Imputing the Physical and Mental Summary Scores (PCS and MCS) for the MOS SF-36 and the Veterans SF-36 Health Survey in the presence of Missing Data

William H. Rogers, Ph.D.
Shirley Qian, M.S.
Lewis Kazis, Sc.D.

Updated and Complete Report*

July 2004*

Technical Report prepared by:

The Health Outcomes Technologies Program, Health Services Department, Boston University School of Public Health, Boston, MA. & The Institute for Health Outcomes and Policy, Center for Health Quality, Outcomes and Economic Research, Department of Veterans Affairs, Bedford, VAMC, MA.

* Note this report is updated from the originally issued white paper on August 29, 2003, with a first update on November 12, 2003 and the second update on July 20, 2004. This report reflects the added validation studies of the Veterans SF-36 Health Survey contained in appendix D. and additional tables reflecting the bias translated to error points in PCS and MCS for both versions of the SF-36.

"Report prepared for the National Committee for Quality Assurance under Contract No. 500-00-0055, entitled Implementing the HEDIS Medicare Health Outcomes Survey, sponsored by the Centers for Medicare & Medicaid Services, U.S. Department of Health and Human Services."

Supported by the Centers for Medicare & Medicaid Services through the National Committee for Quality Assurance; the Health Outcomes Technologies Program, Health Services Department Boston University School of Public Health, Boston MA; and the Center for Health Quality, Outcomes and Economic Research, Veterans Administration Medical Center, Bedford MA..

Questions concerning this work can be e-mailed to Drs. Roger, Kazis or Ms. Qian at:
whrogers@comcast.net; **lek@bu.edu;qian_shirley@yahoo.com**

SF-36® is a registered trademark of the Medical Outcomes Trust

This report does not reflect the viewpoints of any of the organizations supporting this work

The Centers for Medicare & Medicaid Services' Office of Research, Development, and Information (ORDI) strives to make information available to all. Nevertheless, portions of our files including charts, tables, and graphics may be difficult to read using assistive technology.

Persons with disabilities experiencing problems accessing portions of any file should contact ORDI through e-mail at ORDI_508_Compliance@cms.hhs.gov.

Executive Summary

The purpose of this report is to investigate the theory, use, and validation of estimates for the Physical Component Score (PCS), the Mental Component Score (MCS), and the 8 individual scales from the Medical Outcomes Short Form 36 Health Survey (MOS SF-36) and the Veterans SF-36 Health Survey. Algorithms are presented for the MOS SF-36, this report provides the testing or validation using the MOS SF-36 in the main body of this work and using the same methodology presents the results of the validation for the Veterans SF-36 health survey in Appendix D. We focus on 5 methods for handling the missing data. The first strategy deletes all observations with missing data. The second strategy imputes a score if half the items are present. The third strategy imputes scores based on extensions to Item Response Theory for dealing with multivariate concepts, the missing data estimates (MDE). Fourth, we consider a new approach for the SF-36 that uses regression estimates for imputation (RE) and a related fifth method that uses a modified regression estimate (MRE) which is corrected for regression to the mean.

Separate validation studies are conducted that examine the robustness of the RE and MRE approaches when contrasted with the half scale rule. This is the bias relative to specific subgroup comparisons including health plans, disease groups and demographics. The bias for these comparisons is reported in terms of error computed in the PCS and MCS measures. Separately the MRE is compared to the half scale rule and MDE as the difference in the bias.

Results indicate that failure to impute missing data is a major source of bias and that imputation should impact as many cases as possible to minimize bias. Findings reveal that the MRE approach recovers two thirds of the missing cases for PCS that are still missing after the MDE approach is invoked and one third more for MCS cases. Both the MDE and MRE approaches yielded reduced bias when compared with the half scale rule by health plans, disease groups and demographics. Both the MDE and MRE approaches are within one point of each other almost all of the time, with the MRE giving less bias and variation.

We conclude that the MRE and MDE approaches result in almost comparable outcomes using PCS and MCS with the MOS SF-36 (version 1.0). The MRE approach is quite attractive given a recovery of more missing cases than the MDE and half scale rule approaches. Finally, the results presented using the Veterans SF-36 health survey data perform in virtually the same way as the MOS SF-36 using a 1999 large survey of veteran enrollees.

The full report includes the following sections:

1. Introduction to the Problem 5
2. Theory and Methods for Estimates 6
3. Theory and Methods for Validation 10
4. Validation Results of MOS SF-36 14
5. Implications for Analysis 26
6. Conclusions 27
Appendix A: Contents of the CD-ROM 29
Appendix B: Use of Software For HOS and Veterans SF-36 31
Appendix C: Description of Variables and Questions of the SF-36 for the HOS 31
Appendix D: 36

1. Introduction to the Problem

The US Centers for Medicare & Medicaid Services (CMS) is conducting the Medicare Health Outcomes Survey (HOS) to determine the health change of Medicare beneficiaries in a variety of health plans. The process involves surveying beneficiaries before and after a two-year period. A similar process has taken place in the Veterans Administration since 1996 with follow-up periods ranging from 17 months to 5 years.

It would be simple to analyze these data if all beneficiaries answered every question and nobody was lost to follow-up. However, all longitudinal survey work must deal with certain practicalities related to missing data—subjects die, they fail to complete follow-up questionnaires, and they may omit one or more responses even when they do fill out the questionnaires. This report deals with the last cause of missing data and some possible solutions to it.

Ultimately, the success or failure of any set of methods must be judged in terms of its success in any particular application. In statistical terms, is the answer invalid (biased) or imprecise? In order to understand this, we need to appeal to external data of some kind. Fortunately, in the HOS project, replication is part of the longitudinal study design. That is, we can use baseline and follow-up values having complete data as a check on what might be done in cases with incomplete data.

We can translate this into a precise concept. In simplified terms, we can either create artificial missing data by dropping items, or we can study naturally missing data. If we create artificially missing data, we can compare it against known complete data at the same point in time. If we have naturally missing data, we can compare missing data imputations at time 1 against known data at time 2 to figure out the degree of bias, taking true change into account. In the presence of naturally missing data, it is possible that discarding the observation will result in more bias than any imputation.

Once an imputation is created and its bias is known, then strategies for estimating with it can be evaluated. These strategies may include full substitution of a missing value, discarding the observation, or weighting the observation less to lessen the impact of the bias—a framework which includes both full substitution and discarding.

The ultimate accuracy of the imputation method comes from its mean square error in an application, which combines bias and variance. The bias is fixed by the estimator and the nature of the comparison, but the variance depends on the sample size. A slightly biased imputation may be preferred if it can be scored in a larger sample, but this benefit is limited if the sample size is sufficiently large anyway.

A particular imputation method may be very biased in one application, but nearly unbiased in another. For example, an estimate may be biased for determining individual health status, biased for determining the physical and mental summaries from the SF-36 (PCS or MCS) associated with a disease state, but adequate for comparing HOS health plans or VA regions or

VISNs. If the purpose of estimation is general, and it does not matter whether comparisons are made with one scale or another (e.g. physical functioning or bodily pain) and these are conveying roughly the same information, then we are free to impute boldly because there is relatively little bias. However, when the exercise involves PCS and MCS comparisons between health plans, then bias may be important to identify and minimize with methods of imputation.

This report focuses on the MOS SF-36 for the validation studies. We provide the scoring algorithms for the new imputation approach described for the MOS SF-36 and the Veterans SF-36 in the appendices. In this revised report we also present a separate set of studies in the appendix for the Veterans SF-36 using the 1999 Large Health Survey of Veteran Enrollees. A separate estimator is needed for the VA because its survey format differs (that is, the Veterans SF-36 and the MOS SF-36 are not identical), as will be shown in the appendix D. the results are almost comparable.

We begin with the HOS data base because (1) there has been more work done on missing data previously in the HOS, and (2) since the HOS uses the MOS SF-36 or SF-36 version 1.0, it is more pertinent to the immediate needs of the Health Outcomes Survey program.

2. Theory and Methods for Estimates

SF-36 and its Versions

The SF-36[1] is composed of 36 items, one of which measures health change leaving 35 health status items. These items are grouped into 8 scales: Physical Functioning (PF, 10 items), Role Physical (RP, 4 items), Bodily Pain (BP, 2 items), General Health (GH, 5 items), Vitality (VT, 4 items), Social Functioning (SF, 2 items), Role Emotional (RE, 3 items), and Mental Health (MH, 5 items). All of the scales are scored so that the least health has a value of 0 and the greatest health has a value of 100.

From these 8 scales, two linear combinations are commonly computed: a Physical Component Summary (PCS), and a Mental Component Summary (MCS)[2-3]. Based on a general population survey (NORC) conducted in 1990, the average U.S. values for PCS and MCS are 50.

Brief History of the SF-36 versions and adoption by the HOS and VA

Recently, Version 2.0 of the SF-36 was introduced (Ware et al.[3]). The purpose of this version was to correct some outdated language in Version 1.0 and improve the response categories to some questions--particularly the role functioning items. In order to come up with an equivalence, the Version 2.0 survey was fielded along with version 1.0 in a new national survey in 1998 conducted by NRC. The 1998 survey found that the 1990 scale means and standard deviations had shifted. A scoring guide for Version 2.0 -compatible summary measures based on Version 1.0 scales is available[3]. We refer to these as the 1998 rules. The primary impact of the 1998 rules as opposed to the 1990 is a translation (add a constant and multiply by a factor) in PCS and MCS.

The HOS study uses the original version of the questions, but the HOS has now adopted the 1998 scoring rules. At first this happened because the HOS used missing data estimates (MDE, explained in detail below) that were constructed from the 1998 rules. Later, HSAG produced files for Cohort 3 based explicitly on the 1998 rules but did not use the MDE (as of the date that this document was prepared). Because the MDE response scales are based on 1998 data and HOS appears to be using 1998 rules, this report examined the 1998 scoring in detail. However, we also applied our methods to the original Version 1.0 scoring rules and the results would be equivalent. This report is accompanied by a CD-ROM with programs, and the programs score missing data according to both 1990 and 1998 rules (see appendix A).

Meanwhile, the Veterans Administration developed a modification to the MOS SF-36 based on suggestions from Ware. [4] The modifications to the MOS SF-36 are in the response choices of the role physical and role emotional items (RP and RE). The dichotomized two point yes/no choices were changed to five-point Likert scales in order to reduce floor and ceiling effects. With the exception of these role scales and the change items for physical and emotional health, scoring of the Veterans SF-36 scales is the same as that for the MOS SF-36.[1] This process includes a linear transformation from a raw score so that scores range from 0 to 100, where 100 denotes the best health. Scoring of the Veterans SF-36 RP and RE scales uses an algorithm previously developed and validated to ensure comparability with the MOS SF-36.[5-6]

Types of missing data and strategies to deal with them

What are the types of missing data?

There are three types of missing data. *Planned* missing data occurs because the survey instrument did not collect them. *Naturally* missing data occurs because the respondent did not fill out the items. *Intentionally* missing data occurs because in studies such as the HOS,, we often want to observe how estimators would be constructed if we treated certain items as missing, so we can compare with the known value (the gold standard) that was observed. Randomized missing data can either be *systematic* where the same items are deleted for all observations, or *mixed* where different patterns of missing data are created for different observations.

Casewise deletion. The most convenient solution to missing data is simply to delete it. This solution, often referred to as casewise deletion, is a popular default in statistical software. The result of any arithmetic operation is missing if any component is missing. When computing regression estimates, observations are only used if all variables are present in order to assure the numerical stability of the computations.

The problem with casewise deletion is that many observations may be lost even though there are only slight amounts of missing data. For example, a case would be lost if even 1 of the 35 SF-36 items used in the computation were missing. A large fraction of potential cases can be eliminated in this way, about a third of the cases in the HOS for example. The loss of so many observations raises questions about both the bias and the precision of estimates drawn from the complete cases.

Half-scoring rule. The second method of handling missing data studied in this report comes from the original SF-36 reference[7] and has a long history[8]. Under the half-scoring rule, a scale is considered to be scorable if half or more of the items were present. The remaining items are for the most part prorated. The summary methods are considered scorable if all 8 of the scales can be scored.

One major limitation with the half-scoring rule is that scales can be scored usefully with much less data than half. Another limitation is that the method does not take into account "what items are missing." If the items have varying degrees of difficulty (in the Guttman scaling sense), it does not matter if the "easiest" or the "hardest" item is missing, the rule is the same. With regard to scoring the summary scores, the rule is also conservative. Not all scales are really needed for PCS and MCS, particularly if a relatively unimportant scale (e.g. such as social Functioning for PCS) is missing.

Missing Data Estimates (MDE). This method of imputation is based on extensions to Item Response Theory for dealing with multivariate concepts. At least 3 such extensions exist, but at this time all details are unavailable. The methods are proprietary and information is available through QualityMetric™ at www.qualitymetric.com. These approaches have great promise. They are however, proprietary and the documentation on them is limited. These approaches are based upon statistical models that at this time are unknown and the models depend upon training data.

The MDE imputations apply to Version 2.0 of the SF-36 using the 1998 rules for scoring with norms (see earlier section, Brief History of the SF-36 versions and adoption by the HOS and VA). For purposes of this report, we had the MDE estimates available to us but not with the statistical algorithms for creating the models. We did not have MDE imputations available for individual SF-36 items/scales. MDE estimates can be created by going to the QualityMetric™ website.

Based on the observed values, the MDE estimates appear to be tolerant of 1 totally missing scale, so long as it is not the scale on the same end of the SF-36 spectrum as another scale that is missing. For example, the physical summary (PCS) requires at least 1 PF item (physical functioning scale) and the mental summary (MCS) requires at least 1 MH (mental health scale) item.

New approaches for imputing missing values for the SF-36 (Regression Estimates (RE) and Modified Regression Estimates (MRE)

Regression estimates (RE). These are based on breaking each item down into a set of indicator (dummy) variables for the various responses and then regressing PCS and MCS on indicator variables for available items. For example, the PF01 item has three responses (1=limited a lot, 2=limited a little, 3=not limited at all). Indicators are scored for response 2 (pf01r2) or response 3 (pf01r3). If the respondent chooses 2, then pf01r2=1 and pf01r3=0. One indicator in each set is always omitted, it does not matter which one. The method uses complete cases to estimate a regression equation where only those items that are present are used. The following gives the complete equation assuming all items are present.

$$PCS98 = a + b_1*pf01r2 + b_2*pf01r3 + b_3*pf02r2 + ... + b_{109}*mh05r6$$

For each pattern of missing values, this approach gives regression estimates and an R^2 one for each PCS and MCS. The coefficients b_i can then be applied to data from cases with actual missing data to estimate PCS and MCS scores.

The SF-12,[9] a well-established shorter form of the SF-36 that explains more than 90% of the reliable variance in the 36 items, is one such regression estimate based on the assumption that only 12 items are fielded--a situation we refer to as planned missing data. Regression estimates depend on a training data set and so they are data-dependent, similar to the MDE. For the SF-12, the training data came from the 1990 NORC survey. Other subsets have also been fielded in various studies.

This method can be applied to naturally as well as planned missing data. For naturally missing data, the calculations are done by considering each case with missing data in turn. The missing data in that case form a pattern which can be thought of as a 35-bit binary number (each digit is 1 if the corresponding item is present and 0 if missing). A regression estimate would be constructed (based on the data with complete cases) indicating how to score PCS and MCS with exactly the pattern in the candidate case. The coefficients can be computed and a predicted value is scored for the candidate case. The whole process is then repeated for the next observation.

In this native form, the regression method is slow and tedious, but the computation is greatly speeded up by a few ideas. First, regression estimates depend on a cross-product matrix. This matrix is always the same, but we use different subsets depending on which items are used. The matrix can be manipulated to get the correct subset for the pattern. Second, many observations have the same pattern of missing data, and these can be evaluated together. With these two provisions, a computation that would take weeks for the HOS is reduced to minutes.

If a planned subset of the SF-36 is fielded (such as the SF-12[9]), the number of possible patterns may be greatly reduced. A preexisting set of coefficients can be computed for each pattern. Merging that set of coefficients with data and computing the estimates requires only seconds of computer time and no matrix algebra is required. The preexisting coefficient method requires an auxiliary dataset with 2 raised to a power equal to the number of variables available, so as a practical matter the upper limit is 15 to 20 items, but the SF-36 in full is 35 items. Hence, the program supplied with this report uses the matrix method, and that requires a workstation computer with a sufficiently complete installation of SAS.

