges on HHS Total Charges

| | Flat Fee | Non-Flat Fee |
|---|---|---|
| Physician | 24 | 19 |
| Outpatient | 7 | 23 |
| Inpatient | 27 | 27 |
| Hospital | - | 42 |

The remainder of the stratification analysis proceeded by forming a separate categorization of the HHS reported total charge for each utilization/billing group. For a particular group, this was accomplished by inspecting the frequency distribution of HHS reported total charges and constructing categories that were felt to capture the variation in HHS total charges, to separate visits that were essentially different in character, and to still contain enough visits to allow further stratification by other factors. The MPS reported charges were then regressed on various cell mean models defined by cross-classifying the data by HHS reported charge categories and the other candidate stratification variables. These regressions indicated that the three variables most highly correlated with the MPS reported charges were the HHS reported charge categories, insurance coverage and income. The R-squares for these models are reported in Table 3-2.

## 4. EMPIRICAL EVALUATION OF EXPENDITURE IMPUTATION METHODOLOGY

### 4.1 Overview of Evaluation

As has been mentioned, an evaluation of the methods related in this report was undertaken. This was felt to be necessary since many new and innovative techniques were developed and implemented for this task. Theoretical considerations indicated that the procedures possessed many desirable properties in expectation over repeated imputations. This chapter describes an empirical investigation designed to determine how the procedures perform in practice.

The evaluation proceeded by generally considering only Household Survey respondents who were selected for provider checking. This set

Table 3-2 Stratification Analysis Results

| Type of Visit | Dependent Variable | Percent R-square |
|---|---|---|
| Physician Flat Fee Visits | Total Bill | 60 |
| | Out-of-Pocket | 35 |
| | Priv. Insur. | 29 |
| | Medicaid | 59 |
| | Medicare | 41 |
| | Other Public | 39 |
| | Other Payment | 70 |
| | Amt. not Expected | 15 |
| Physician Non-Flat Fee Visits | Total Bill | 21 |
| | Out-of-Pocket | 36 |
| | Priv. Insur. | 17 |
| | Medicaid | 24 |
| | Medicare | 5 |
| | Other Public | 4 |
| | Other Payment | 3 |
| | Amt. not Expected | 4 |
| Outpatient Flat Fee Visit | Total Bill | 69 |
| | Out-of-Pocket | 81 |
| | Priv. Insur. | 58 |
| | Medicaid | 19 |
| | Medicare | 63 |
| | Other Public | 67 |
| | Other Payment | 18 |
| | Amt. not Expected | 49 |
| Outpatient Non-Flat Fee Visits | Total Bill | 27 |
| | Out-of-Pocket | 21 |
| | Priv. Insur. | 31 |
| | Medicaid | 25 |
| | Medicare | 30 |
| | Other Public | 3 |
| | Other Payment | 9 |
| | Amt. not Expected | 5 |
| Inpatient Flat Fee Visits | Total Bill | 53 |
| | Out-of-Pocket | 65 |
| | Priv. Insur. | 29 |
| | Medicaid | 56 |
| | Medicare | 50 |
| | Other Public | 23 |
| | Other Payment | 13 |
| | Amt. not Expected | 29 |
| Inpatient Non-Flat Fee Visits | Total Bill | 29 |
| | Out-of-Pocket | 37 |
| | Priv. Insur. | 25 |
| | Medicaid | 14 |
| | Medicare | 17 |
| | Other Public | 7 |
| | Other Payment | 5 |
| | Amt. not Expected | 11 |
| Hospital Non-Flat Fee Visits | Total Bill | 58 |
| | Out-of-Pocket | 21 |
| | Priv. Insur. | 51 |
| | Medicaid | 49 |
| | Medicare | 54 |
| | Other Public | 41 |
| | Other Payment | 10 |
| | Amt. not Expected | 35 |

of persons was then randomly subsampled into two disjoint sets, one being designated as the donor set and the other as the recipient set. The MPS data for those persons designated as recipients were then ignored and data imputed to them from the donors using the same stratification scheme employed for production imputation. By independently replicating samples and imputations within samples it was possible to obtain valid estimates of the variances of imputation based estimates and the components of variance due to sampling and imputation. It should be noted that variance estimates obtained in this manner are conditional on the initial sample of persons and do not incorporate the household sampling variance inherent in the full NMCES sample. However, these variance estimates are exactly the quantities required to assess the ability of the MPS subsampling/imputation process to reproduce complete (i.e., non-subsample) medical provider record check results.

