

# UCLA

## UCLA Previously Published Works

### Title

Imputation of SF-12 Health Scores for Respondents with Partially Missing Data

### Permalink

<https://escholarship.org/uc/item/2v5559n2>

### Journal

Health Services Research, 40(3)

### ISSN

0017-9124

### Authors

Liu, Honghu  
Hays, Ron D  
Adams, John L  
[et al.](#)

### Publication Date

2005-06-01

### DOI

10.1111/j.1475-6773.2005.00391.x

Peer reviewed

# Imputation of SF-12 Health Scores for Respondents with Partially Missing Data

*Honghu Liu, Ron D. Hays, John L. Adams, Wen-Pin Chen, Diana Tisnado, Carol M. Mangione, Cheryl L. Damberg, and Katherine L. Kahn*

---

**Objective.** To create an efficient imputation algorithm for imputing the SF-12 physical component summary (PCS) and mental component summary (MCS) scores when patients have one to eleven SF-12 items missing.

**Study Setting.** Primary data collection was performed between 1996 and 1998.

**Study Design.** Multi-pattern regression was conducted to impute the scores using only available SF-12 items (simple model), and then supplemented by demographics, smoking status and comorbidity (enhanced model) to increase the accuracy. A cut point of missing SF-12 items was determined for using the simple or the enhanced model. The algorithm was validated through simulation.

**Data Collection.** Thirty-thousand-three-hundred and eight patients from 63 physician groups were surveyed for a quality of care study in 1996, which collected the SF-12 and other information. The patients were classified as “chronic” patients if they reported that they had diabetes, heart disease, asthma/chronic obstructive pulmonary disease, or low back pain. A follow-up survey was conducted in 1998.

**Principal Findings.** Thirty-one percent of the patients missed at least one SF-12 item. Means of variance of prediction and standard errors of the mean imputed scores increased with the number of missing SF-12 items. Correlations between the observed and the imputed scores derived from the enhanced models were consistently higher than those derived from the simple model and the increments were significant for patients with  $\geq 6$  missing SF-12 items ( $p < .03$ ).

**Conclusion.** Missing SF-12 items are prevalent and lead to reduced analytical power. Regression-based multi-pattern imputation using the available SF-12 items is efficient and can produce good estimates of the scores. The enhancement from the additional patient information can significantly improve the accuracy of the imputed scores for patients with  $\geq 6$  items missing, leading to estimated scores that are as accurate as that of patients with  $< 6$  missing items.

**Key Words.** Health related quality of life, SF-12 health survey, imputation, validation

---

The SF-12 health survey encompasses all eight SF-36 scales (physical function, role-physical, bodily pain, general health, vitality, social functioning, role-emotional and mental health) and can be administered in 2–3 minutes (Ware

et al. 1993; Ware, Kosinski, and Keller 1995). Four of the eight health concepts are measured using two items each and the other four concepts are measured using one item each. Previous studies have found that the SF-12 can reproduce more than 90 percent of the variance in both the physical component summary (PCS) and mental component summary (MCS) scores of the SF-36 (Ware, Kosinski, and Keller 1995; Ware et al. 2002).

Item missing data is not uncommon in self-report surveys. The standard SF-12 scoring algorithm, however, is a weighted sum of the SF-12 items and requires answers to all 12 individual items in order to calculate PCS and MCS summary scores. However, missing items are likely to occur in any study that collects data by self-administration. Studies have shown that among patients with missing SF-12 items, approximately 50 percent have missing data on only one of the 12 items (Ware et al. 2002). Yet, by using the standard SF-12 scoring algorithm, even one missing item will yield missing data on the SF-12 PCS and MCS. The loss of observations because of missing SF-12 items can reduce analytical power, increase variation in parameter estimates, and potentially lead to bias, particularly when the SF-12 is the primary outcome of interest in a study. Therefore, an algorithm that can efficiently and accurately estimate PCS and MCS scores with partially missing SF-12 data would be valuable to researchers.

Since the SF-12 PCS and MCS scores are directly weighted functions of SF-12 items, the amount of missing data will have a significant impact on the imputed value. Hence, the imputation algorithm may need to apply different models according to the degree of "missingness." As the number of missing SF-12 items increases, the available information and predicting power from SF-12 items diminishes. When the number of missing SF-12 items reaches a certain level, the available SF-12 items may no longer carry sufficient information to produce accurate parameter estimates of SF-12 PCS and MCS scores. In these situations, additional patient information may help in estimation.