Modified Regression Estimate. The regression estimates are pulled toward the mean of the particular training data set, depending on the number and usefulness of the items available. This creates bias if the estimates are extended to outside populations or even distinct subpopulations in the original sample. This modification corrects for this regression-to-the mean effect:

$$Y_{modified} = (average) + (Y_{regression} - average)/R$$

where R is the square root of R-squared (percent variance explained) in the regression model used and average is the average value in the training dataset. Possibly, this could extend the usefulness down to very tiny subsets of the SF-36. The benefits of doing this are discussed in the results section.

Other Considerations: Training datasets

The first two estimates (casewise deletion and half-scoring) do not depend on data. They can be achieved abstractly in computer code. The last three methods (Missing Data Estimates (MDE), Regression Estimates (RE) and Modified Regression Estimates (MRE) depend on models and training data.

The MDE method uses a proprietary dataset available to QualityMetric™ for training data. The RE and MRE methods in this paper are based on the complete cases in the first two cohorts of the HOS study (elderly and non-elderly combined), 289,650 individuals. Since the training data are based on the HOS, the regression and modified regression estimates have an advantage over the other three methods in present and future HOS data, whereas in non-HOS samples, this should be construed as a disadvantage, depending on how different the samples are from the HOS. The CD-ROM companion includes the estimates for the RE and MRE methods.

We have also applied the regression and modified regression method to the 1999 VA large survey data based on 539,000 training cases, and the algorithms used are also included on the CD-ROM companion and outlined in Appendix A and B. This revised report also documents the validity of the Veterans SF-36 Health Survey for the purposes of imputation and is presented in Appendix D. We do note that the MDE comparison is not included for the Veterans SF-36 validation work as the specifics of their algorithm is not available on the Quality Metric Web site and computes scores based upon the MOS version of the SF-36. The tables presented in Appendix D reflect comparisons of the half scale rule and the RE and MRE approaches.

3. Theory and Methods for Validation

An important issue for the MOS SF-12 in its early days was whether the performance described under a randomized systematic experiment would apply under planned conditions. That is, did the presence of some items cause respondents to change their answers to other questions? Several researchers concluded that there were no large biases when comparing these two methods, but it was also found that the R-square values were less than described in the original documentation (.93 instead of .98). This problem is sometimes known as *serendipity*, which refers to a result that is "too good" because of efforts to optimize it in the face of uncertainty.

Further discussion of the benefits of "imputing" scores for missing data depends on two error concepts--*bias and variation*. *Bias* occurs because the estimate used differs systematically from what we would have obtained with complete data. *Variation* occurs because an estimate varies around the expected answer, due to sampling. Theoretically, it helps to conceptualize what

the answer would have been if there were an infinite number of observations with the same missing data phenomena that are seen in the finite data.

Error = bias + variation = (infinite answer - true answer) + (sample answer - infinite answer).

As the sample size increases, the first term remains the same, but the last term approaches zero according to the law of large numbers. Accordingly, bias is much more of a threat in large samples, but variation is more of a threat in small samples. In large samples, we need to take care with imputation or case exclusion because of the dangers of drawing an incorrect conclusion with a false sense of precision. In small samples we need to be concerned with the unnecessary deletion of observations. Whether the sample is large or small depends on both the study and what is being compared. In the case of the HOS, the sample is very large if we are following the health of patients in HMOs generally, but smaller if we are comparing health plans.

In a given situation, bias and variance arise because of different aspects of analysis, so we can create a formal trade-off and attempt to minimize a combination of the two. The combination usually encountered is Mean Squared Error (MSE) which is defined

$$MSE = bias^2 + variation^2$$

To give this problem more analytic structure, we have two options for each missing data strategy--we can include the observations with their missing data estimates, or we can exclude them. In addition, we can weight them. A weight of 0 corresponds to excluding them, and a weight of 1.0 is equivalent to including them. Given N1 samples with complete cases and N2 samples which could be imputed with squared *bias h* and variations with *variance v*:

$$\text{Bias contribution} = (N2*W/(N1+W*N2))^2\ h$$
$$\text{Variation contribution} = (N1+W^2\ N2)/(N1+W\ N2)^2\ v$$

N1: samples with complete cases
N2: samples with incomplete case
W: Weight
h = squared bias
v = squared standard deviation

This assumes that the variation contributes about the same amount for complete as well as incomplete cases. Unless the amount of missing data is extreme, the variation of the imputed observations is about the same as the complete cases. In addition, it is helpful to express both h and v in standard terms--the only thing that really matters is the ratio h/v and the sample sizes N1 and N2.

Illustration with systematic planned missing data. To illustrate these terms with practical data, imagine that we simulate the planned omission of the pf01 item under the half-scale rule and we evaluate PCS98 for a population mean, and a comparison of health plan baseline scores.

For the population mean, the bias in half-scale PCS98 was $B_K = 0.6079$ points on average and the standard deviation of (half-scale PCS98 - true value) was $SD_K = 0.550$. The standard deviation of PCS98 was 11.74. The ratio h/v is $.6079^2/11.74^2$ or 0.00268. Suppose that N1 is 650 and N2 is 350. Then W should be about 0.5 and imputation is better than no imputation. However, if N1=6500 and N2=3500 we are much better off not imputing.

For comparison of health plan baseline scores, we run an analysis of variance (ANOVA) of the difference (PCS98 with half-scale scoring - PCS98 gold standard or complete data) on health plan baseline ID. We get a SS(plan) of 1036.77 with F = 10.83, so

h/v = (SS of effect) * (F-1)/F / (N * SD^2)
h/v = 1036.77 * ((10.83-1)/10.83)/ (289650 * 11.74^2) = 0.0000236

For health plan sizes of about 1000, the optimal value of W is very close to 1--that is, we should impute and use the observations. This does not mean that half-scale imputing is better than other types of imputing (see results), but it does mean that if we are missing PF01 and we are given a choice of casewise deletion or using half-scale, we should use half-scale

In the following tables (Tables 2A and 2B), we show h/v for three types of problems: comparison of health plans, comparison of disease groups, and demographic comparisons. The health plans are defined by the initial contract number. The disease groups are defined by indicator variables for the number of serious conditions (CHF, stroke, COPD, arthritis of the hip, sciatica, diabetes, or cancer), the number of minor chronic conditions (hypertension, coronary heart disease, AMI survivor, other heart condition, GI problems, arthritis of the hand), and depression. The demographics include age group (<65, 65-75, 75-85, 85+), gender, non-white vs. white race, income (groups 1-2, 3-4, 5-9, 10 or missing), education (1-2, 3-6, missing) and marital status (married vs. not married). Tables 2C and 2D give the actual bias in points of PCS and MCS.

Problems with various imputation methods can be traced mostly to the fact that items have unique content as well as error. For example, within the PF scale of the SF-36, PF8 is an item that describes walking several blocks and PF9 describes limitations in walking one block. Both are part of the physical functioning scale. The Pearson correlation (same as Spearman) between the two items at time 1 is 0.58. Comparing the two waves, the two PF9 items are correlated 0.61, and the two PF8 items 0.65, but the cross correlations are 0.54 and 0.61. Cross

correlations are only slightly lower, suggesting that just over 90% of the variance is shared (.58/.63) and a little less than 10% is unique. For PF6 (bending, kneeling, and stooping) in relation to PF9, a similar technique tells us that two-thirds of the variance is overlapping and one-third is unique.

Extension to naturally missing data. In addition, we have not considered potential biases from omitting the cases via casewise deletion. Naturally missing data is potentially not random. The fundamental problem is that there is no gold standard to compare against. There are two ways to create what we term here a "gold standard." The first approach is through a trusted missing value method. For example, if we were to use the MDE as the standard, we could apply it to determine how the less-trusted half scale rule performs with respect to the values it cannot estimate. The other approach is to use longitudinal data for comparison purposes.

To illustrate this logic, consider naturally missing pf01 cases. In the first two baseline cohorts of the HOS, 4037 cases are missing only pf01. Cases that were naturally missing pf01 had MDE estimates 6.12 points higher than the HOS average for complete cases. However, an analysis of variance (ANOVA) including plan ID and a missing value indictor for pf01 (all other incomplete cases excluded) had a non-significant plan-missing value interaction (F=0.77). Failing to impute these cases would bias a national mean but would not bias a plan comparison for sample sizes in the HOS range. Any bias we had found would be subtracted from h in our optimality formula, possibly leading to a situation where both bias and variation are improved by using imputation.

If we use the longitudinal data for pf01, we find that missing cases were 4.77 points higher at follow-up and the interaction with plan is not significant (F=0.82), so the information is qualitatively similar. But the observations with missing pf01 decline faster, according to the MDE (-2.84 points vs. -1.94, or 0.90 points (46%) faster than average). Does the faster decline mean that pf01 is a harbinger of decline or is this a defect in the MDE estimate? That is, have we used the information in the missing value to its maximum benefit? Looking at the items, the closest item pf02 declines 59% faster but pf03 declines 24% faster when pf01 is missing. Therefore, it is logical to consider approaches that may be able to capture more of the missing data information and the pattern of missing data that may influence the rate of decline.

These examples illustrate basic dilemmas. Without a missing value estimator we trust, it is hard to grasp the impact of naturally missing data on other estimators. But if we cannot examine the limitations of the estimator, how do we come to consider it a standard? And if we are going to use longitudinal data for validation purposes, how de we factor in the relationship between missing data and subsequent change?

Fortunately there are also some saving graces. Actual naturally missing data patterns have different kinds of bias implications which tend to cancel each other out. The tests with systematic planned missing data are more stringent tests. In actual practice, all of the methods are probably more subtly biased than systematic planned tests suggest. And finally, the customer may not have a highly refined sense of what the outcome should be. A biased result may be equally desirable or even more desirable than the complete-case formula.

4. Validation Results of MOS SF-36

Observations imputable

For HOS cohorts 1and 2, observations were scorable according to Table 1:

Method	Number Scorable	% of possible cases
Casewise deletion	289650	75.4
Half-scale	353426	92.1
MDE (PCS)	370692	96.6
MRE (PCS, r^2>0.5)	383656	99.0
MDE (MCS)	371442	96.8
MRE (MCS, r^2>0.5)	376115	98.0

The cases that are scorable with the MDE are a perfect subset of those scorable with the MRE using a cutoff of r^2>0.5. For PCS, the cases that are scorable with the MRE but not the MDE range up to r^2>0.98 and average over 0.85. For MCS, the additional cases are many fewer, but the average over 0.82 and range up to r^2>0.98. *The results suggest that cases with partially missing SF-36 data are being missed by the MDE approach and have potential value. Results show that two thirds of the missing cases using the MDE approach are captured for PCS by the MRE approach and one third are captured for MCS using the same approach.*

Systematic Planned Missing Data and Gold Standard Validation

Tables 2A, 2B, 2C and 2D. in this section considers several situations where we intentionally make some values missing. The MDE cannot be included in this table because the statistical algorithm was unavailable to experimentally manipulate. The table includes the bias and variation (sd^2) for imputation by the half scale test, RE and MRE. Based upon the means and standard deviation of the bias for each of the missing items, the bias properties and PCS and MCS error points are computed for comparisons between the health plans, disease groups and demographic groups. There are well over 600 regression models for each table.

The bias properties can be converted to the error in PCS and MCS points, using the following formula:

Error in PCS or MCS points = Square root (bias property value/1000) standard deviation (PCS or MCS).*

The following table may be used as an interpretation guideline for Tables 2A and 2B. While one can examine the actual bias in PCS and MCS points in tables 2C and 2D.

Table Entry (all are x1000) Bias	Typical PCS comparison bias (in points)	Typical MCS comparison Bias (in points)
0.0010	0.01174	0.01080
0.0100	0.03713	0.03415
0.1000	0.1174	0.1080
1.0000	0.3713	0.3415
4.0000	0.7426	0.6830
10.000	1.1174	1.0800

For example, if you are comparing the health of those in a health plan with the overall average, the potential bias from using a method with a bias of 4.0000 is 0.7426 PCS points. The typical health plan effect is 1.65 points (standard deviation of the health plans). The signal to noise ratio based upon a variance components analysis is about 5 or equal to (SD of the health plans/SD of the bias)^2 = (1.65/0.74).^{2} These results would be understandable, but somewhat flawed. For disease group comparisons, the average differences are about 5 points, so even a bias of 10.000 would not seriously distort or invalidate the comparison

Table 2A. Bias properties of PCS estimates*

	Means and Standard Deviations						Bias Properties (x1000)								
	Half-Scale		Regression		MRE		Health Plans			Disease Groups			Demographic Groups		
Missing	bias	sd	bias	sd	bias	sd	HS	RE	MRE	HS	RE	MRE	HS	RE	MRE
pf01	0.61	0.55	-0.00	0.47	-0.00	0.47	0.0236	0.0032	0.0033	0.0558	0.0034	0.0026	0.0983	0.0103	0.0107
pf02	0.01	0.41	-0.00	0.39	-0.00	0.39	0.0022	0.0015	0.0015	0.0107	0.0001	0.0003	0.0118	0.0027	0.0024
pf04	0.20	0.43	-0.00	0.40	0.00	0.40	0.0075	0.0032	0.0031	0.0233	0.0015	0.0010	0.0087	0.0017	0.0017
pf06	0.11	0.46	0.00	0.46	0.00	0.46	0.0047	0.0026	0.0024	0.0121	0.0083	0.0070	0.0080	0.0021	0.0024
pf09	-0.30	0.39	0.00	0.33	0.00	0.33	0.0048	0.0009	0.0009	0.0162	0.0003	0.0002	0.0077	0.0006	0.0006
pf10	-0.50	0.50	-0.00	0.36	-0.00	0.36	0.0098	0.0016	0.0015	0.1702	0.0003	0.0005	0.0739	0.0066	0.0064
5pfA	1.13	1.24	-0.00	1.07	-0.00	1.07	0.0929	0.0336	0.0330	0.1688	0.0447	0.0256	0.2358	0.0382	0.0403
5pfB	-0.36	1.01	0.00	0.94	-0.00	0.94	0.0191	0.0189	0.0169	0.0103	0.0157	0.0085	0.0267	0.0118	0.0075
rp2	0.17	0.77	-0.00	0.73	0.00	0.73	0.0077	0.0045	0.0043	0.0032	0.0007	0.0014	0.0132	0.0057	0.0057
rp3	0.07	0.71	0.00	0.68	0.00	0.68	0.0009	0.0002	0.0002	0.0132	0.0008	0.0005	0.0058	0.0038	0.0036
2rp	0.37	1.27	-0.00	1.13	-0.00	1.13	0.0251	0.0113	0.0108	0.0413	0.0031	0.0046	0.0364	0.0090	0.0090
bp1	0.14	1.43	-0.00	1.31	0.00	1.31	0.0269	0.0163	0.0158	0.2130	0.1472	0.0855	0.1013	0.0434	0.0562
bp2	0.32	1.09	0.00	0.84	0.00	0.84	0.0111	0.0034	0.0032	0.0447	0.0042	0.0021	0.0540	0.0041	0.0039
gh1	0.16	0.44	0.00	0.41	0.00	0.41	0.0073	0.0057	0.0055	0.0130	0.0097	0.0082	0.0374	0.0214	0.0204
gh3	0.02	0.54	0.00	0.52	0.00	0.52	0.0040	0.0037	0.0035	0.0020	0.0008	0.0005	0.0126	0.0174	0.0176
2gh	-0.34	0.76	0.00	0.67	0.00	0.67	0.0137	0.0065	0.0065	0.0731	0.0200	0.0139	0.0715	0.0318	0.0314
vt2	0.03	0.06	0.00	0.13	0.00	0.13	0.0000	0.0000	0.0000	0.0003	0.0000	0.0000	0.0002	0.0001	0.0001
vt4	-0.02	0.06	-0.00	0.13	-0.00	0.13	0.0001	0.0000	0.0000	0.0002	0.0000	0.0000	0.0001	0.0001	0.0001
2vt	0.03	0.09	-0.00	0.14	-0.00	0.14	0.0001	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0002	0.0001
sf1	-0.00	0.03	0.00	0.12	0.00	0.12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001
sf2	0.00	0.03	0.00	0.12	-0.00	0.12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001
re1	0.09	0.55	0.00	0.51	0.00	0.51	0.0025	0.0013	0.0013	0.0063	0.0066	0.0073	0.0079	0.0024	0.0027
re2	-0.17	0.58	-0.00	0.56	-0.00	0.56	0.0035	0.0008	0.0009	0.0093	0.0011	0.0018	0.0085	0.0032	0.0041
mh1	0.09	0.50	0.00	0.48	0.00	0.48	0.0028	0.0026	0.0027	0.0099	0.0029	0.0034	0.0133	0.0109	0.0115
mh2	0.35	0.40	-0.00	0.35	0.00	0.35	0.0027	0.0005	0.0005	0.0124	0.0108	0.0112	0.0073	0.0014	0.0015
mh3	-0.41	0.53	0.00	0.49	0.00	0.49	0.0030	0.0025	0.0023	0.0183	0.0021	0.0024	0.0095	0.0054	0.0055
mh4	0.18	0.41	-0.00	0.38	-0.00	0.38	0.0010	0.0005	0.0004	0.0076	0.0132	0.0136	0.0057	0.0050	0.0052
mh5	-0.21	0.50	0.00	0.47	0.00	0.47	0.0032	0.0021	0.0019	0.0051	0.0125	0.0123	0.0117	0.0031	0.0031
2mhA	-0.42	0.79	-0.00	0.73	0.00	0.73	0.0086	0.0082	0.0080	0.0493	0.0116	0.0138	0.0491	0.0404	0.0424
2mhB	0.70	0.82	0.00	0.64	0.00	0.64	0.0104	0.0033	0.0033	0.0686	0.1344	0.1387	0.0259	0.0106	0.0119
2mhC	-0.04	0.71	0.00	0.63	-0.00	0.63	0.0050	0.0043	0.0039	0.0103	0.0662	0.0675	0.0306	0.0110	0.0116
list1	1.30	2.51	-0.00	2.19	-0.00	2.20	0.2740	0.1327	0.1185	0.7157	0.3223	0.1953	0.8709	0.2168	0.2228
list2	-1.35	2.60	-0.00	2.33	-0.00	2.34	0.1283	0.0533	0.0396	0.3097	0.5732	0.2914	0.4848	0.1098	0.1568

* The superior imputation method is shaded in the table. Note that the variable names for the SF-36 are explained in Appendix C, that gives the SF-36 questionnaire.