4.2 Study Design and Analysis Methodology

From the set of Household Survey respondents selected for provider checking, eight independent (i.e., with replacement) samples of persons were selected. The first four samples were drawn at a 50 percent rate, while the second four were drawn at a 75 percent rate. This defined eight pairs of visit sets $[S_i, S_i']$ ($i = 1,2,\ldots,8$), where $S_i$ is the set of visits reported by persons in the i-th subsample and $S_i'$ is the complementary set of visits. Five independent imputations were then carried out for each pair using $S_i$ as the donor set and $S_i'$ as the recipient set. The four 50 percent samples and their associated imputations were designed to yield some insight into the effectiveness of the NMCES imputations, while the four 75 percent samples indicate the consequences of a larger provider record check sample and of using the weighted hot-deck imputation method in a more typical nonresponse setting.

This subsampling/imputation design yielded the equivalent of 40 data sets from which 40 sets of population estimates were formed, one set for each subsample/imputation combination. The estimates were calculated using actual non-imputed provider data for the donor visits and imputed provider data for the recipient visits. In addition, full sample estimates were calculated using the actual non-imputed provider data for every visit.

Since the analysis was conducted separately and identically for the data arising from the two different donor group sampling rates the analysis methodology will be described in terms of four donor subsamples drawn at the same rate. Let $y_{ij}$ be the imputation-based estimate of some population value obtained from the i-th subsample (at a particular rate) and the j-th imputation ($i = 1,2,3,4$; $j = 1,2,3,4,5$). For this with replacement sampling and imputation scheme, $y_{ij}$ can be viewed as a random variable arising from the following random effects model:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

where

$$\mu = \text{fixed constant mean,}$$

$$\alpha_i = \text{random effect of subsample-i,}$$

$$\varepsilon_{ij} = \text{random effect of imputation-j within subsample-i.}$$

The following analysis of variance arises naturally in this situation:

| Source | d.f. | Mean Square |
|---|---|---|
| Subsampling | 3 | $MS_S$ |
| Imputation | 16 | $MS_I$ |

In this case, the usual assumptions of homogenous variances and independence are appropriate and it can be shown that

$$E[MS_S] = \sigma_I^2 + 5\,\sigma_S^2$$

and

$$E[MS_I] = \sigma_I^2$$

where $\sigma_I^2 = Var(\varepsilon_{ij})$ and $\sigma_S^2 = Var(\alpha_i)$ for all i and j. The parameter $\sigma_I^2$ is the component of variance due to imputation, while $\sigma_S^2$ is the component due to subsampling. Thus, by equating observed to expected mean squares, estimates of the variance components can be obtained.

The main thrust of the evaluation study was to assess the ability of the combined subsampling/imputation process to reproduce, in expectation, complete (i.e., non-subsample) provider record check results. In statistical terms, it is desired to test the hypothesis

$$H_o: E[\bar{Y}_I] = \bar{Y}_{MPS} \quad \text{versus} \quad H_A: E[\bar{Y}_I] \neq \bar{Y}_{MPS}$$

where $\bar{Y}_I$ is the average of the imputation-based estimates and $\bar{Y}_{MPS}$ is the estimate obtained using the original non-imputed MPS data. The above expectation is taken over repeated subsamples and impuations. The appropriate test statistic for this hypothesis is

$$t = (\bar{Y}_I - \bar{Y}_{MPS})/\sqrt{MS_S/20}$$

which, in this case, has approximately a t distribution with 3 degrees of freedom and a five percent critical value of 3.182.