We applied a multi-pattern regression-based algorithm for imputing SF-12 PCS and MCS scores. This algorithm fits different imputation models

---

Address correspondence to Honghu Liu, Ph.D., Department of Medicine, Division of General Internal Medicine and Health Services Research, University of California, Los Angeles, 911 Broxton Avenue, Los Angeles, CA 90024. Dr. Liu, Ron D. Hays, Ph.D., Wen-Pin Chen, M.S., Diana Tisnado, Ph.D., Carol M. Mangione, M.D., M.P.H. and Katherine L. Kahn, M.D., are with the Department of Medicine, Division of General Internal Medicine and Health Services Research, University of California, Los Angeles, CA. Ron D. Hays, Ph.D., John L. Adams, Ph.D., Carol M. Mangione, M.D., M.P.H., Cheryl L. Damberg, Ph.D., Katherine L. Kahn, M.D., are with the RAND Corporation, Santa Monica, CA. Cheryl L. Damberg, Ph.D., is with the PBGH, San Francisco, CA.

according to the amount of SF-12 information missing. These imputation models are either a function of the available SF-12 items alone, or are a function of the available SF-12 items plus additional patient characteristics. Through simulation analyses using baseline and follow-up data 2 years later, we have evaluated and validated the imputation algorithm.

## METHODS

### *Patient Sample and Data Collection*

The analysis uses data collected in 1996 by the Pacific Business Group on Health (PBGH), a large nonprofit business coalition of 47 private and public sector purchasers of health care. The Physician Value Check Survey (PVCS) Project surveyed patients who were 18–70-years-old and received care from 63 managed care physician groups (medical groups and Independent Practice Organizations in California, Oregon and Washington). The self-administered mail survey is a full length survey with a total of 131 items, including the SF-12, and information about patient demographics, smoking habit, comorbidity, satisfaction with care, and receipt of preventive care services. As an incentive to complete the survey, respondents were eligible to receive one of 50 \$100 cash prizes (Kahn et al. 2003). PBGH surveyed the same cohort of patients 2 years later in 1998 with the goal of assessing changes in satisfaction, process and outcomes of care between different physician groups (Damberg and Bloomfield 1997).

The 1996 baseline study population was defined by patients between the ages of 18 and 70 enrolled in any of the 63 physician groups, who had at least one physician encounter during calendar year 1995. The size of the 1996 study population (i.e., the sampling frame) was 1,170,242 patients (Damberg and Bloomfield 1997). From each physician group, 1,000 patients were randomly selected, with over-sampling of individuals between ages 50 and 70 years to increase the likelihood of detecting 2-year changes in the SF-12. Of the 1,000 enrollees associated with each medical organization, 700 were drawn from 50- to 70-year-old group and 300 from 18- to 49-year-old group. The study assumed a 50 percent response rate at baseline and approximately 20 percent attrition in enrollment over time (Damberg and Bloomfield 1997; Kahn et al. 2003). Patients who indicated in the 1996 survey that they had at least one of four conditions (diabetes, ischemic heart disease, asthma or chronic obstructive pulmonary disease, or low back pain) were classified as “chronic” ( $n = 9,980$ ); other patients were considered “nonchronic” ( $n = 20,324$ ). In 1998, a subset of the patients who responded to 1996

baseline survey were invited for the follow-up survey and 1,661 chronic and 6,985 nonchronic patients responded (with about 60 percent response rate). A detailed description of the study can be found elsewhere (Kahn et al. 2003).

#### *Algorithm and Statistical Analyses*

To understand the characteristics of the baseline patient sample, the distribution of demographics and SF-12 PCS and MCS scores as well as the distribution of number of missing SF-12 items were calculated and evaluated for chronic and nonchronic patients. A multiple-step process was implemented to establish and validate the imputation algorithm. In step one, multi-pattern (different permutation patterns of missing SF-12 items) regressions were performed to impute SF-12 PCS and MCS scores (Little and Rubin 1987; Goldstein 1996; StataCorp 1999). Multi-pattern imputation is a regression-based imputation approach that fits one regression imputation model for each missing pattern. The *STATA* "impute" procedure was used to implement the algorithm. An imputed value was calculated as the predicted value from the imputation model. These imputation regression models were fitted first using a simple model of only SF-12 items only (simple model), and then using SF-12 items plus patient demographics (age, gender, race/ethnicity, education, income), current smoking status and medical comorbidities (enhanced model) to increase the predicting power of the imputation model and to improve the accuracy of the imputed values.