Table 2B: Bias properties of MCS estimates*

	Means and Standard Deviations						Bias Properties (x1000)								
	Half-Scale		Regression		Modified		Health Plans			Disease Groups			Demographic Groups		
Missing	**bias**	**sd**	**bias**	**sd**	**bias**	**sd**	**HS**	**RE**	**MRE**	**HS**	**RE**	**MRE**	**HS**	**RE**	**MRE**
pf01	-0.33	0.30	0.00	0.25	0.00	0.25	0.0082	0.0011	0.0011	0.0195	0.0011	0.0012	0.0343	0.0038	0.0037
pf02	-0.00	0.22	-0.00	0.21	0.00	0.21	0.0008	0.0006	0.0006	0.0037	0.0000	0.0000	0.0041	0.0009	0.0010
pf04	-0.11	0.23	-0.00	0.21	-0.00	0.21	0.0026	0.0011	0.0011	0.0081	0.0005	0.0006	0.0030	0.0006	0.0006
pf06	-0.06	0.25	-0.00	0.24	-0.00	0.24	0.0016	0.0009	0.0009	0.0042	0.0028	0.0029	0.0028	0.0007	0.0007
pf09	0.16	0.21	-0.00	0.17	0.00	0.17	0.0017	0.0003	0.0003	0.0057	0.0001	0.0001	0.0027	0.0002	0.0002
pf10	0.27	0.27	0.00	0.19	0.00	0.19	0.0034	0.0005	0.0005	0.0594	0.0001	0.0001	0.0258	0.0022	0.0022
5pfA	-0.61	0.67	0.00	0.58	0.00	0.58	0.0324	0.0118	0.0115	0.0589	0.0153	0.0173	0.0823	0.0137	0.0133
5pfB	0.19	0.55	-0.00	0.51	-0.00	0.51	0.0067	0.0065	0.0067	0.0036	0.0054	0.0064	0.0093	0.0039	0.0048
rp2	-0.06	0.27	-0.00	0.26	0.00	0.26	0.0011	0.0007	0.0006	0.0005	0.0001	0.0001	0.0019	0.0008	0.0008
rp3	-0.03	0.25	-0.00	0.24	-0.00	0.24	0.0001	0.0000	0.0000	0.0019	0.0001	0.0001	0.0008	0.0006	0.0006
2rp	-0.13	0.45	0.00	0.39	-0.00	0.39	0.0037	0.0017	0.0016	0.0060	0.0004	0.0005	0.0053	0.0013	0.0014
bp1	-0.04	0.44	0.00	0.40	0.00	0.40	0.0030	0.0018	0.0018	0.0237	0.0164	0.0175	0.0113	0.0048	0.0047
bp2	-0.10	0.33	-0.00	0.26	0.00	0.26	0.0012	0.0004	0.0004	0.0050	0.0005	0.0005	0.0060	0.0005	0.0005
gh1	-0.01	0.03	-0.00	0.04	-0.00	0.04	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0001	0.0001
gh3	-0.00	0.03	0.00	0.05	-0.00	0.05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001
2gh	0.02	0.05	-0.00	0.06	0.00	0.06	0.0001	0.0000	0.0000	0.0003	0.0001	0.0001	0.0003	0.0002	0.0002
vt2	0.27	0.52	0.00	0.44	-0.00	0.44	0.0035	0.0030	0.0031	0.0244	0.0002	0.0004	0.0170	0.0061	0.0066
vt4	-0.14	0.49	0.00	0.42	-0.00	0.42	0.0049	0.0025	0.0025	0.0124	0.0006	0.0004	0.0113	0.0013	0.0014
2vt	0.21	0.73	-0.00	0.65	0.00	0.65	0.0075	0.0043	0.0047	0.0057	0.0017	0.0020	0.0113	0.0136	0.0150
sf1	0.03	1.20	-0.00	0.97	-0.00	0.97	0.0246	0.0143	0.0160	0.0244	0.0031	0.0048	0.0401	0.0351	0.0414
sf2	-0.03	1.20	0.00	0.97	0.00	0.97	0.0246	0.0107	0.0090	0.0244	0.0106	0.0039	0.0401	0.0165	0.0140
re1	-0.21	1.24	-0.00	1.11	-0.00	1.11	0.0150	0.0077	0.0065	0.0381	0.0399	0.0145	0.0479	0.0175	0.0145
re2	0.38	1.32	0.00	1.23	-0.00	1.23	0.0214	0.0049	0.0053	0.0566	0.0063	0.0051	0.0516	0.0206	0.0199
mh1	-0.21	1.10	-0.00	1.02	-0.00	1.03	0.0159	0.0162	0.0155	0.0567	0.0165	0.0092	0.0763	0.0613	0.0566
mh2	-0.76	0.88	-0.00	0.73	-0.00	0.73	0.0153	0.0028	0.0024	0.0715	0.0618	0.0443	0.0418	0.0090	0.0081
mh3	0.91	1.17	0.00	1.05	-0.00	1.05	0.0171	0.0133	0.0149	0.1051	0.0118	0.0038	0.0544	0.0276	0.0293
mh4	-0.39	0.91	-0.00	0.79	-0.00	0.79	0.0056	0.0027	0.0026	0.0437	0.0753	0.0527	0.0330	0.0299	0.0286
mh5	0.45	1.10	-0.00	1.00	-0.00	1.00	0.0186	0.0118	0.0123	0.0294	0.0706	0.0457	0.0670	0.0168	0.0168
2mhA	0.93	1.73	-0.00	1.59	-0.00	1.60	0.0497	0.0469	0.0499	0.2831	0.0662	0.0258	0.2822	0.2265	0.2216
2mhB	-1.54	1.80	0.00	1.38	0.00	1.38	0.0597	0.0190	0.0161	0.3941	0.7718	0.5551	0.1488	0.0643	0.0520
2mhC	0.08	1.57	0.00	1.36	-0.00	1.36	0.0287	0.0245	0.0253	0.0590	0.3787	0.2496	0.1757	0.0637	0.0591
list1	0.59	2.62	-0.00	2.42	-0.00	2.43	0.0884	0.0477	0.0529	0.6437	0.4848	0.1790	0.5876	0.1725	0.1867
list2	0.77	2.66	0.00	2.38	0.00	2.40	0.1140	0.0246	0.0401	0.7812	0.3459	0.2558	0.2862	0.0848	0.1208

* The superior imputation method is shaded in the table. Note that the variable names for the SF-36 are explained in Appendix C, that gives the SF-36 questionnaire.

Table 2C. Error Points of PCS estimates *

	Means and Standard Deviations						PCS Error Points								
	Half-Scale		Regression		MRE		Health Plans			Disease Groups			Demographic Groups		
Missing	bias	sd	bias	sd	bias	sd	HS	RE	MRE	HS	RE	MRE	HS	RE	MRE
pf01	0.61	0.55	-0.00	0.47	-0.00	0.47	0.0570	0.0210	0.0213	0.0877	0.0216	0.0189	0.1164	0.0377	0.0384
pf02	0.01	0.41	-0.00	0.39	-0.00	0.39	0.0174	0.0144	0.0144	0.0384	0.0037	0.0064	0.0403	0.0193	0.0182
pf04	0.20	0.43	-0.00	0.40	0.00	0.40	0.0322	0.0210	0.0207	0.0567	0.0144	0.0117	0.0346	0.0153	0.0153
pf06	0.11	0.46	0.00	0.46	0.00	0.46	0.0255	0.0189	0.0182	0.0408	0.0338	0.0311	0.0332	0.0170	0.0182
pf09	-0.30	0.39	0.00	0.33	0.00	0.33	0.0257	0.0111	0.0111	0.0473	0.0064	0.0053	0.0326	0.0091	0.0091
pf10	-0.50	0.50	-0.00	0.36	-0.00	0.36	0.0368	0.0149	0.0144	0.1532	0.0064	0.0083	0.1009	0.0302	0.0297
5pfA	1.13	1.24	-0.00	1.07	-0.00	1.07	0.1132	0.0681	0.0674	0.1525	0.0785	0.0594	0.1803	0.0726	0.0745
5pfB	-0.36	1.01	0.00	0.94	-0.00	0.94	0.0513	0.0510	0.0483	0.0377	0.0465	0.0342	0.0607	0.0403	0.0322
rp2	0.17	0.77	-0.00	0.73	0.00	0.73	0.0326	0.0249	0.0243	0.0210	0.0098	0.0139	0.0427	0.0280	0.0280
rp3	0.07	0.71	0.00	0.68	0.00	0.68	0.0111	0.0053	0.0053	0.0427	0.0105	0.0083	0.0283	0.0229	0.0223
2rp	0.37	1.27	-0.00	1.13	-0.00	1.13	0.0588	0.0395	0.0386	0.0754	0.0207	0.0252	0.0708	0.0352	0.0352
bp1	0.14	1.43	-0.00	1.31	0.00	1.31	0.0609	0.0474	0.0467	0.1713	0.1424	0.1086	0.1182	0.0773	0.0880
bp2	0.32	1.09	0.00	0.84	0.00	0.84	0.0391	0.0216	0.0210	0.0785	0.0241	0.0170	0.0863	0.0238	0.0232
gh1	0.16	0.44	0.00	0.41	0.00	0.41	0.0317	0.0280	0.0275	0.0423	0.0366	0.0336	0.0718	0.0543	0.0530
gh3	0.02	0.54	0.00	0.52	0.00	0.52	0.0235	0.0226	0.0220	0.0166	0.0105	0.0083	0.0417	0.0490	0.0493
2gh	-0.34	0.76	0.00	0.67	0.00	0.67	0.0435	0.0299	0.0299	0.1004	0.0525	0.0438	0.0993	0.0662	0.0658
vt2	0.03	0.06	0.00	0.13	0.00	0.13	0.0000	0.0000	0.0000	0.0064	0.0000	0.0000	0.0053	0.0037	0.0037
vt4	-0.02	0.06	-0.00	0.13	-0.00	0.13	0.0037	0.0000	0.0000	0.0053	0.0000	0.0000	0.0037	0.0037	0.0037
2vt	0.03	0.09	-0.00	0.14	-0.00	0.14	0.0037	0.0000	0.0000	0.0037	0.0037	0.0037	0.0037	0.0053	0.0037
sf1	-0.00	0.03	0.00	0.12	0.00	0.12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0037	0.0037
sf2	0.00	0.03	0.00	0.12	-0.00	0.12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0037	0.0037
re1	0.09	0.55	0.00	0.51	0.00	0.51	0.0186	0.0134	0.0134	0.0295	0.0302	0.0317	0.0330	0.0182	0.0193
re2	-0.17	0.58	-0.00	0.56	-0.00	0.56	0.0220	0.0105	0.0111	0.0358	0.0123	0.0158	0.0342	0.0210	0.0238
mh1	0.09	0.50	0.00	0.48	0.00	0.48	0.0196	0.0189	0.0193	0.0369	0.0200	0.0216	0.0428	0.0388	0.0398
mh2	0.35	0.40	-0.00	0.35	0.00	0.35	0.0193	0.0083	0.0083	0.0413	0.0386	0.0393	0.0317	0.0139	0.0144
mh3	-0.41	0.53	0.00	0.49	0.00	0.49	0.0203	0.0186	0.0178	0.0502	0.0170	0.0182	0.0362	0.0273	0.0275
mh4	0.18	0.41	-0.00	0.38	-0.00	0.38	0.0117	0.0083	0.0074	0.0324	0.0427	0.0433	0.0280	0.0263	0.0268
mh5	-0.21	0.50	0.00	0.47	0.00	0.47	0.0210	0.0170	0.0162	0.0265	0.0415	0.0412	0.0402	0.0207	0.0207
2mhA	-0.42	0.79	-0.00	0.73	0.00	0.73	0.0344	0.0336	0.0332	0.0824	0.0400	0.0436	0.0823	0.0746	0.0764
2mhB	0.70	0.82	0.00	0.64	0.00	0.64	0.0379	0.0213	0.0213	0.0972	0.1361	0.1383	0.0597	0.0382	0.0405
2mhC	-0.04	0.71	0.00	0.63	-0.00	0.63	0.0263	0.0243	0.0232	0.0377	0.0955	0.0965	0.0649	0.0389	0.0400
list1	1.30	2.51	-0.00	2.19	-0.00	2.20	0.1943	0.1352	0.1278	0.3141	0.2108	0.1641	0.3465	0.1729	0.1752
list2	-1.35	2.60	-0.00	2.33	-0.00	2.34	0.1330	0.0857	0.0739	0.2066	0.2811	0.2004	0.2585	0.1230	0.1470

* The superior imputation method is shaded in the table. Note that the variable names for the SF-36 are explained in Appendix C, that gives the SF-36 questionnaire.