### 4.3 Results

In order to contain the size of the evaluation study, analysis was restricted to all combinations of three utilization types (outpatient non-flat fee, hospital non-flat fee and physician flat fee) crossed with the five subpopulations reported on in Table 4-1. For the 50 percent donor sample analysis, results for both mean expenditure per visit and proportion of visits with charges falling in certain ranges were obtained. Only mean expenditures were analyzed for the 75 percent donor sample analysis. This implies that a total of 930 hypotheses of the form shown above were tested, many more than can be reported here. To circumvent this problem, the key results of the evaluation study have been summarized in Table 4-1.

Table 4-1. Observed Number of Rejected Hypotheses and Expected Number of Rejection Assuming the Null Hypothesis of No Imputation Bias ($H_o$)

**50 Percent Donor Group**

Mean Estimates (15 hypotheses tested per domain)

| Domain | Number of Hypotheses Actually Rejected | |
|---|---|---|
| Total Population | 2 | Expected number of |
| Black Population | 2 | rejections per domain |
| Non-Black Population | 1 | under $H_o$ = .75 |
| Age 0-18 | 6 | Upper 5 percent critical |
| Age 18-65 | 3 | value for rejecting the |
| Age 65+ | 7 | hypothesis of no imputation bias = 2.14 |

Distributional Estimates (55 hypotheses tested per domain)

| Domain | Number of Hypotheses Actually Rejected | |
|---|---|---|
| Total Population | 4 | Expected number of |
| Black Population | 6 | rejections per domain |
| Non-Black Population | 4 | under $H_o$ = 2.75 |
| Age 0-18 | 14 | Upper 5 percent critical |
| Age 18-65 | 8 | value for rejecting the |
| Age 65+ | 5 | hypothesis of no imputation bias = 5.41 |

Overall (70 hypotheses tested per domain)

| Domain | Number of Hypotheses Actually Rejected | |
|---|---|---|
| Total Population | 6 | Expected number of |
| Black Population | 8 | rejections per domain |
| Non-Black Population | 5 | under $H_o$ = 3.50 |
| Age 0-18 | 20 | Upper 5 percent critical |
| Age 18-65 | 11 | value for rejecting the |
| Age 65+ | 12 | hypothesis of no imputation bias = 6.50 |

**75 Percent Donor Group**

Mean Estimates (15 hypotheses tested per domain)

| Domain | Number of Hypotheses Actually Rejected | |
|---|---|---|
| Total Population | 2 | Expected number of |
| Black Population | 1 | rejections per domain |
| Non-Black Population | 1 | under $H_o$ = .75 |
| Age 0-18 | 2 | Upper 5 percent critical |
| Age 18-65 | 0 | value for rejecting the |
| Age 65+ | 2 | hypothesis of no imputation bias = 2.14 |

Table 4-1 presents the number of hypotheses actually rejected at a 5 percent confidence level (i.e., $|t| > 3.182$) for various groupings. Recall that in any statistical hypothesis testing situation there is always a chance of rejecting the null hypothesis when in fact the null is true and that the probability of this occurring is called the significance level of the test. Thus, for the present case, even if the hypothesis that $E[\bar{Y}_I] = \bar{Y}_{MPS}$ is true, approximately five percent of the tests would still be declared significant with the exact number being a binomial random variable under the assumption of independent tests. Thus, Table 4-1 also contains the expected number of hypothesis rejections and an upper five percent critical value for rejecting the global hypothesis that no more significant results have been observed than should be expected under the null hypothesis of no imputation bias. The upper five percent critical value is based on the Gaussian approximation to the distribution of a binomial frequency.