After imputation, means, medians, quantiles (5, 25, 75 and 95 percent) and variations of the imputed scores were evaluated. The underlying uncertainty because of missing SF-12 items was examined by calculating two measures: the means of variance of prediction from imputation models and the standard errors of the mean imputed values by number of missing SF-12 items in the sample (variance prediction measures the variation of a predicted value from the imputation models). The means of variance of prediction were calculated by averaging the patient level variance of prediction within each number of missing SF-12 items. Because the mean of imputed PCS or MCS scores for a given number of missing SF-12 items was calculated from different imputation models, the close-form solution (i.e., the exact formula for computation) of the standard error of the mean imputed score is practically unobtainable.

We estimated the standard errors of mean imputed SF-12 PCS and MCS scores using bootstrapping methods with 1,000 replicates of samples drawn

(Efron 1979; Efron and Tibshirani 1986). The confidence intervals of means of the imputed scores were then constructed using the bootstrapped standard errors by the number of missing SF-12 items. Based on variations of the mean variance of prediction and the confidence bands formed by confidence intervals of the mean imputed SF-12 values, a cut point of number of missing SF-12 items was determined for using the simple versus the enhanced model to impute SF-12 PCS and MCS scores.

In step two, the 69 percent of patients who had no missing data on SF-12 items were identified and their observed SF-12 PCS and MCS scores were calculated. These patients were then randomly assigned to 11 equal subgroups corresponding to the number of missing SF-12 items from one to 11. Permutations of missing patterns (missing one to 11 out of the total 12 items constitutes a total of 4,094 possible missing patterns) on SF-12 items were generated by simulation for these patients to create the distribution of “missingness.” Multi-pattern imputation was then employed to impute the PCS and MCS scores using observed SF-12 items only for simple models and then using observed SF-12 items plus patient demographics, smoking status and medical comorbidity for enhanced models. The imputed PCS and MCS scores were then compared with the observed PCS and MCS scores for these patients by the identified cut point of number of missing SF-12 items for using simple or enhanced imputation models. Correlations between the observed and the imputed PCS and MCS scores were calculated and the differences between correlations were evaluated by the cut point and tested using the asymptotic distribution of the difference of correlation with bootstrapped standard errors of the differences.

Multiple imputation (MI) (Rubin 1987, 1996) can fill in missing data with a set of plausible values that represent the uncertainty about the correct value to impute. Although the SF-12 items and the other measures used in the enhanced models do not meet the assumption of multivariate joint normal distribution, which are generally used for the computation in MI, we also conducted MI in the validation step using *SAS* MI procedure with the Markov Chain Monte Carlo method and five imputed values for each missing value. Point and variance estimates of the imputed PCS or MCS were calculated for each of the five imputed values from MI. The final point and variance estimates were obtained by averaging the five estimates and then comparing it to that from the multi-pattern regression imputation.

This simulation-based validation analysis was repeated again using 1998 follow-up data to confirm the validation performed using 1996 baseline data.

## RESULTS

Of the 30,308 respondents who completed the 1996 survey, 9,431 (31 percent) had missing SF-12 items. The mean ages (SD) were 53 (13) and 54 (12) for respondents with and without missing SF-12 items, respectively. A higher percentage of respondents with missing SF-12 items were female (67 percent) compared with percentage among respondents without missing SF-12 items (60 percent). A majority (62 and 65 percent for with and without missing SF-12 item[s]) of respondents had higher than high school education, with 10 and 7 percent having less than high school education; 27 and 28 percent with high school education. The majority of the respondents had an annual household income of more than \$30,000: 73 percent for patients with and 76 percent for patients without missing SF-12 item(s). The mean number of medical comorbidities (SD) was 1.3 (1.3) and 1.5 (1.3) for patients with and without missing SF-12 item(s), respectively (Table 1). Of the 112,008 SF-12 item responses covered by the 9,334 patients who missed one to 11 items ( $9,334 \times 12 = 112,008$ ), there were actually only 19,767 (18 percent) missing data items. Since the standard SF-12 algorithm produces missing summary PCS and MCS scores for patients with any missing SF-12 items, the 18 percent missing data