Table 2D: Error Points of MCS estimates *

	Means and Standard Deviations						MCS Error Points								
	Half-Scale		Regression		Modified		Health Plans			Disease Groups			Demographic Groups		
Missing	bias	sd	bias	sd	bias	sd	HS	RE	MRE	HS	RE	MRE	HS	RE	MRE
pf01	-0.33	0.30	0.00	0.25	0.00	0.25	0.0309	0.0113	0.0113	0.0476	0.0113	0.0118	0.0631	0.0210	0.0207
pf02	-0.00	0.22	-0.00	0.21	0.00	0.21	0.0096	0.0084	0.0084	0.0207	0.0000	0.0000	0.0218	0.0102	0.0108
pf04	-0.11	0.23	-0.00	0.21	-0.00	0.21	0.0174	0.0113	0.0113	0.0307	0.0076	0.0084	0.0187	0.0084	0.0084
pf06	-0.06	0.25	-0.00	0.24	-0.00	0.24	0.0136	0.0102	0.0102	0.0221	0.0180	0.0184	0.0180	0.0090	0.0090
pf09	0.16	0.21	-0.00	0.17	0.00	0.17	0.0141	0.0059	0.0059	0.0257	0.0034	0.0034	0.0177	0.0048	0.0048
pf10	0.27	0.27	0.00	0.19	0.00	0.19	0.0199	0.0076	0.0076	0.0831	0.0034	0.0034	0.0548	0.0160	0.0160
5pfA	-0.61	0.67	0.00	0.58	0.00	0.58	0.0614	0.0370	0.0366	0.0827	0.0422	0.0448	0.0978	0.0399	0.0393
5pfB	0.19	0.55	-0.00	0.51	-0.00	0.51	0.0279	0.0275	0.0279	0.0205	0.0251	0.0273	0.0329	0.0213	0.0236
rp2	-0.06	0.27	-0.00	0.26	0.00	0.26	0.0113	0.0090	0.0084	0.0076	0.0034	0.0034	0.0149	0.0096	0.0096
rp3	-0.03	0.25	-0.00	0.24	-0.00	0.24	0.0034	0.0000	0.0000	0.0149	0.0034	0.0034	0.0096	0.0084	0.0084
2rp	-0.13	0.45	0.00	0.39	-0.00	0.39	0.0207	0.0141	0.0136	0.0264	0.0068	0.0076	0.0248	0.0123	0.0128
bp1	-0.04	0.44	0.00	0.40	0.00	0.40	0.0187	0.0145	0.0145	0.0525	0.0437	0.0451	0.0362	0.0236	0.0234
bp2	-0.10	0.33	-0.00	0.26	0.00	0.26	0.0118	0.0068	0.0068	0.0241	0.0076	0.0076	0.0264	0.0076	0.0076
gh1	-0.01	0.03	-0.00	0.04	-0.00	0.04	0.0000	0.0000	0.0000	0.0034	0.0034	0.0034	0.0048	0.0034	0.0034
gh3	-0.00	0.03	0.00	0.05	-0.00	0.05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0034	0.0034	0.0034
2gh	0.02	0.05	-0.00	0.06	0.00	0.06	0.0034	0.0000	0.0000	0.0059	0.0034	0.0034	0.0059	0.0048	0.0048
vt2	0.27	0.52	0.00	0.44	-0.00	0.44	0.0202	0.0187	0.0190	0.0532	0.0048	0.0068	0.0444	0.0266	0.0277
vt4	-0.14	0.49	0.00	0.42	-0.00	0.42	0.0239	0.0170	0.0170	0.0380	0.0084	0.0068	0.0362	0.0123	0.0128
2vt	0.21	0.73	-0.00	0.65	0.00	0.65	0.0295	0.0224	0.0234	0.0257	0.0141	0.0152	0.0362	0.0398	0.0418
sf1	0.03	1.20	-0.00	0.97	-0.00	0.97	0.0535	0.0408	0.0431	0.0532	0.0190	0.0236	0.0683	0.0639	0.0694
sf2	-0.03	1.20	0.00	0.97	0.00	0.97	0.0535	0.0353	0.0323	0.0532	0.0351	0.0213	0.0683	0.0438	0.0403
re1	-0.21	1.24	-0.00	1.11	-0.00	1.11	0.0418	0.0299	0.0275	0.0665	0.0681	0.0410	0.0746	0.0451	0.0410
re2	0.38	1.32	0.00	1.23	-0.00	1.23	0.0499	0.0239	0.0248	0.0811	0.0271	0.0243	0.0774	0.0489	0.0481
mh1	-0.21	1.10	-0.00	1.02	-0.00	1.03	0.0430	0.0434	0.0424	0.0812	0.0438	0.0327	0.0942	0.0844	0.0811
mh2	-0.76	0.88	-0.00	0.73	-0.00	0.73	0.0422	0.0180	0.0167	0.0912	0.0847	0.0717	0.0697	0.0323	0.0307
mh3	0.91	1.17	0.00	1.05	-0.00	1.05	0.0446	0.0393	0.0416	0.1105	0.0370	0.0210	0.0795	0.0566	0.0584
mh4	-0.39	0.91	-0.00	0.79	-0.00	0.79	0.0255	0.0177	0.0174	0.0713	0.0935	0.0783	0.0619	0.0589	0.0577
mh5	0.45	1.10	-0.00	1.00	-0.00	1.00	0.0465	0.0370	0.0378	0.0585	0.0906	0.0729	0.0882	0.0442	0.0442
2mhA	0.93	1.73	-0.00	1.59	-0.00	1.60	0.0760	0.0738	0.0761	0.1814	0.0877	0.0548	0.1811	0.1622	0.1605
2mhB	-1.54	1.80	0.00	1.38	0.00	1.38	0.0833	0.0470	0.0433	0.2140	0.2995	0.2540	0.1315	0.0864	0.0777
2mhC	0.08	1.57	0.00	1.36	-0.00	1.36	0.0578	0.0534	0.0542	0.0828	0.2098	0.1703	0.1429	0.0860	0.0829
list1	0.59	2.62	-0.00	2.42	-0.00	2.43	0.1014	0.0745	0.0784	0.2735	0.2374	0.1442	0.2613	0.1416	0.1473
list2	0.77	2.66	0.00	2.38	0.00	2.40	0.1151	0.0535	0.0683	0.3013	0.2005	0.1724	0.1824	0.0993	0.1185

* The superior imputation method is shaded in the table. Note that the variable names for the SF-36 are explained in Appendix C, that gives the SF-36 questionnaire.

If an item is named in the above list, then only that item is missing. If the entry starts with a number, then that number of items is missing from that scale. List1 and List2 are two lists of 16 items that press the half-scale limits by deleting the largest number of responses from each scale possible before the half-scale cannot score it at all. All of the scenarios in the above are situations which the half-scale rule can score.

Overall, in situations with only moderate amounts of missing data -those that could be scored with the half-scale rule, the regression estimates and the modified regression estimates show no bias. They have lower variance with respect to the true score too. The zero biases mean that in HOS-like population studies, the regression estimators can be relied on to give good means. The rest of the table speaks to the question of how generalizable or valid this result might be. All three estimators have low h/v ratios--the numbers shown in the table had to be multiplied by 1000 just to be readable. A value of 0.1 means that PCS is typically accurate to 0.12 points when looking at a subgroup $[(h/v)^{0.5}*11.74]$. The low h/v values imply that the two unbiased estimates would be equally unbiased in any subgroup of the elderly described by the disease and demographic variables, and the half-scale estimate would be equally biased. For calculating population averages, only the regression estimators (simple and modified) are suitable.

Since the h/v ratios are small, any of the three estimators would work in comparative studies. But the regression and modified regression estimates are clearly better than the half-scale for almost all patterns of missing data. Between the two regression estimators, they are fairly equal, though the MRE was superior most often (those superior methods are shaded cells in the tables).

The more interesting analyses comparing simple and modified regression estimators are those with greater amounts of missing data--beyond what the half-scale rule can accommodate. A number of these missing data patterns have practical importance because they have been used in several studies. While tables 2A and 2B and 2C and 2D provided missing value cases that are less extreme, table 3A and 3B provides the bias properties using the RE and MRE approaches for selected missing value conditions that are more extreme. Tables 3C and 3D give the bias properties in PCS and MCS points. For these selected cases, the half scale rule is not appropriate for the selected scenarios related to these more extreme missing value conditions.

Table 3A:PCS bias properties (h/v) of Regression Estimates (x1000)

		Health Plans		Health Conditions		Demographics	
Values	**R-sq**	**RE**	**MRE**	**RE**	**MRE**	**RE**	**MRE**
list1	96.5	0.1327	0.1185	0.3223	0.1953	0.2168	0.2228
list2	96.0	0.0533	0.0396	0.5732	0.2914	0.1098	0.1568
NoPF*	92.0	0.2906	0.2185	1.0895	0.4623	3.6327	2.7901
OnlyPF7	96.2	0.1371	0.1131	0.2181	0.0783	0.7438	0.5483
OnlyPF10	93.5	0.1805	0.1450	0.8774	0.3463	2.4218	1.8958
NoRERP*	94.6	0.0550	0.0767	0.2478	0.4812	0.1219	0.2543
SF18	93.7	0.1589	0.1587	1.1052	0.4349	0.1877	0.1022
SF12	93.1	0.1661	0.1610	1.5652	0.8879	0.1419	0.1079
SF8	89.4	0.3840	0.3163	2.1740	1.5139	0.8158	0.5538
SF6a*	84.8	0.5430	0.4100	5.8209	3.3593	1.0113	0.7319
SF6b*	85.5	0.3623	0.3730	2.8183	2.8832	1.4615	1.5611
SF3a*	76.2	1.3640	1.0055	8.3625	3.2857	7.1235	4.2326
SF3b*	74.0	1.0716	0.5092	13.8076	6.1482	6.8464	2.4793
SF1*	47.4	4.1002	4.3167	49.6548	22.7774	18.8844	13.9126

Table 3B: MCS bias properties (h/v) of Regression Estimates (x1000)

		Health Plans		Health Conditions		Demographics	
Values	**R-sq**	**RE**	**MRE**	**RE**	**MRE**	**RE**	**MRE**
list1	95.0	0.0477	0.0529	0.4848	0.1790	0.1725	0.1867
list2	95.1	0.0246	0.0401	0.3459	0.2558	0.0848	0.1208
NoPF	97.2	0.1010	0.1221	0.3792	0.4953	1.2650	1.4732
OnlyPF7	98.7	0.0475	0.0534	0.0758	0.1025	0.2583	0.3002
OnlyPF10	97.7	0.0628	0.0719	0.3052	0.3884	0.8437	0.9649
NoRERP*	88.1	0.1360	0.0989	1.1625	0.3010	0.6813	0.5108
SF18	97.0	0.0514	0.0607	0.2583	0.2518	0.1175	0.1831
SF12	92.7	0.0664	0.0653	1.8937	0.8988	0.3972	0.4656
SF8	86.1	0.3168	0.4406	4.6375	2.0138	0.6473	0.9330
SF6a*	71.8	0.7858	0.5743	14.6012	4.9038	2.6499	1.8169
SF6b*	67.7	0.4514	0.3972	17.2094	5.2446	1.5398	0.5661
SF3a*	59.5	0.5881	0.3797	23.9794	2.5605	3.2436	0.8088
SF3b*	58.9	0.9602	0.3253	28.7366	5.0421	4.8791	2.1762
SF1*	22.2	3.1993	4.9513	196.2021	132.6086	15.0800	10.6728

Table 3C: PCS Error Points of Regression Estimates

		Health Plans		Health Conditions		Demographics	
Values	**R-sq**	**RE**	**MRE**	**RE**	**MRE**	**RE**	**MRE**
list1	96.5	0.1352	0.1278	0.2108	0.1641	0.1729	0.1752
list2	96.0	0.0857	0.0739	0.2811	0.2004	0.1230	0.1470
NoPF*	92.0	0.2001	0.1735	0.3875	0.2524	0.7076	0.6201
OnlyPF7	96.2	0.1375	0.1249	0.1734	0.1039	0.3202	0.2749
OnlyPF10	93.5	0.1577	0.1414	0.3477	0.2185	0.5777	0.5112
NoRERP*	94.6	0.0871	0.1028	0.1848	0.2575	0.1296	0.1872
SF18	93.7	0.1480	0.1479	0.3903	0.2448	0.1608	0.1187
SF12	93.1	0.1513	0.1490	0.4645	0.3498	0.1398	0.1219
SF8	89.4	0.2301	0.2088	0.5474	0.4568	0.3353	0.2763
SF6a*	84.8	0.2736	0.2377	0.8957	0.6804	0.3733	0.3176
SF6b*	85.5	0.2235	0.2267	0.6232	0.6304	0.4488	0.4639
SF3a*	76.2	0.4336	0.3723	1.0736	0.6729	0.9909	0.7638
SF3b*	74.0	0.3843	0.2649	1.3795	0.9205	0.9714	0.5846
SF1*	47.4	0.7517	0.7713	2.6161	1.7718	1.6133	1.3848

Table 3D: MCS Error Points of Regression Estimates

		Health Plans		Health Conditions		Demographics	
Values	**R-sq**	**RE**	**MRE**	**RE**	**MRE**	**RE**	**MRE**
list1	95.0	0.0745	0.0784	0.2374	0.1442	0.1416	0.1473
list2	95.1	0.0535	0.0683	0.2005	0.1724	0.0993	0.1185
NoPF	97.2	0.1083	0.1191	0.2099	0.2399	0.3834	0.4138
OnlyPF7	98.7	0.0743	0.0788	0.0939	0.1091	0.1733	0.1868
OnlyPF10	97.7	0.0854	0.0914	0.1883	0.2125	0.3131	0.3349
NoRERP*	88.1	0.1257	0.1072	0.3675	0.1870	0.2814	0.2436
SF18	97.0	0.0773	0.0840	0.1733	0.1711	0.1169	0.1459
SF12	92.7	0.0878	0.0871	0.4691	0.3232	0.2148	0.2326
SF8	86.1	0.1919	0.2263	0.7341	0.4838	0.2743	0.3293
SF6a*	71.8	0.3022	0.2583	1.3026	0.7549	0.5549	0.4595
SF6b*	67.7	0.2290	0.2148	1.4142	0.7807	0.4230	0.2565
SF3a*	59.5	0.2614	0.2101	1.6693	0.5455	0.6139	0.3066
SF3b*	58.9	0.3340	0.1944	1.8274	0.7655	0.7530	0.5029
SF1*	22.2	0.6097	0.7585	4.7750	3.9256	1.3238	1.1137

'List1' and 'list2' are the same as in the earlier tables--they are maximal subsets that could have been scored with the half-scale rule. 'NoPF' is a scenario where all PF (physical function) items are missing but all other scales are present. 'OnlyPF7' and 'OnlyPF10' are two scenarios where just one PF item is present, along with all other scales. 'NoRERP' is a version where (role limitations due to physical or emotional problems) RE and RP are completely missing, but all other items are present. 'SF12' is the classical SF-12.[9] 'SF18' is a special case where a short version of the SF-36 called the SF-12 is included plus unused vitality and mental health items (vt1,vt3,vt4 and mh1, mh2, mh5,) --often used in mental health studies. 'SF8' is a scale that combined 1 item from each of the other scales (pf08, rp2, bp1, gh1, vt2, sf2, re2, and mh2). 'SF6a' is a 6-item version (pf02, pf04, bp2, vt2, mh1, and mh4) used in prior work related to managed care populations. 'SF6b' is a set of 6 items used in ongoing studies of prescription drug benefits (pf04, pf09, bp1, gh1, vt2, and mh2). 'SF3a' is the best 3-item subset of 'SF6b' (pf4, bp1, mh2) and an even shorter version of health, that includes a single item each for physical, psychological and pain. 'SF3b' is another 3 item version (pf06, bp2, and mh4) of the same three concepts to show what happens in another casually-chosen minimal case. 'SF1' is the single gh1 item of overall health from excellent to poor. The combinations marked by an asterisk (*) are combinations that cannot be scored with the MDE estimator. There are a total of six combinations each for PCS and MCS that cannot be scored using the MDE approach.

Considering what they are, the missing value estimators seem to be usable if as few as 3 items are present, as long as they draw from the three main concepts, physical, bodily pain and mental health --PF, BP, and MH. It is possible that other configurations would work, but we did not test them. With an h/v of 4.3×10^{-3}, the SF-1 (gh1 item) would have a typical *error* of 0.76 points as an estimate of PCS, about a third of the health plan PCS effect (determined to be 1.65 points by variance components). This means that about 21% of the variation is "off concept" relative to the PCS, but 79% is on-concept in this extreme case. For determination of disease means, the error is 1.77 points, but these means often differ by 5-10 points.

The advantage of the MRE compared to the simple regression estimator becomes more important when we are dealing with more extreme imputation of missing values, particularly for MCS. 'SF1' would lead to errors of several points, but 'SF3a' seems to be quite usable with errors of about 0.5 points, typically.

We can't say much about the MDE in these analyses because we did not have access to a convenient algorithm to score it the millions of times needed for simulation. The web sight referred to by QualityMetric™ did not allow us to perform this exercise. We can however, compare the behavior of the MDE and MRE in naturally missing data. These indicate that the two estimators are fairly close, differing by a mean of -0.012 (MDE is lower) with a SD of 0.40 between them. That suggests they will be within 1 point of each other almost all the time. For example, when 'PF01' is missing, MDE is lower by 0.13 points, and if 'PF10' is missing, MDE is higher by .067 points. *A multiple regression of the MDE on the half-scale rule and the MRE suggests that the MDE is closer to the half-scale rule than it is to the MRE, particularly for MCS. However, the correlation between the half-scale rule, the MDE, and the MRE gives coefficients of 0.9997 and higher.*

Randomized Planned Missing Data

No tables are currently included with Randomized Planned Missing Data. If we had them, they would support the idea that randomized subsets have less bias than planned subsets.