Inspection of Table 4-1 reveals that for the 75 percent donor group portion of the study, the observed number of significant tests is always less than the five percent critical value. This indicates that the imputation procedure preformed as expected for the 75 percent case. However, a different picture is apparent for the 50 percent donor group. For the 50 percent donor group, only the large total and non-black populations always had fewer rejected hypotheses than their associated five percent critical values. For virtually every other domain, too many hypotheses were rejected to be consonant with the premise of no imputation bias. This disturbing result can be understood if it is remembered that the weighted hot-deck procedure is only unbiased within imputation strata. Thus, an analysis domain which corresponds to a stratum or a group of strata will receive an unbiased imputation. In the results above, this fact is exemplified by the total population, which is a combination of all the strata. Both the theory and the empirical results point to the conclusion that the imputation for the total population is unbiased. This does not seem to be the case for most of the remaining analysis domains included in the 50 percent donor group analysis. Since these domains do not coincide with imputation strata, the theory does not necessary guarantee that the imputation should be unbiased and the results in Table 4-2 suggest that it is not. The 75 percent donor group imputation was less sensitive to the effects of imputation than the 50 percent donor imputation. This is probably a consequence of the smaller proportion of visits which received imputed data for the 75 percent donor group analysis.

The results presented above indicate that the NMCES medical provider data imputations may not be strictly unbiased for domains not defined in terms of imputation strata. This emphasizes the importance of making a judicious selection of strata. Chapter 3 outlined two competing criteria for stratum formation; namely, imputation strata should explain significant variations in the items being imputed but should not over stratify the donors into groups which are too sparse to characterize the imputation item distributions. The stratification analysis, documented in section 4.2, identified the household reported total visit charge, personal insurance coverage, and family income as the three highest correlates of the imputation items. At the same time, virtually every other factor considered in the stratification regression analysis made a statistically significant contribution to the explanation of the MPS data while only marginally improving the fit of the models, as measured by the model R-squares, over that obtained using the three main variables only. To obtain strictly unbiased imputation based estimates of complete provider record check values would require stratifying on every factor found to be statistically significant. Clearly, this was not possible since this would have defined more strata than donor visits. Rather, the course taken was to tightly stratify on the three factors which accounted for the mass of the explainable variation in imputation items to obtain the maximum reduction in the imputation component of variance. It was felt that while this might introduce some statistically significant imputation bias in uncontrolled (i.e., non-stratum) domains, the bias would be small enough to be substantially unimportant. Thus, the fact that the imputation is not unbiased does not necessarily imply that the imputation is unusable provided that bias is small. Unfortunately, this was not always the case. The imputation approach described in this paper was originally conceived when it was felt that multiple R-squares of 80 to 90 percent between the imputation itmes and the stratification variables would be obtainable. The inability to obtain such strong relationships apparently led to the downfall of this approach.

## 5. CONCLUSIONS

The primary lesson to be learned for this style of imputation is the importance of properly selecting an imputation stratification scheme. The evaluation study demonstrated the importance of controlling for analysis groups. The stratification described in this report tightly controlled the household reported total visit charge, the variable most highly related to the imputation items, in an attempt to build an imputation that was universally applicable to all analyses. This attempt was a mixed success. The method suggested for further consideration is to identify the subpopulations to be analyzed and to build the stratification scheme around them. The total household reported visit charge could then be used as either a low level stratification variable or as a sorting variable to help reduce the variance of imputation. If including every domain simultaneously defines too many strata, several separately stratified imputations could be performed for subsets of the domains. The basic principle here is to build the imputation around specific analyses.


## ACKNOWLEDGEMENT

This research was performed for the National Center for Health Services Research (NCHSR) under Contract No. HRA-230-76-0268. The views expressed in this paper are those of the authors and no official endorsement by NCHSR is intended or should be inferred.


## REFERENCES


Chapman, David W. (1979). "A Survey of Nonresponse Imputation Procedures." Proceedings of the American Statistical Assocation, Social Statistics Section.

Cox, Brenda G. (1980). "The Weighted Sequential Hot Deck Imputation Procedure." Proceedings of the American Statistical Association, Section on Survey Research Methods.

Folsom, Ralph E. and Rick L. Williams (1977). Design of a Medical Provider Survey: Sample Design and Analysis Recommendations. Working Paper prepared for the Health Resources Administration under Contract No. HRA-230-76-2102.