Table 1: Distribution of Patient Characteristics at Baseline by SF-12 Missing Status

	<i>Missing 1-12 items</i> ( <i>n = 9,431</i> )	<i>None Missing</i> ( <i>n = 20,877</i> )	<i>p-Value</i>
Age, mean (SD)	53 (13)	54 (12)	.0001
Gender, % female	67	60	.0001
Race, %			
White	72	81	.0001
Black/African American	4	3	
Hispanic/Latino	12	8	
Asian/Pacific Islander	9	7	
Other	3	2	
Education, %			
<High school	11	7	.0001
High school	27	28	
>High school	62	65	
Income, %			
≤ \$30,000	27	24	.0001
>\$30,000	73	76	
# of comorbidity conditions (SD)	1.3 (1.3)	1.5 (1.3)	.0001

SD, standard deviation.

leads to a huge loss of 92,241 (112,008 – 19,767) collected SF-12 items from these subjects.

Among the respondents, 9,984 (33 percent) identified themselves as having at least one of the four chronic diseases (heart, lung, diabetic, or back pain). The chronic cohort had a mean age (SD) of 57 (10) compared with 52 (13) for nonchronic cohort. There was a higher percentage of female patients in the nonchronic cohort (64 percent) versus the chronic cohort (58 percent). The education and income distributions are similar across chronic and nonchronic patients. The nonchronic patients had better health than the chronic patients in terms of number of comorbidities and SF-12 physical and mental summary scores. The mean number of comorbidities (SD) was 1.9 (1.5) and 1.2 (1.2) for chronic and nonchronic patients, respectively ( $p < .001$ ). The chronic cohort had a mean PCS score (SD) of 41.8 (11.5) and a mean MCS score of 50.8 (8.2), which is 18 and 3 percent lower than that of nonchronic cohort, respectively ( $p < .001$ ).

Missing SF-12 items were fairly common—29 percent (2,945) of the chronic cohort and 32 percent (6,486) of the nonchronic cohort had at least one missing SF-12 item. The majority of the patients were missing only a few items (less than six of the 12) and the proportion of patients decreases as the number of missing items increases. Among the patients with missing SF-12 items, 52 percent (1,543) and 50 percent (3,231) were missing one item, while 24 percent (704) and 22 percent (1,400) were missing two items for chronic and nonchronic patients, respectively. There were 76 percent (2,247) and 83 percent (5,402) of patients missing three or less SF-12 items, and .4 percent (12) and 3 percent (170) missing 10 or more SF-12 items for chronic and nonchronic patients, respectively (Table 2).

For patients with a given number of missing SF-12 items, the patterns of “missingness” was fairly spread out across different missing patterns. For example, there were 1,543 patients in the chronic cohort and 3,231 patients in the nonchronic cohort who were missing only one SF-12 item. Among these patients with one item missing, the percent of patients across the 12 items ranged from 4 percent (question #9) to 20 percent (question #7) for chronic patients and from 4 percent (questions #1) to 20 percent (question #7) for nonchronic patients. The three questions that were missing for more than 10 percent of respondents with only one item missing were: “limited in climbing several flights of stairs”; “limited in the kind of work or other activities”; and “did work or other activities less carefully than usual as a result of emotional problems.” Although not a uniform distribution, the proportion of patients missing early SF-12 items was about the same as the proportion who were missing later SF-12 items (Table 3).