Naturally Missing Data

We can correlate imputed baseline observations with follow-up values over time that are scored with all items present. For this analysis the 1998 and 1999 baseline and 2000 and 2001 follow-up data was used from the HOS. If the summaries are scorable with half scale rules, then corr($PCS_{baseline\text{-}HS}$,$PCS_{F\text{-}up\text{-}HS}$) = 0.6774, corr($PCS_{baseline\text{-}MDE}$, $PCS_{F\text{-}up\text{-}MDE}$) = 0.6778, and corr($PCS_{Baseline\text{-}MRE}$, $PCS_{F\text{-}up\text{-}MRE}$) = 0.6792. If the summaries are not scorable with half-scale rules but the MDE can be calculated, then corr($PCS_{baseline\text{-}MDE}$, $PCS_{F\text{-}up\text{-}MDE}$) = 0.6688 and corr($PCS_{baseline\text{-}MRE}$, $PCS_{F\text{-}up\text{-}MRE}$) = 0.6807. If only the MRE can be scored and its r^2>0.5, then corr($PCS_{baselineMRE}$, $PCS_{F\text{-}up\text{-}MRE}$) = 0.6861 *The MRE is consistent across the three scenarios and is always the best available approach* ***for PCS*** *and the MDE is consistently better than the half scale rule.*

For MCS, if the summaries are scorable with half scale rules, then corr($MCS_{baseline\text{-}HS}$,$MCS_{F\text{-}up\text{-}HS}$) = 0.5344, corr($MCS_{baseline\text{-}MDE}$, $MCS_{F\text{-}up\text{-}MDE}$) = 0.5351, and corr($MCS_{baseline\text{-}MRE}$, $MCS_{F\text{-}up\text{-}MRE}$) = 0.5368. If the summaries are not scorable with half-scale rules but the MDE can be calculated, then corr($MCS_{baselineMDE}$, $MCS_{F\text{-}up\text{-}MDE}$) = 0.5379 and corr($MCS_{baseline\text{-}MRE}$, $MCS_{F\text{-}up\text{-}MRE}$) = 0.5446. If only the MRE can be scored and its r^2>0.5, then corr($MCS_{baseline\text{-}MRE}$, $MCS_{F\text{-}up\text{-}MRE}$) = 0.5453. *The MRE is consistent across the three scenarios and is always the best available approach* ***for MCS,*** *and the MDE consistently better than the half-scoring rule.*

If the follow-up over time is scored with the MDE instead of complete cases, the above relative statements remain true, but the actual correlations differ a little. That is, *the MRE is consistent across the three scenarios and has the highest correlation with follow-up.*

Given that the MRE seems to be a reasonable estimator and can be used as a standard, we can use it to evaluate how naturally missing data *bias* the conclusions if they do not lead to an imputation. We can evaluate 3 missing value imputation rules in this way where we compare the complete case rule, half scale rule and MDE to the MRE as a kind of standard. Tables 4A and 4B give the differences in the bias between the cases that can be scored using imputation from the cases that cannot be scored. This is reflected using the mean bias of this difference. Tables 4C and 4D give the differences in the bias expressed as PCS and MCS in points between cases scored from those that cannot be scored.

Table 4A: PCS Bias due to naturally missing data Compared with the MRE Approach as the Standard

Imputation Algorithm	MeanBias	Bias (h/v) x 1000 Health Plans	Conditions	Demographics
Complete case	-1.94	0.2096	0.6470	1.5821
Half-scale rule	-3.26	0.5555	0.3043	1.4547
MDE	-2.42	0.2654	0.4839	0.7555

Table 4B: MCS Bias due to naturally missing data Compared with the MRE Approach as the Standard

Imputation Algorithm	MeanBias	Bias (h/v) x 1000 Health Plans	Conditions	Demographics
Complete case	-2.20	0.0000	0.9806	0.2919
Half-scale rule	-3.49	0.3733	0.4370	0.2203
MDE	-4.37	0.5516	0.1136	0.0860

Table 4C: PCS Error Points due to naturally missing data Compared with the MRE Approach as the Standard

Imputation Algorithm	MeanBias	Bias (h/v) x 1000 Health Plans	Conditions	Demographics
Complete case	-1.94	0.1700	0.2986	0.4670
Half-scale rule	-3.26	0.2767	0.2048	0.4478
MDE	-2.42	0.1913	0.2583	0.3227

Table 4D: MCS Error Points due to naturally missing data Compared with the MRE Approach as the Standard

Imputation Algorithm	MeanBias	Bias (h/v) x 1000 Health Plans	Conditions	Demographics
Complete case	-2.20	0.0000	0.3376	0.1842
Half-scale rule	-3.49	0.2083	0.2254	0.1600
MDE	-4.37	0.2532	0.1149	0.1000

The "MeanBias" column in table 4A – 4D describes how cases that cannot be scored with the imputation algorithm differ from those that can be scored. The impact is proportional to this number times the percentage that is missing. For table 4A and 4C that describes PCS, the half-scale rule gives about 92% of the cases for PCS, so the mean bias associated with not scoring it is 8% times -3.26 or about -0.26 bias points. For PCS the equivalent error in points is 0.15. For MDE, the MDE gives 96.6% cases, so the mean bias associated with not scoring is 3.4% times -2.42 or about -0.08 bias points which is equivalent to about the same PCS error in points. For health plan comparisons the MDE approach gives about 0.19 point error for PCS compared with the MRE approach. For table 4B and 4D, the mean bias associated with not scoring using the half scale rule is 8% times -3.49 or about -0.28 bias points for MCS which is equivalent to about 0.18 point error. For health plan comparisons for the MDE approach, there is an error of 0.25 points for MCS compared with the MRE approach.

The remaining columns should be interpreted similar to the systematic planned tables above. That is, the biases shown have been multiplied by 1000. Although none of these biases is serious, they offset typical biases from imputation. They also suggest that the MRE approach is less biased compared to the half scale rule and the MDE, although marginally for the MDE.

5. Implications for Analysis

Based on the above analysis, the MRE is our most preferred method if the goal is to replicate original values of the SF-36 summaries in a point in time. For follow-up data over time the MRE provides slightly better estimates than the MDE approach. The MDE while almost

comparable to the MRE is not fully available to us since it requires enormous computer resource requirements for scoring and unknown algorithms that are not available to the public. We conclude that the MRE method is the more reasonable approach for estimating individual scale values (e.g. PF, RP, etc) of the SF-36.

Given, that we have selected a preferred method and know about the bias typically associated with it, how should estimation be done using this approach? The following points should be kept in mind:

a. For complete cases, we use the complete case value.

b. For incomplete cases that can use the half scale rule we do NOT follow the half scale rule

c. We should score the incomplete cases using the MRE method so long as the MRE reaches the threshold of acceptable performance--we suggest an R-squared of greater than 0.5.

As the MRE approach results in very little bias--even when we have used fairly extreme cases of missing values. We do not suggest weighting the imputed data. The observations imputed with the MRE should be used without weights..

6. Conclusions

When the SF-36 was originally proposed, the half-scale rule seemed like a good method for imputation of missing values. It was straightforward and easily programmed. It handled most practical circumstances.

In recent years other more sophisticated approaches have been developed. The MDE is rooted in a sound and currently popular theory of scale psychometrics (Item Response Theory).[10] Its main disadvantages are the complicated and proprietary nature of the software. The regression imputation is based on older regression technologies, but is an order of magnitude more complicated than the half-scale rule. The MDE requires a complicated software program to run effectively, and the means to do that within popular computer software has evolved with the speed of the computers and the sophistication of software programs (e.g. SAS and STATA) The MRE, described here, employs a simple yet effective correction for regression to the mean that makes the regression estimate more general (and therefore less biased) than it would otherwise be.

We found that failing to impute resulted in more bias in the results than imputing the results. The MRE has relatively small imputation biases which cancel out in naturally missing data, but the biases due to not imputing (and losing the cases) are consistent. The more that data are imputed, the less biased the overall answers will be, and they will also be more accurate due to the additional sample size. Although this statement applies most to the MRE, it would also apply to the MDE when compared to the half-scale rule.

Our ability to directly compare the MRE and the MDE was limited given that the algorithm was not available to us. The evidence though in the results available to us using the Web site that calculates results using the MDE approach through QualityMetric™ suggests the MRE methodology is better. The MRE method imputes more cases and so should be both less biased and lower in variance. In addition, the correlation analysis produced better agreement between the MRE and follow-up data than between the MDE and follow-up data, even if the MDE was used for follow-up. The MDE appears to retain some affinity to the half-scale rule--though it is far better than the half-scale rule. We do not take the use of Item Response Theory (IRT) to be an advantage of the MDE, but we do not know it is a disadvantage either. The MDE is just another approach using an IRT statistical model that needs to be trained. These negatives could be offset by possible advantages of the MDE in non-HOS populations, since the MDE presumably had a more diverse training set and therefore might be more generalizable.

The mean bias of unimputed cases was negative in all cases. This implies that when patients don't fill out lots of items, their health is typically poorer than when they do fill out all or most of the items. However, the illustration in methods for this report suggested that was not true for every item. Neither the MRE (nor we think the MDE) address the fundamental question of whether the naturally missing nature of the items conveys information beyond being missing at random, once the values of the other items have been properly taken into account. Nor have we addressed the interesting question of whether missing data somehow signals impending change in the SF-36.

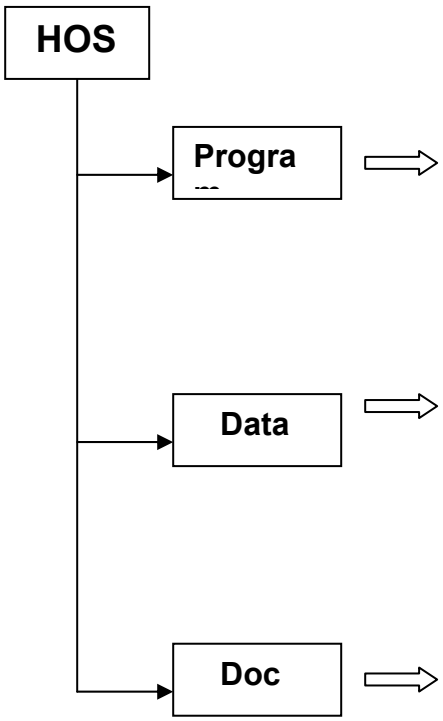
Last, appendix D gives the same tables with some description that provide validation of the Veterans SF-36. The Veterans SF-36 performed as well as the HOS SF-36 using the same imputation approaches. The MRE approach is strongly recommended for the Veterans SF-36 over the half scale rule.

Appendices A.-D.

The following appendices give: The contents of the CD-Rom (appendix A), The use of software for the HOS (appendix B), Description of Variables and Questions of the SF-36 for the HOS (appendix C), and Appendix D: 1. Validation Studies of the Physical and Mental Summary Scores of the Veterans SF-36 Health Survey (appendix D1.) and 2. Description of Variables and Questions of the Veterans SF-36 (appendix D2.)

Appendix A: Contents of the CD-ROM

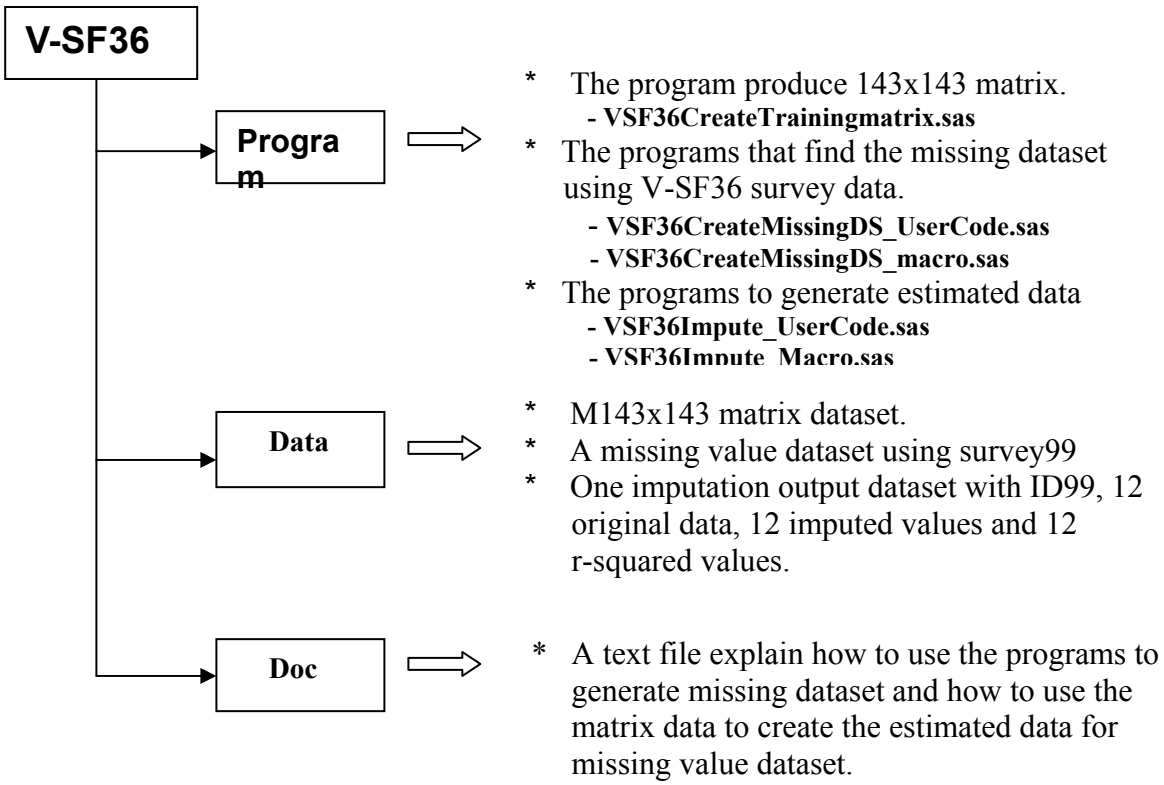
The CD-ROM contains 2 folders, one for HOS and one for the V/SF36. Each folder contains programs, documentation, and test data--the structure of the two is similar. The HOS test data include regression and modified regression missing data estimates for all HOS cohorts available to the Veterans Administration evaluation group as of July 1, 2003. The following diagram shows this:



- The program produce 122x122 matrix.
  - **HOSCreateTraningMatrix.sas**
- The programs that find the missing dataset using HOS data.
  - **HOSCreateMissingDS_UserCode.sas**
  - **HOSCreateMissingDS_macro.sas**
- The Programs generate estimated pcs mcs pcs90 mcs90 and SF36 scores for HOS data. (PCS MCS are based upon the '98 scoring rule')

- M122x122 matrix dataset using cohort 1 and 2 combined baseline data.
- 5 Missing value datasets for all 5 data files (3 baselines + 2 follow-ups) with PATID.
- One imputation output dataset with PATID, original data, imputed data and r-squared values for baseline4 data.

- A text file explain how to use the programs to generate missing dataset and how to use the matrix data to generate the estimates for missing value dataset.



Appendix B: Use of Software For HOS and Veterans SF-36

The theory and use of the software is described on the CD-ROM.

Briefly, the estimation process consists of three phases. These were programmed in SAS Version 8 using PROC IML. In the first phase we create a training matrix. We have created it based on HOS cohort 1 and 2 baselines, but we included the program so if the user wants to use another training dataset, our method can be applied. The second and third phases must be run each time a new dataset with missing data is to be processed. First, the data must be regularized (CreateMissingDataset_UserCode) and then the missing data need to be imputed (HOSImpute_UserCode). A similar structure is used for the VA data (Veterans SF-36).

Appendix C: Description of Variables and Questions of the SF-36 for the HOS

Description of the variables and questions of the SF-36 for the HOS referred to in the text of this document is given in this appendix.

1. In general, would you say your health is: (GH1)

Excellent Very	good	Good	Fair	Poor
1☐	2☐	3☐	4☐	5☐

2. The following items are about activities you might do during a typical day. Does **your health now limit you** in these activities? If so, how much?