Table 2: Distribution of Patients with Number of Missing SF-12 Items

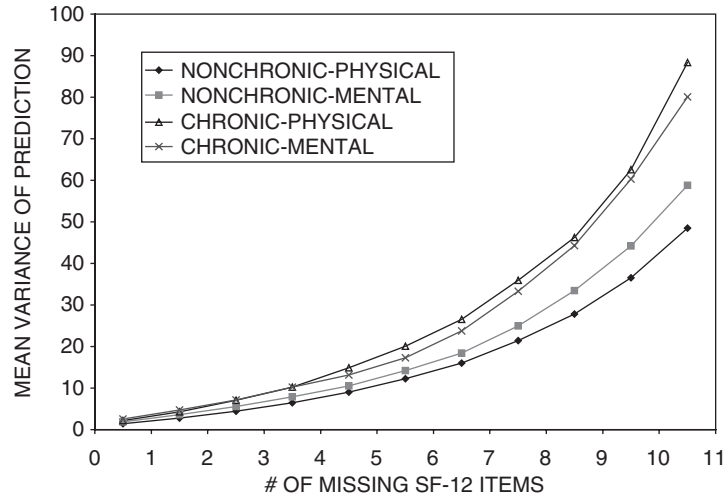
<i># of Missing SF-12 Items</i>	<i>Chronic Cohort Responders in 1996 (%) n = 9,984</i>	<i>Nonchronic Cohort Responders in 1996 (%) n = 20,324</i>
0	7,039 (70.5)	13,838 (68.09)
1	1,543 (15.45)	3,231 (15.9)
2	704 (7.05)	1,400 (6.89)
3	349 (3.5)	771 (3.79)
4	155 (1.55)	366 (1.8)
5	98 (0.98)	272 (1.34)
6	36 (0.36)	119 (0.59)
7	26 (0.26)	82 (0.4)
8	10 (0.10)	46 (0.23)
9	12 (0.12)	29 (0.14)
10	7 (0.07)	30 (0.15)
11	4 (0.04)	44 (0.22)
12	1 (0.01)	96 (0.47)

The means of variance of prediction of imputed PCS and MCS increased almost monotonically by the number of missing SF-12 items. For chronic patients, the means of variance of prediction of imputed range from 2.18 to 88.37 for PCS and from 2.58 to 80.08 for MCS, respectively. For nonchronic patients, the means range from 1.42 (1.89) to 48.53 (58.81) for PCS and from 1.89 to 58.81 for MCS, respectively. The increments of mean

Table 3: Distribution of Missing Patterns for the Patients Who Missed One SF-12 Item

<i>SF-12 item</i>	<i>Chronic (%) (n) n = 1,543</i>	<i>Nonchronic (%) (n) n = 3,231</i>
General health (#1)	4.34 (67)	3.71 (120)
Limited on moderate activities (# 2)	6.35 (98)	6.38 (206)
Limited on climbing flights of stairs (#3)	13.16 (203)	12.69 (410)
Accomplished less because of physical health (#4)	9.92 (153)	8.67 (280)
Limited in the kind of work because of physical health (#5)	14 (216)	12.84 (415)
Accomplished less because of emotional problem (#6)	8.23 (127)	7.67 (247)
Did work or other activities less carefully because of emotional problem (#7)	20.16 (311)	19.72 (637)
Pain interfered with normal work (#8)	6.61 (102)	8.29 (268)
Felt calm and peaceful (#9)	3.63 (56)	4.61 (149)
Had a lot of energy (#10)	4.67 (72)	5.20 (168)
Felt downhearted and blue (#11)	4.93 (76)	5.91 (191)
Problems with social activities (#12)	4.02 (62)	4.33 (140)

Figure 1: Distribution Variation of Imputation



variance of prediction also increase with the number of missing SF-12 items (see Figure 1). For both chronic and nonchronic patients, when the number of missing items was six or more, the increments of the mean variance of prediction compared with fewer missing items were generally above three and the means of the variance of prediction themselves reached above 10, indicating that the uncertainty because of missing SF-12 items was quite large. The means of imputed PCS and MCS scores showed a slightly downward trend as the number of missing SF-12 items increased. The width of confidence bands of mean imputed values significantly increased when the number of missing items was six or more, signifying that additional information beyond the nonmissing SF-12 items may be needed to increase the accuracy of imputed PCS and MCS scores. The distributions of standard errors of means of imputed SF-12 PCS or MCS scores derived from enhanced models are similar to that derived from the simple models but have smaller ranges except for MCS of the chronic cohort. The standard errors corresponding to a different number of missing items ranged from .28 to 4.77 for PCS and from .26 to 3.31 for MCS in the chronic cohort, and from .15 to 1.44 for PCS and from .15 to 1.13 for MCS in the nonchronic cohort.