ACTIVITIES	Yes, limited a lot	Yes, limited a little	No, not limited at all
a. **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports.....(PF1)	1☐	2☐	3☐
b. **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.....(PF2)	1☐	2☐	3☐
c. Lifting or carrying groceries.....(PF3)	1☐	2☐	3☐
d. Climbing **several** flights of stairs.....(PF4)	1☐	2☐	3☐
e. Climbing **one** flight of stairs.....(PF5)	1☐	2☐	3☐
f. Bending, kneeling, or stooping.....(PF6)	1☐	2☐	3☐
g. Walking **more than a mile**.....(PF7)	1☐	2☐	3☐
h. Walking **several blocks**.....(PF8)	1☐	2☐	3☐
i. Walking **one block**.....(PF9)	1☐	2☐	3☐
j. Bathing or dressing yourself.....(PF10)	1☐	2☐	3☐

3. During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities **as a result of your physical health**?

	Yes	No
a. Cut down on the **amount of time** you spent on work or other activities.....(RP1)	1☐	2☐
b. **Accomplished less** than you would like .....(RP2)	1☐	2☐
c. Were limited in the **kind** of work or other activities (RP3)	1☐	2☐

d. Had **difficulty** performing the work or other activities (*for example*, it took extra effort) .....(RP4) ☐1 ☐2

4. During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities **as a result of any emotional problems** (such as feeling depressed or anxious)?

		Yes	No
a.	Cut down on the **amount of time** you spent on work or other activities.....(VRE1)	☐1	☐2
b.	**Accomplished less** than you would like .....(VRE2)	☐1	☐2
c.	Didn't do work or other activities as **carefully** as usual.....(VRE3)	☐1	☐2

5. During the **past 4 weeks**, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups? (SF1)

Not at all	Slightly	Moderately	Quite a bit	Extremely
☐1	☐2	☐3	☐4	☐5

6. How much **bodily** pain have you had during the **past 4 weeks**? (BP1)

None	Very mild	Mild	Moderate	Severe	Very severe
☐1	☐2	☐3	☐4	☐5	☐6

7. During the **past 4 weeks**, how much did **pain** interfere with your normal work (including both work outside the home and housework)? (BP2)

Not at all	A little bit	Moderately	Quite a bit	Extremely
☐1	☐2	☐3	☐4	☐5

8. These questions are about how you feel and how things have been with you during the **past 4 weeks**. For each question, please give the one answer that comes closest to the way you have been feeling.

How much of the time during the **past 4 weeks**...	**All of the time**	**Most of the time**	**A good bit of the time**	**Some of the time**	**A little of the time**	**None of the time**
a. did you feel full of pep? .....(VT1)	1☐	2☐	3☐	4☐	5☐	6☐
b. have you been a very nervous person? .....(MH1)	1☐	2☐	3☐	4☐	5☐	6☐
c. have you felt so down in the dumps that nothing could cheer you up? .....(MH2)	1☐	2☐	3☐	4☐	5☐	6☐
d. have you felt calm and peaceful?....	1☐	2☐	3☐	4☐	5☐	6☐
e. did you have a lot of energy? (MH3) (VT2)	1☐	2☐	3☐	4☐	5☐	6☐
f. have you felt downhearted and blue? .....(MH4)	1☐	2☐	3☐	4☐	5☐	6☐
g. did you feel worn out? .....(VT3)	1☐	2☐	3☐	4☐	5☐	6☐
h. have you been a happy person?	1☐	2☐	3☐	4☐	5☐	6☐
i. did you feel tired? .....(MH5) (VT4)	1☐	2☐	3☐	4☐	5☐	6☐

9. During the **past 4 weeks**, how much of the time has your **physical health or emotional problems** interfered with your social activities (like visiting with friends, relatives, etc.)? (SF2)

All of the time	**Most of the time**	**Some of the time**	**A little of the time**	**None of the time**
1☐	2☐	3☐	4☐	5☐

10. How TRUE or FALSE is **each** of the following statements for you?

	Definitely true	**Mostly true**	**Don't know**	**Mostly false**	**Definitely false**
a. I seem to get sick a little easier than other people .....(GH2)	1☐	2☐	3☐	4☐	5☐
b. I am as healthy as anybody I know (GH3)	1☐	2☐	3☐	4☐	5☐
c. I expect my health to get worse.....(GH4)	1☐	2☐	3☐	4☐	5☐

d. My health is excellent.....(GH5) 1☐ 2☐ 3☐ 4☐ 5☐

Now we'd like to ask you some questions about how your health may have changed.

11. **Compared to one year ago,** how would you rate your physical health in general **now**?

Much better	Somewhat better	About the same	Somewhat worse	Much worse
1☐	2☐	3☐	4☐	5☐

12. **Compared to one year ago,** how would you rate your emotional problems (such as feeling anxious, depressed or irritable) **now**?

Much better	Somewhat better	About the same	Somewhat worse	Much worse
1☐	2☐	3☐	4☐	5☐

Appendix D:

D.1. Validation Studies of the Physical and Mental Summary Scores of the Veterans SF-36 Health Survey

Description of Validation Results of Veterans SF-36

Methods used to validate the Veterans SF-36 are the same as those described in the body of this report for the MOS SF-36. The differences with the two versions are in the role physical and emotional scales, where the Veterans SF-36 include 5 point ordinal choices from "no, none of the time to yes all of the time" compared to the MOS SF-36 which uses yes or no dichotomous choices for each of these items. The number of dummy variable estimates in the complete data model is 130 for the Veterans SF-36 while 109 for the MOS SF-36 version plus the intercept.

The results of the eight tables presented gives a very similar picture as the MOS SF-36 validation results in the body of this report. Table 1 describes the observations imputable which were about 95% using the half-scale rule and close to all of the cases were imputable for PCS and MCS using the MRE approach. Tables 2A, 2B, 2C and 2D describe a similar pattern of bias and error points for PCS and MCS estimates as in the MOS SF-36 version. The RE and MRE approaches were superior in limiting the bias as compared with the half scale rule as in the MOS SF-36 results. For tables 3A, 3B, 3C and 3D results were also comparable for the VISNs and demographics as compared with the MOS SF-36 results using bias and converted to error points in PCS and MCS. The only exception was in the health conditions where the bias and error points, respectively were somewhat larger for the Veterans SF-36. However, this was probably due to the differences in the samples. The means were quite variable to begin with and suggested greater standard deviation of PCS and MCS. Analysis of variance of the disease conditions gave a standard deviation that was larger than the MOS SF-36 administered in the HOS sample (PCS 7.2 SD for Veterans SF-36 vs. 6.5 for MOS SF-36; and MCS 8.6 SD for Veterans SF-36 vs. 6.6 for MOS SF-36). Tables 4A, 4B, 4C and 4D indicate that the PCS and MCS bias and error points, respectively due to naturally occurring data compared with the MRE as the standard is superior to the half scale rule for VISNs, conditions and demographics. We were unable to compare this to the MDE approach given the unavailability of the program to compute the Veterans SF-36 scores. However, the results would very likely be comparable to the findings described in the body of this report for the MOS SF-36.

To summarize and conclude, the Veterans SF-36 performed as well as the HOS SF-36 using the same imputation approaches. The MRE approach is strongly recommended for the Veterans SF-36 as the method for imputation.

(Note that the following tables reflect over 1200 regression models, to complete the validation studies for the Veterans SF-36.)

Table 1: Observations imputable.

Method	Number Scorable	% of possible cases
Casewise deletion	587,642	68.04
Half-scale	824,301	95.44
MRE (PCS, r^2>0.5)	863,565	99.99
MRE (MCS, r^2>0.5)	861,704	99.78

Table 2A: Bias properties of PCS estimates.*

	Means and Standard Deviations						Bias Properties(x1000)								
	Half-Scale		Regression		MRE		VISN			Disease Groups			Demographic Groups		
Missing	bias	sd	bias	sd	bias	sd	HS	RE	MRE	HS	RE	MRE	HS	RE	MRE
pf01	1.48	1.50	0.00	0.42	0.00	0.42	0.0261	0.0009	0.0010	0.8651	0.0046	0.0039	0.4492	0.0074	0.0072
pf02	0.92	1.38	-0.00	0.39	0.00	0.39	0.0214	0.0003	0.0003	0.7389	0.0006	0.0006	0.3071	0.0007	0.0006
pf04	1.12	1.43	0.00	0.39	0.00	0.39	0.0363	0.0010	0.0009	0.8616	0.0041	0.0039	0.3650	0.0005	0.0005
pf06	1.03	1.41	-0.00	0.47	-0.00	0.47	0.0293	0.0019	0.0016	0.9137	0.0279	0.0269	0.3132	0.0035	0.0033
pf09	0.55	1.28	-0.00	0.34	-0.00	0.34	0.0137	0.0003	0.0003	0.5676	0.0008	0.0008	0.2449	0.0003	0.0002
pf10	0.34	1.28	-0.00	0.40	-0.00	0.40	0.0070	0.0002	0.0002	0.4467	0.0039	0.0039	0.1639	0.0040	0.0038
5pfA	2.12	2.00	0.00	1.07	0.00	1.07	0.0637	0.0072	0.0055	1.5220	0.0552	0.0378	0.7092	0.0400	0.0369
5pfB	0.67	1.65	-0.00	0.98	0.00	0.98	0.0547	0.0117	0.0088	0.7858	0.0386	0.0319	0.2271	0.0094	0.0071
rp2	0.94	1.36	-0.00	0.50	0.00	0.50	0.0218	0.0003	0.0003	0.6421	0.0004	0.0006	0.3025	0.0016	0.0016
rp3	0.95	1.38	-0.00	0.48	-0.00	0.48	0.0224	0.0001	0.0001	0.7606	0.0007	0.0004	0.3023	0.0005	0.0005
2rp	1.06	1.49	0.00	0.79	0.00	0.79	0.0305	0.0008	0.0009	0.7234	0.0019	0.0017	0.3258	0.0035	0.0034
bp1	1.69	1.99	0.00	1.15	0.00	1.15	0.0561	0.0092	0.0076	1.3597	0.2333	0.2005	0.4208	0.0350	0.0409
bp2	1.32	1.66	-0.00	0.82	0.00	0.82	0.0366	0.0023	0.0015	0.6712	0.0280	0.0254	0.4077	0.0086	0.0087
gh1	1.03	1.41	-0.00	0.43	-0.00	0.43	0.0205	0.0014	0.0013	0.9137	0.0174	0.0162	0.4944	0.0190	0.0183
gh3	0.98	1.42	-0.00	0.51	-0.00	0.51	0.0324	0.0025	0.0022	0.7781	0.0019	0.0014	0.3059	0.0041	0.0042
2gh	0.52	1.50	-0.00	0.73	0.00	0.73	0.0118	0.0015	0.0013	0.4226	0.0421	0.0359	0.1417	0.0269	0.0281
vt2	0.92	1.30	0.00	0.10	0.00	0.10	0.0190	0.0000	0.0000	0.7008	0.0000	0.0000	0.3041	0.0000	0.0000
vt4	0.87	1.29	0.00	0.10	0.00	0.10	0.0180	0.0000	0.0000	0.6772	0.0000	0.0000	0.2871	0.0000	0.0000
2vt	0.91	1.30	0.00	0.12	0.00	0.12	0.0178	0.0000	0.0000	0.6914	0.0001	0.0001	0.2982	0.0001	0.0001
sf1	0.88	1.30	0.00	0.09	0.00	0.09	0.0182	0.0000	0.0000	0.6813	0.0000	0.0000	0.2886	0.0001	0.0001
sf2	0.89	1.30	0.00	0.09	0.00	0.09	0.0194	0.0000	0.0000	0.6938	0.0000	0.0000	0.2996	0.0000	0.0000
re1	0.93	1.33	0.00	0.42	0.00	0.42	0.0217	0.0000	0.0003	0.7549	0.0018	0.0017	0.3458	0.0007	0.0007
re2	0.78	1.33	-0.00	0.42	0.00	0.42	0.0127	0.0000	0.0002	0.6686	0.0005	0.0007	0.2780	0.0013	0.0014
mh1	1.01	1.40	0.00	0.51	0.00	0.51	0.0180	0.0000	0.0030	0.6738	0.0279	0.0294	0.2116	0.0078	0.0085
mh2	1.22	1.39	-0.00	0.37	-0.00	0.37	0.0238	0.0000	0.0001	0.7803	0.0025	0.0025	0.3138	0.0006	0.0006
mh3	0.45	1.38	-0.00	0.45	0.00	0.45	0.0152	0.0000	0.0007	0.5500	0.0040	0.0038	0.2714	0.0068	0.0065
mh4	1.07	1.39	0.00	0.39	0.00	0.39	0.0249	0.0000	0.0006	0.7650	0.0043	0.0043	0.3429	0.0042	0.0040
mh5	0.67	1.37	0.00	0.46	0.00	0.46	0.0255	0.0000	0.0007	0.7191	0.0123	0.0116	0.3817	0.0120	0.0113
2mhA	0.48	1.50	0.00	0.73	0.00	0.73	0.0115	0.0000	0.0051	0.4747	0.0462	0.0491	0.1716	0.0212	0.0218
2mhB	1.58	1.60	0.00	0.70	0.00	0.70	0.0401	0.0000	0.0035	0.9253	0.0375	0.0378	0.4030	0.0171	0.0161
2mhC	0.84	1.49	0.00	0.64	0.00	0.64	0.0352	0.0000	0.0025	0.8472	0.0393	0.0377	0.4962	0.0359	0.0336
list1	2.19	2.72	-0.00	1.99	0.00	2.00	0.0688	0.0000	0.0229	1.4423	0.3850	0.2380	0.9870	0.2369	0.1914
List2	0.60	2.53	0.00	2.18	0.00	2.19	0.0104	0.0000	0.0220	0.8694	0.6916	0.4731	0.1228	0.0922	0.1043

* The superior imputation method is shaded in the table. Note that the variable names for the Veterans SF-36 are explained in Appendix D.2, that gives the Veterans SF-36 questionnaire. The explanation for the missing item nomenclature is explained as follows: if an item is named in the above list, then only that item is missing. If the entry starts with a number, then that number of items is missing from that scale. List1 and List2 are two lists of 16 items that press the half-scale limits by deleting the largest number of responses from each scale possible before the half-scale cannot score it at all. All of the scenarios in the above are situations which the half-scale rule can score.