The results of imputation of PCS and MCS by using multi-pattern regression based imputation were similar to that from MI. For chronic patients missing fewer than six items, the means (variances) for PCS and MCS were

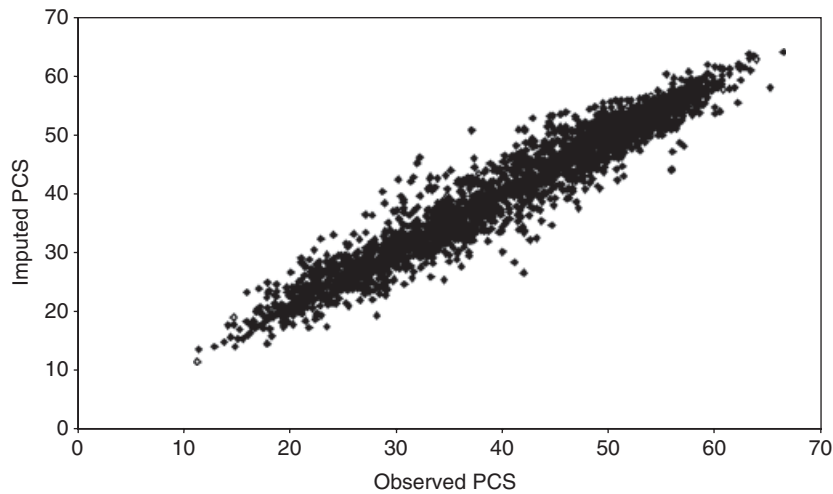
41.79 (133.16) and 50.28 (108.88) from MI compared with 41.83 (125.97) and 50.29 (101.33) from multi-pattern regression, respectively. For chronic patients missing six or more items, the means (variances) for PCS and MCS were 41.84 (133.12) and 50.65 (107.59) from MI compared with 42.07 (93.57) and 50.73 (66.48) from multi-pattern regression. The results of nonchronic patients were similar to that of the chronic patients. For nonchronic patients missing fewer than six items, the means (variances) for PCS and MCS were 50.87 (66.94) and 51.94 (80.50) from MI compared with 50.87 (62.05) and 51.92 (74.37) from multi-pattern imputation. For missing six or more items, the means (variances) were 50.82 (64.28) and 51.79 (78.12) from MI compared with 50.91 (41.55) and 52.11 (47.33) from multi-pattern imputation for PCS and MCS.

The relationship between PCS or MCS and other measures that were outside the imputation models remain similar before and after imputation. For those missing fewer than six items, the correlations between satisfaction and use of hospital with imputed PCS (MCS) were .09 (.15) and  $-.06 (-.02)$ , compared with .11 (.16) and  $-.06 (-.02)$  with observed PCS and MCS. For those missing six or more items, the correlations were .09 (.19) and  $-.09 (-.06)$  for imputed, compared with .07(.22) and  $-.12 (-.02)$  for observed, respectively.

In the simulation analysis, means of the imputed PCS and MCS were very close to that of the observed and the differences between the imputed and observed means were quite small, ranging from  $-.114$  to .158. The difference in absolute values were smaller for missing fewer than six items (with the simple imputation, range from .006 to .057) compared with that for missing  $\geq 6$  items (with the enhanced models, range from .051 to .158). There were no systematic directions of these differences for either the simple or the enhanced models. Stratified by missing  $<6$  or  $\geq 6$ , the statistical tests for the null hypotheses that the differences between the observed and the imputed PCS or MCS scores equal to zero yielded nonsignificant results ( $p > .05$ ) for both simple and enhanced imputation models.

Product-moment correlations between the imputed and the observed SF-12 PCS and MCS scores were larger for those missing fewer than six items, ranging from .96 to .97, compared with the correlations that were calculated for those missing six or more items (.77 to .83). The relationships between the imputed and the observed PCS or MCS scores were similar across the range of the scores but slightly tighter at the two ends than at the middle ranges of the scores. As an example, Figure 2 shows a scatter-plot of the imputed and the observed PCS for chronic patients missing fewer than six items. The difference between correlation coefficients derived by the enhanced and the simple

Figure 2: Scatter Plot of Observed versus Imputed Physical Component Summary (PCS) for Chronic Disease with Missing < 6 SF-12 Items



models for missing six or more items was consistently larger than the difference of correlations for those missing fewer than six items. The differences of correlation between the observed and the imputed SF-12 PCS and MCS derived from the simple models and the enhanced models were all positive. Except for the MCS score in chronic patients, all these increments of correlations between the imputed and observed SF-12 PCS or MCS scores were statistically significant for patients missing six or more SF-12 items ( $p < .05$ ) (Table 4).