Table 2B: Bias properties of MCS estimates.*

	Means and Standard Deviations						MCS Bias Properties(x1000)								
	Half-Scale		Regression		MRE		VISNS			Disease Groups			Demographic Groups		
Missing	Bias	sd	bias	sd	bias	sd	HS	RE	MRE	HS	RE	MRE	HS	RE	MRE
pf01	0.32	2.47	0.00	0.22	0.00	0.22	0.1349	0.0002	0.0002	5.3924	0.0011	0.0011	0.9947	0.0017	0.0017
pf02	0.63	2.42	-0.00	0.21	-0.00	0.21	0.1003	0.0001	0.0001	4.8312	0.0001	0.0002	0.7973	0.0002	0.0002
pf04	0.52	2.43	0.00	0.21	0.00	0.21	0.0965	0.0002	0.0002	4.9103	0.0010	0.0010	0.8255	0.0001	0.0001
pf06	0.56	2.45	-0.00	0.25	-0.00	0.25	0.0981	0.0004	0.0005	5.1053	0.0064	0.0065	0.8669	0.0008	0.0008
pf09	0.83	2.46	-0.00	0.18	-0.00	0.18	0.1139	0.0001	0.0001	5.2734	0.0002	0.0002	0.8136	0.0001	0.0001
pf10	0.94	2.49	-0.00	0.21	-0.00	0.21	0.1457	0.0000	0.0000	5.6852	0.0009	0.0009	0.8760	0.0009	0.0009
5pfA	-0.02	2.54	-0.00	0.58	-0.00	0.58	0.0974	0.0017	0.0018	5.1487	0.0127	0.0134	1.1227	0.0092	0.0086
5pfB	0.76	2.48	-0.00	0.53	-0.00	0.53	0.0743	0.0027	0.0029	4.8622	0.0089	0.0091	0.7821	0.0022	0.0022
rp2	0.63	2.43	-0.00	0.18	-0.00	0.18	0.1116	0.0000	0.0000	5.1194	0.0000	0.0000	0.8519	0.0002	0.0001
rp3	0.62	2.44	-0.00	0.17	-0.00	0.17	0.1082	0.0000	0.0000	5.0795	0.0001	0.0001	0.8353	0.0000	0.0000
2rp	0.58	2.45	0.00	0.28	0.00	0.28	0.1106	0.0001	0.0001	5.1105	0.0002	0.0002	0.8757	0.0003	0.0003
bp1	0.40	2.56	0.00	0.35	0.00	0.35	0.1621	0.0007	0.0007	6.2373	0.0171	0.0174	1.0552	0.0026	0.0026
bp2	0.51	2.41	-0.00	0.25	-0.00	0.25	0.0980	0.0002	0.0002	4.8365	0.0021	0.0021	0.7724	0.0006	0.0006
gh1	0.64	2.43	-0.00	0.04	-0.00	0.04	0.1094	0.0000	0.0000	5.1053	0.0001	0.0001	0.8174	0.0001	0.0001
gh3	0.64	2.43	-0.00	0.04	-0.00	0.04	0.1075	0.0000	0.0000	5.0739	0.0000	0.0000	0.8262	0.0000	0.0000
2gh	0.67	2.43	-0.00	0.05	-0.00	0.05	0.1096	0.0000	0.0000	5.0185	0.0001	0.0001	0.8180	0.0001	0.0001
vt2	0.93	2.53	0.00	0.42	0.00	0.42	0.1044	0.0006	0.0005	5.3166	0.0006	0.0006	0.8879	0.0023	0.0024
vt4	0.49	2.45	0.00	0.44	0.00	0.44	0.0994	0.0006	0.0006	4.8129	0.0008	0.0008	0.7526	0.0013	0.0014
2vt	0.83	2.53	0.00	0.65	-0.00	0.65	0.0823	0.0003	0.0004	4.9805	0.0028	0.0030	0.7766	0.0064	0.0069
sf1	0.76	2.87	0.00	1.03	-0.00	1.03	0.1309	0.0036	0.0041	5.5762	0.0196	0.0181	0.8339	0.0309	0.0307
sf2	0.53	2.64	0.00	1.07	-0.00	1.07	0.1000	0.0061	0.0056	4.5952	0.0276	0.0171	0.8796	0.0293	0.0251
re1	0.55	2.49	0.00	0.93	-0.00	0.93	0.1138	0.0010	0.0010	5.2902	0.0068	0.0034	0.8668	0.0030	0.0020
re2	0.88	2.61	-0.00	0.92	-0.00	0.92	0.1391	0.0008	0.0009	5.6184	0.0019	0.0011	0.8741	0.0056	0.0059
mh1	0.36	2.68	0.00	1.12	-0.00	1.12	0.1514	0.0107	0.0099	5.3732	0.1070	0.0820	0.8967	0.0300	0.0307
mh2	-0.09	2.60	-0.00	0.80	-0.00	0.80	0.1341	0.0006	0.0007	5.4851	0.0091	0.0057	1.0395	0.0025	0.0021
mh3	1.60	2.76	-0.00	0.98	-0.00	0.98	0.1273	0.0029	0.0035	5.2018	0.0146	0.0123	0.6905	0.0252	0.0251
mh4	0.23	2.60	0.00	0.83	-0.00	0.83	0.1225	0.0024	0.0025	5.0214	0.0160	0.0108	0.9574	0.0163	0.0151
mh5	1.13	2.64	0.00	0.99	-0.00	0.99	0.0697	0.0033	0.0038	4.4483	0.0465	0.0372	0.7646	0.0457	0.0416
2mhA	1.54	3.06	0.00	1.60	-0.00	1.60	0.1820	0.0192	0.0190	5.5685	0.1773	0.1114	0.8126	0.0807	0.0824
2mhB	-0.89	3.05	0.00	1.52	-0.00	1.52	0.1870	0.0145	0.0156	5.5262	0.1434	0.0911	1.3516	0.0665	0.0569
2mhC	0.73	2.86	0.00	1.40	-0.00	1.40	0.0706	0.0112	0.0126	4.2154	0.1498	0.1102	0.9998	0.1380	0.1244
list1	0.99	3.47	-0.00	2.27	-0.00	2.28	0.1216	0.0142	0.0209	5.5545	0.2682	0.1619	1.1568	0.1700	0.1349
list2	1.23	3.52	0.00	2.41	-0.00	2.42	0.1842	0.0121	0.0162	5.8435	0.5093	0.3524	0.9540	0.1084	0.0950

* The superior imputation method is shaded in the table. Note that the variable names for the Veterans SF-36 are explained in Appendix D.2, that gives the Veterans SF-36 questionnaire. The explanation for the missing item nomenclature is explained as follows: if an item is named in the above list, then only that item is missing. If the entry starts with a number, then that number of items is missing from that scale. List1 and List2 are two lists of 16 items that press the half-scale limits by deleting the largest number of responses from each scale possible before the half-scale cannot score it at all. All of the scenarios in the above are situations which the half-scale rule can score

Table 2C: Error Points of PCS Estimates.*

	Means and Standard Deviations						PCS Error Points								
	Half-Scale		Regression		MRE		VISN			Disease Groups			Demographic Groups		
Missing	bias	sd	bias	sd	bias	sd	HS	RE	MRE	HS	RE	MRE	HS	RE	MRE
pf01	1.48	1.50	0.00	0.42	0.00	0.42	0.0628	0.0117	0.0123	0.3615	0.0264	0.0243	0.2605	0.0334	0.0330
pf02	0.92	1.38	-0.00	0.39	0.00	0.39	0.0569	0.0067	0.0067	0.3341	0.0095	0.0095	0.2154	0.0103	0.0095
pf04	1.12	1.43	0.00	0.39	0.00	0.39	0.0740	0.0123	0.0117	0.3607	0.0249	0.0243	0.2348	0.0087	0.0087
pf06	1.03	1.41	-0.00	0.47	-0.00	0.47	0.0665	0.0169	0.0155	0.3715	0.0649	0.0637	0.2175	0.0230	0.0223
pf09	0.55	1.28	-0.00	0.34	-0.00	0.34	0.0455	0.0067	0.0067	0.2928	0.0110	0.0110	0.1923	0.0067	0.0055
pf10	0.34	1.28	-0.00	0.40	-0.00	0.40	0.0325	0.0055	0.0055	0.2598	0.0243	0.0243	0.1573	0.0246	0.0240
5pfA	2.12	2.00	0.00	1.07	0.00	1.07	0.0981	0.0330	0.0288	0.4795	0.0913	0.0756	0.3273	0.0777	0.0747
5pfB	0.67	1.65	-0.00	0.98	0.00	0.98	0.0909	0.0420	0.0365	0.3445	0.0764	0.0694	0.1852	0.0377	0.0327
rp2	0.94	1.36	-0.00	0.50	0.00	0.50	0.0574	0.0067	0.0067	0.3114	0.0078	0.0095	0.2138	0.0155	0.0155
rp3	0.95	1.38	-0.00	0.48	-0.00	0.48	0.0582	0.0039	0.0039	0.3389	0.0103	0.0078	0.2137	0.0087	0.0087
2rp	1.06	1.49	0.00	0.79	0.00	0.79	0.0679	0.0110	0.0117	0.3306	0.0169	0.0160	0.2218	0.0230	0.0227
bp1	1.69	1.99	0.00	1.15	0.00	1.15	0.0921	0.0373	0.0339	0.4532	0.1877	0.1740	0.2521	0.0727	0.0786
bp2	1.32	1.66	-0.00	0.82	0.00	0.82	0.0744	0.0186	0.0151	0.3184	0.0650	0.0619	0.2482	0.0360	0.0363
gh1	1.03	1.41	-0.00	0.43	-0.00	0.43	0.0556	0.0145	0.0140	0.3715	0.0513	0.0495	0.2733	0.0536	0.0526
gh3	0.98	1.42	-0.00	0.51	-0.00	0.51	0.0700	0.0194	0.0182	0.3428	0.0169	0.0145	0.2150	0.0249	0.0252
2gh	0.52	1.50	-0.00	0.73	0.00	0.73	0.0422	0.0151	0.0140	0.2526	0.0797	0.0736	0.1463	0.0637	0.0651
vt2	0.92	1.30	0.00	0.10	0.00	0.10	0.0536	0.0000	0.0000	0.3253	0.0000	0.0000	0.2143	0.0000	0.0000
vt4	0.87	1.29	0.00	0.10	0.00	0.10	0.0521	0.0000	0.0000	0.3198	0.0000	0.0000	0.2082	0.0000	0.0000
2vt	0.91	1.30	0.00	0.12	0.00	0.12	0.0519	0.0000	0.0000	0.3232	0.0039	0.0039	0.2122	0.0039	0.0039
sf1	0.88	1.30	0.00	0.09	0.00	0.09	0.0524	0.0000	0.0000	0.3208	0.0000	0.0000	0.2088	0.0039	0.0039
sf2	0.89	1.30	0.00	0.09	0.00	0.09	0.0541	0.0000	0.0000	0.3237	0.0000	0.0000	0.2127	0.0000	0.0000
re1	0.93	1.33	0.00	0.42	0.00	0.42	0.0573	0.0000	0.0067	0.3377	0.0165	0.0160	0.2285	0.0103	0.0103
re2	0.78	1.33	-0.00	0.42	0.00	0.42	0.0438	0.0000	0.0055	0.3178	0.0087	0.0103	0.2049	0.0140	0.0145
mh1	1.01	1.40	0.00	0.51	0.00	0.51	0.0521	0.0000	0.0213	0.3190	0.0649	0.0666	0.1788	0.0343	0.0358
mh2	1.22	1.39	-0.00	0.37	-0.00	0.37	0.0600	0.0000	0.0039	0.3433	0.0194	0.0194	0.2177	0.0095	0.0095
mh3	0.45	1.38	-0.00	0.45	0.00	0.45	0.0479	0.0000	0.0103	0.2882	0.0246	0.0240	0.2025	0.0320	0.0313
mh4	1.07	1.39	0.00	0.39	0.00	0.39	0.0613	0.0000	0.0095	0.3399	0.0255	0.0255	0.2276	0.0252	0.0246
mh5	0.67	1.37	0.00	0.46	0.00	0.46	0.0621	0.0000	0.0103	0.3296	0.0431	0.0419	0.2401	0.0426	0.0413
2mhA	0.48	1.50	0.00	0.73	0.00	0.73	0.0417	0.0000	0.0278	0.2678	0.0835	0.0861	0.1610	0.0566	0.0574
2mhB	1.58	1.60	0.00	0.70	0.00	0.70	0.0778	0.0000	0.0230	0.3738	0.0753	0.0756	0.2467	0.0508	0.0493
2mhC	0.84	1.49	0.00	0.64	0.00	0.64	0.0729	0.0000	0.0194	0.3577	0.0770	0.0755	0.2738	0.0736	0.0712
list1	2.19	2.72	-0.00	1.99	0.00	2.00	0.1019	0.0000	0.0588	0.4667	0.2411	0.1896	0.3861	0.1892	0.1700
List2	0.60	2.53	0.00	2.18	0.00	2.19	0.0396	0.0000	0.0576	0.3624	0.3232	0.2673	0.1362	0.1180	0.1255

* The superior imputation method is shaded in the table. Note that the variable names for the Veterans SF-36 are explained in Appendix D.2, that gives the Veterans SF-36 questionnaire. The explanation for the missing item nomenclature is explained as follows: if an item is named in the above list, then only that item is missing. If the entry starts with a number, then that number of items is missing from that scale. List1 and List2 are two lists of 16 items that press the half-scale limits by deleting the largest number of responses from each scale possible before the half-scale cannot score it at all. All of the scenarios in the above are situations which the half-scale rule can score

Table 2D: Error Points of MCS Estimates.*

	Means and Standard Deviations						MCS Error Points								
	Half-Scale		Regression		MRE		VISNS			Disease Groups			Demographic Groups		
Missing	bias	sd	bias	sd	bias	sd	HS	RE	MRE	HS	RE	MRE	HS	RE	MRE
pf01	0.32	2.47	0.00	0.22	0.00	0.22	0.1609	0.0062	0.0062	1.0170	0.0145	0.0145	0.4368	0.0181	0.0181
pf02	0.63	2.42	-0.00	0.21	-0.00	0.21	0.1387	0.0044	0.0044	0.9627	0.0044	0.0062	0.3911	0.0062	0.0062
pf04	0.52	2.43	0.00	0.21	0.00	0.21	0.1361	0.0062	0.0062	0.9705	0.0139	0.0139	0.3979	0.0044	0.0044
pf06	0.56	2.45	-0.00	0.25	-0.00	0.25	0.1372	0.0088	0.0098	0.9896	0.0350	0.0353	0.4078	0.0124	0.0124
pf09	0.83	2.46	-0.00	0.18	-0.00	0.18	0.1478	0.0044	0.0044	1.0058	0.0062	0.0062	0.3951	0.0044	0.0044
pf10	0.94	2.49	-0.00	0.21	-0.00	0.21	0.1672	0.0000	0.0000	1.0443	0.0131	0.0131	0.4099	0.0131	0.0131
5pfA	-0.02	2.54	-0.00	0.58	-0.00	0.58	0.1367	0.0181	0.0186	0.9938	0.0494	0.0507	0.4641	0.0420	0.0406
5pfB	0.76	2.48	-0.00	0.53	-0.00	0.53	0.1194	0.0228	0.0236	0.9658	0.0413	0.0418	0.3873	0.0205	0.0205
rp2	0.63	2.43	-0.00	0.18	-0.00	0.18	0.1463	0.0000	0.0000	0.9910	0.0000	0.0000	0.4042	0.0062	0.0044
rp3	0.62	2.44	-0.00	0.17	-0.00	0.17	0.1441	0.0000	0.0000	0.9871	0.0044	0.0044	0.4003	0.0000	0.0000
2rp	0.58	2.45	0.00	0.28	0.00	0.28	0.1457	0.0044	0.0044	0.9901	0.0062	0.0062	0.4099	0.0076	0.0076
bp1	0.40	2.56	0.00	0.35	0.00	0.35	0.1763	0.0116	0.0116	1.0938	0.0573	0.0578	0.4499	0.0223	0.0223
bp2	0.51	2.41	-0.00	0.25	-0.00	0.25	0.1371	0.0062	0.0062	0.9632	0.0201	0.0201	0.3849	0.0107	0.0107
gh1	0.64	2.43	-0.00	0.04	-0.00	0.04	0.1449	0.0000	0.0000	0.9896	0.0044	0.0044	0.3960	0.0044	0.0044
gh3	0.64	2.43	-0.00	0.04	-0.00	0.04	0.1436	0.0000	0.0000	0.9866	0.0000	0.0000	0.3981	0.0000	0.0000
2gh	0.67	2.43	-0.00	0.05	-0.00	0.05	0.1450	0.0000	0.0000	0.9812	0.0044	0.0044	0.3961	0.0044	0.0044
vt2	0.93	2.53	0.00	0.42	0.00	0.42	0.1415	0.0107	0.0098	1.0099	0.0107	0.0107	0.4127	0.0210	0.0215
vt4	0.49	2.45	0.00	0.44	0.00	0.44	0.1381	0.0107	0.0107	0.9608	0.0124	0.0124	0.3800	0.0158	0.0164
2vt	0.83	2.53	0.00	0.65	-0.00	0.65	0.1256	0.0076	0.0088	0.9774	0.0232	0.0240	0.3860	0.0350	0.0364
sf1	0.76	2.87	0.00	1.03	-0.00	1.03	0.1585	0.0263	0.0280	1.0342	0.0613	0.0589	0.4000	0.0770	0.0767
sf2	0.53	2.64	0.00	1.07	-0.00	1.07	0.1385	0.0342	0.0328	0.9389	0.0728	0.0573	0.4108	0.0750	0.0694
re1	0.55	2.49	0.00	0.93	-0.00	0.93	0.1477	0.0139	0.0139	1.0074	0.0361	0.0255	0.4078	0.0240	0.0196
re2	0.88	2.61	-0.00	0.92	-0.00	0.92	0.1633	0.0124	0.0131	1.0381	0.0191	0.0145	0.4095	0.0328	0.0336
mh1	0.36	2.68	0.00	1.12	-0.00	1.12	0.1704	0.0453	0.0436	1.0152	0.1433	0.1254	0.4147	0.0759	0.0767
mh2	-0.09	2.60	-0.00	0.80	-0.00	0.80	0.1604	0.0107	0.0116	1.0258	0.0418	0.0331	0.4465	0.0219	0.0201
mh3	1.60	2.76	-0.00	0.98	-0.00	0.98	0.1563	0.0236	0.0259	0.9989	0.0529	0.0486	0.3639	0.0695	0.0694
mh4	0.23	2.60	0.00	0.83	-0.00	0.83	0.1533	0.0215	0.0219	0.9814	0.0554	0.0455	0.4285	0.0559	0.0538
mh5	1.13	2.64	0.00	0.99	-0.00	0.99	0.1156	0.0252	0.0270	0.9237	0.0944	0.0845	0.3830	0.0936	0.0893
2mhA	1.54	3.06	0.00	1.60	-0.00	1.60	0.1868	0.0607	0.0604	1.0335	0.1844	0.1462	0.3948	0.1244	0.1257
2mhB	-0.89	3.05	0.00	1.52	-0.00	1.52	0.1894	0.0527	0.0547	1.0296	0.1659	0.1322	0.5092	0.1129	0.1045
2mhC	0.73	2.86	0.00	1.40	-0.00	1.40	0.1164	0.0464	0.0492	0.8992	0.1695	0.1454	0.4379	0.1627	0.1545
list1	0.99	3.47	-0.00	2.27	-0.00	2.28	0.1527	0.0522	0.0633	1.0322	0.2268	0.1762	0.4711	0.1806	0.1609
list2	1.23	3.52	0.00	2.41	-0.00	2.42	0.1880	0.0482	0.0557	1.0587	0.3126	0.2600	0.4278	0.1442	0.1350

* The superior imputation method is shaded in the table. Note that the variable names for the Veterans SF-36 are explained in Appendix D.2, that gives the Veterans SF-36 questionnaire. The explanation for the missing item nomenclature is explained as follows: if an item is named in the above list, then only that item is missing. If the entry starts with a number, then that number of items is missing from that scale. List1 and List2 are two lists of 16 items that press the half-scale limits by deleting the largest number of responses from each scale possible before the half-scale cannot score it at all. All of the scenarios in the above are situations which the half-scale rule can score

Table 3A: PCS bias properties (h/v) of Regression Estimates (x1000).*

		VISNS		Health Conditions		Demographics	
Values	**R-sq**	**RE**	**MRE**	**RE**	**MRE**	**RE**	**MRE**
list1	97.4	0.0430	0.0229	0.3850	0.2380	0.2369	0.1914
list2	96.8	0.0442	0.0220	0.6916	0.4731	0.0922	0.1043
NoPF*	92.4	0.0750	0.0653	1.6538	0.8499	1.3510	0.9122
OnlyPF7	96.3	0.0317	0.0184	0.4818	0.2682	0.3102	0.2151
OnlyPF10	94.2	0.0524	0.0392	0.8960	0.3794	0.7206	0.4826
NoRERP*	95.8	0.0252	0.0368	0.2608	0.3168	0.2080	0.1765
SF18	94.4	0.0983	0.0588	1.2707	0.5468	0.1619	0.0983
SF12	93.8	0.1215	0.0623	1.7225	0.9292	0.1720	0.0890
SF8	91.0	0.1529	0.0763	2.9646	1.5327	0.9706	0.5755
SF6a*	85.8	0.4766	0.1514	7.2899	3.4836	2.3308	1.3022
SF6b*	86.9	0.1905	0.0812	4.1871	2.4703	1.8468	1.1833
SF3a*	78.0	1.2443	0.4265	16.8720	7.8409	8.6421	4.7350
SF3b*	73.2	1.4315	0.3347	38.2210	22.3220	12.4230	7.0427
SF1*	51.9	7.3217	1.3029	110.1200	38.3420	22.6710	10.5930

* See 1st paragraph page 19 for description of variables in this table.

Table 3B: MCS bias properties (h/v) of Regression Estimates (x1000).*

		VISNS		Health Conditions		Demographics	
Values	**R-sq**	**RE**	**MRE**	**RE**	**MRE**	**RE**	**MRE**
list1	97.3	0.0142	0.0209	0.2682	0.1619	0.1700	0.1349
list2	97.0	0.0121	0.0162	0.5093	0.3524	0.1084	0.0950
NoPF	98.2	0.0163	0.0192	0.3601	0.3826	0.2942	0.2798
OnlyPF7	99.1	0.0071	0.0090	0.1083	0.1127	0.0697	0.0668
OnlyPF10	98.7	0.0116	0.0137	0.1979	0.2169	0.1593	0.1483
NoRERP*	92.3	0.0275	0.0732	0.6250	0.2951	0.8508	0.8964
SF18	98.3	0.0155	0.0189	0.2786	0.2689	0.1292	0.1211
SF12	95.2	0.0191	0.0111	1.5226	0.9417	0.5541	0.4587
SF8	91.4	0.0529	0.0716	3.4403	2.0171	1.6493	1.3201
SF6a*	79.4	0.1876	0.4814	11.2830	4.6998	2.5209	0.9871
SF6b*	76.0	0.1463	0.3638	18.2640	6.1967	3.5233	1.3223
SF3a*	68.8	0.5385	0.3093	29.1140	6.2885	4.4823	1.2355
SF3b*	68.0	0.4632	0.4344	34.4130	8.6081	3.9673	0.7940
SF1*	26.0	6.4660	4.2440	682.1100	154.1600	138.7700	52.0790

* See 1st paragraph page 19 for description of variables in this table.

Table 3C: PCS Error Points of Regression Estimates.*

		VISNS		Health Conditions		Demographics	
Values	**R-sq**	**RE**	**MRE**	**RE**	**MRE**	**RE**	**MRE**
list1	97.4	0.0806	0.0588	0.2411	0.1896	0.1892	0.1700
list2	96.8	0.0817	0.0576	0.3232	0.2673	0.1180	0.1255
NoPF*	92.4	0.1064	0.0993	0.4998	0.3583	0.4517	0.3712
OnlyPF7	96.3	0.0692	0.0527	0.2698	0.2013	0.2165	0.1802
OnlyPF10	94.2	0.0890	0.0769	0.3679	0.2394	0.3299	0.2700
NoRERP*	95.8	0.0617	0.0746	0.1985	0.2187	0.1772	0.1633
SF18	94.4	0.1219	0.0942	0.4381	0.2874	0.1564	0.1219
SF12	93.8	0.1355	0.0970	0.5101	0.3746	0.1612	0.1159
SF8	91.0	0.1520	0.1074	0.6692	0.4811	0.3829	0.2948
SF6a*	85.8	0.2683	0.1512	1.0493	0.7254	0.5933	0.4435
SF6b*	86.9	0.1696	0.1107	0.7953	0.6108	0.5282	0.4228
SF3a*	78.0	0.4335	0.2538	1.5964	1.0883	1.1425	0.8457
SF3b*	73.2	0.4650	0.2248	2.4027	1.8362	1.3698	1.0314
SF1*	51.9	1.0516	0.4436	4.0784	2.4065	1.8505	1.2649

* See 1st paragraph page 19 for description of variables in this table.

Table 3D: MCS Error Points of Regression Estimates.*

		VISNS		Health Conditions		Demographics	
Values	**R-sq**	**RE**	**MRE**	**RE**	**MRE**	**RE**	**MRE**
list1	97.3	0.0522	0.0633	0.2268	0.1762	0.1806	0.1609
list2	97.0	0.0482	0.0557	0.3126	0.2600	0.1442	0.1350
NoPF	98.2	0.0559	0.0607	0.2628	0.2709	0.2376	0.2317
OnlyPF7	99.1	0.0369	0.0416	0.1441	0.1470	0.1156	0.1132
OnlyPF10	98.7	0.0472	0.0513	0.1948	0.2040	0.1748	0.1687
NoRERP*	92.3	0.0726	0.1185	0.3463	0.2379	0.4040	0.4147
SF18	98.3	0.0545	0.0602	0.2312	0.2271	0.1574	0.1524
SF12	95.2	0.0605	0.0461	0.5404	0.4250	0.3260	0.2966
SF8	91.4	0.1007	0.1172	0.8124	0.6220	0.5625	0.5032
SF6a*	79.4	0.1897	0.3039	1.4712	0.9495	0.6954	0.4351
SF6b*	76.0	0.1675	0.2642	1.8717	1.0903	0.8221	0.5036
SF3a*	68.8	0.3214	0.2436	2.3632	1.0983	0.9273	0.4868
SF3b*	68.0	0.2981	0.2887	2.5693	1.2850	0.8724	0.3903
SF1*	26.0	1.1137	0.9023	11.4387	5.4380	5.1594	3.1607

* See 1st paragraph page 19 for description of variables in this table.

Table 4A : PCS Bias due to naturally missing data compared with the MRE approach as the standard

		PCS Bias (h/v) x 1000		
Imputation Algorithm	**MeanBias VISNS**		**Conditions**	**Demographics**
Complete case	-3.17 0.1526		1.4263	2.0598
Half-Scale rule	-4.28 0.0946		0.4201	1.5154

Table 4B: MCS Bias due to naturally missing data compared with the MRE approach as the standard

		PCS Bias (h/v) x 1000		
Imputation Algorithm	**MeanBias VISNS**		**Conditions**	**Demographics**
Complete case	-2.55 0.1350		0.1253	0.3098
Half-Scale rule	-2.93 0.0884		0.1175	0.6197

Table 4C : PCS Error Points due to naturally missing data compared with the MRE approach as the standard

		PCS Error Points		
Imputation Algorithm	**MeanBias VISNS**		**Conditions**	**Demographics**
Complete case	-3.17 0.1518		0.4641	0.5578
Half-Scale rule	-4.28 0.1195		0.2519	0.4784

Table 4D: MCS Error Points due to naturally missing data compared with the MRE approach as the standard

		MCS Error Points		
Imputation Algorithm	**MeanBias VISNS**		**Conditions**	**Demographics**
Complete case	-2.55 0.1609		0.1550	0.2438
Half-Scale rule	-2.93 0.1302		0.1501	0.3448

Appendix D2. Description of Variables and Questions
Veterans SF-36 Health Survey

1. In general, would you say your health is: (GH1)

Excellent Very	good	Good	Fair	Poor
1☐	2☐	3☐	4☐	5☐

2. The following items are about activities you might do during a typical day. Does **your health now limit you** in these activities? If so, how much?

ACTIVITIES	Yes, limited a lot	Yes, limited a little	No, not limited at all
a. **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports.....(PF1)	1☐	2☐	3☐
b. **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.....(PF2)	1☐	2☐	3☐
c. Lifting or carrying groceries.....(PF3)	1☐	2☐	3☐
d. Climbing **several** flights of stairs.....(PF4)	1☐	2☐	3☐
e. Climbing **one** flight of stairs.....(PF5)	1☐	2☐	3☐
f. Bending, kneeling, or stooping.....(PF6)	1☐	2☐	3☐
g. Walking **more than a mile**.....(PF7)	1☐	2☐	3☐
h. Walking **several blocks**.....(PF8)	1☐	2☐	3☐
i. Walking **one block**.....(PF9)	1☐	2☐	3☐
j. Bathing or dressing yourself.....(PF10)	1☐	2☐	3☐

3. During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities **as a result of your physical health**?

	Yes	No
a. Cut down on the **amount of time** you spent on work or other activities.....(VRP1)	1☐	2☐
b. **Accomplished less** than you would like .....(VRP2)	1☐	2☐
c. Were limited in the **kind** of work or other activities (VRP3)	1☐	2☐

d. Had **difficulty** performing the work or other activities (*for example*, it took extra effort) .....(VRP4) ☐1 ☐2

4. During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities **as a result of any emotional problems** (such as feeling depressed or anxious)?

		Yes	No
a.	Cut down on the **amount of time** you spent on work or other activities.....(VRE1)	☐1	☐2
b.	**Accomplished less** than you would like .....(VRE2)	☐1	☐2
c.	Didn't do work or other activities as **carefully** as usual.....(VRE3)	☐1	☐2

5. During the **past 4 weeks**, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups? (SF1)

Not at all	Slightly	Moderately	Quite a bit	Extremely
☐1	☐2	☐3	☐4	☐5

6. How much **bodily** pain have you had during the **past 4 weeks**? (BP1)

None	Very mild	Mild	Moderate	Severe	Very severe
☐1	☐2	☐3	☐4	☐5	☐6

7. During the **past 4 weeks**, how much did **pain** interfere with your normal work (including both work outside the home and housework)? (BP2)

Not at all	A little bit	Moderately	Quite a bit	Extremely
☐1	☐2	☐3	☐4	☐5

8. These questions are about how you feel and how things have been with you during the **past 4 weeks**. For each question, please give the one answer that comes closest to the way you have been feeling.

How much of the time during the **past 4 weeks**...	**All of the time**	**Most of the time**	**A good bit of the time**	**Some of the time**	**A little of the time**	**None of the time**
a. did you feel full of pep? .....(VT1)	1☐	2☐	3☐	4☐	5☐	6☐
b. have you been a very nervous person? .....(MH1)	1☐	2☐	3☐	4☐	5☐	6☐
c. have you felt so down in the dumps that nothing could cheer you up? .....(MH2)	1☐	2☐	3☐	4☐	5☐	6☐
d. have you felt calm and peaceful?....	1☐	2☐	3☐	4☐	5☐	6☐
e. did you have a lot of energy? (MH3) (VT2)	1☐	2☐	3☐	4☐	5☐	6☐
f. have you felt downhearted and blue? .....(MH4)	1☐	2☐	3☐	4☐	5☐	6☐
g. did you feel worn out? .....(VT3)	1☐	2☐	3☐	4☐	5☐	6☐
h. have you been a happy person?	1☐	2☐	3☐	4☐	5☐	6☐
i. did you feel tired? .....(MH5) (VT4)	1☐	2☐	3☐	4☐	5☐	6☐

9. During the **past 4 weeks**, how much of the time has your **physical health or emotional problems** interfered with your social activities (like visiting with friends, relatives, etc.)? (SF2)

All of the time	**Most of the time**	**Some of the time**	**A little of the time**	**None of the time**
1☐	2☐	3☐	4☐	5☐

10. How TRUE or FALSE is **each** of the following statements for you?

	Definitely true	**Mostly true**	**Don't know**	**Mostly false**	**Definitely false**
a. I seem to get sick a little easier than other people .....(GH2)	1☐	2☐	3☐	4☐	5☐
b. I am as healthy as anybody I know (GH3)	1☐	2☐	3☐	4☐	5☐
c. I expect my health to get worse.....(GH4)	1☐	2☐	3☐	4☐	5☐

d. My health is excellent.....(GH5) 1☐ 2☐ 3☐ 4☐ 5☐

Now we'd like to ask you some questions about how your health may have changed.

11. **Compared to one year ago,** how would you rate your physical health in general **now**?

Much better	Somewhat better	About the same	Somewhat worse	Much worse
1☐	2☐	3☐	4☐	5☐

12. **Compared to one year ago,** how would you rate your emotional problems (such as feeling anxious, depressed or irritable) **now**?

Much better	Somewhat better	About the same	Somewhat worse	Much worse
1☐	2☐	3☐	4☐	5☐

References

1. Ware JE, Kosinski M, Keller SD. *SF-36 Physical and Mental Health Summary Scales, A User's Manual.* The Health Institute, New England Medical Center. Boston, MA, 1994.

2. Ware JE, Kosinski M: *SF-36 Physical & Mental Health Summary Scores: A Manual for Users of Version 1. 2nd Edition.* QualityMetric Incorporated, Lincoln RI. 2001.
3. Ware JE, Kosinski M, Dewey JE. How to Score Version 2 of the SF-36 Health Survey. Medical Outcomes Trust and QualityMetric Incorporated 2002.

4. Ware JE. Modifications to the Role Functioning Scales of the Veterans SF-36 (Private Communication 1996)

5. Kazis LE. The veterans SF-36 health status questionnaire: Development and application in the Veterans Health Administration. Medical Outcomes Trust Monitor 2000; 5 (1).

6. Kazis L., Ren XS, Lee A, et al. Health status in VA patients: Results from the Veterans Health Study. American Journal of Medical Quality 1999; 14(1):28-38.

7. Ware JE, Snow KK, Kosinski M, Gandek B: *SF-36 Health Status Survey Manual and Interpretation Guide.* Boston. The Health Institute, New England Medical Center. Boston, MA, 1993. Available from the Medical Outcomes Trust.

8. Ware JE. Scales for measuring general health perceptions. Health Services Research 1976; 11, 396-415.

9. Ware JE, Kosinski M, Keller SD: *How to Score the SF-12 Physical & Mental health Summary Scales.* The Health Institute, New England Medical Center. Boston, MA, 1995.

10. Little JA, Rubin, DB: *Statistical Analysis with Missing Data.* 278 pp. Wiley, New York, 1987.

11. Embretson SE and Reise SP. Item Response Theory for Psychologists. Erlbaum Associates Publishers, London, 2000.