The validation analyses using 1998 follow-up data produced similar results as the 1996 data analyses and further confirmed the findings. The validation analyses with follow-up data again showed that the correlation coefficients between the observed and the imputed SF-12 summary scores derived from the enhanced models have significant improvement over that from the simple models if the number of missing SF-12 items was six or greater (Table 5).

## DISCUSSION

For the large cohort of patients in the PVCS ( $n = 30,308$ ), about 30 percent (9,431) of patients submitted a survey with one or more SF-12 items missing.

Table 4: Correlations between Observed and Imputed SF-12 Summary Scores (Simulation Using 1996 Baseline Patients)

# of Missing SF-12 Items	SF-12 Score	Model	Chronic		Nonchronic		
			Diff	p-Value	Diff	p-Value	
< 6	PCS	Simple model	0.9722		0.9642		
		Enhanced model	0.9741	0.0019	.120	0.9665	0.0023
≥ 6	MCS	Simple model	0.9661		0.9612		
		Enhanced model	0.9662	0.0001	.841	0.9616	0.0004
	PCS	Simple model	0.7959		0.7666		
		Enhanced model	0.8331	0.0372	.003	0.8042	0.0376
MCS	Simple model	0.7757		0.7662			
	Enhanced model	0.7829	0.0072	.311	0.7796	0.0134	.022

PCS, physical component summary; MCS, mental component summary; Diff, difference.

The relatively high missing rate of SF-12 items in the PVCS cohort could be because of a number of factors, such as patient characteristics, content, and the length of survey. The PVCS surveys were full-length questionnaires consisting of as many as 131 items. The self-administered mail mode of data collection used in the PVCS is more likely to lead to missing data than a telephone or face-to-face survey. Although some other studies have reported lower rates of missing data for the SF-12 items than were observed in this study (e.g., NCQA 2002) reported that Medical Health of Seniors survey yielded one or more missing SF-36 items for about 25 percent (NCQA 1998), missing SF-12 items are not uncommon. Because of the large sample size, the differences in distribution of patient characteristics at baseline between those with and without

Table 5: Correlations between Observed and Imputed SF-12 Summary Scores (Simulation Using 1998 Follow-up Patients)

# of Missing SF-12 Items	SF-12 Score	Model	Chronic		Nonchronic		
			Diff	p-Value	Diff	p-Value	
< 6	PCS	Simple model	0.9746		0.9665		
		Enhanced model	0.9750	0.0004	.78	0.9691	0.0026
≥ 6	MCS	Simple model	0.9689		0.9574		
		Enhanced model	0.9698	0.0009	.97	0.9579	0.0005
≥ 6	PCS	Simple model	0.7928		0.7828		
		Enhanced model	0.8332	0.0404	.026	0.8354	0.0526
	MCS	Simple model	0.8132		0.7573		
		Enhanced model	0.8238	0.0100	.288	0.7701	0.0128

PCS, physical component summary; MCS, mental component summary; Diff, difference.

missing SF-12 item were significant, the actual differences were not substantial; for example, mean age differed by only 1 year. Among the patients who had missing SF-12 item(s), most of the patients (> 80 percent) missed fewer than three items. Since the standard SF-12 algorithm produces missing PCS and MCS scores for patients with any missing SF-12 item, the nearly one-third missing SF-12 scores would significantly weaken the analytical power and could produce inaccurate or biased estimates. Although the 12 SF items vary somewhat with respect to rates of missing, overall, the missingness was fairly spread across many patterns. For patients with fewer than six items missing, all the missing patterns were present (e.g., there are 220 missing patterns with three items missing and all the 220 patterns were present in the data). The number of patients with a large number of missing items is small, and only a subset of the possible missing patterns was observed for those with six or more missing items.

When the number of missing SF-12 items increases, the available information for imputing SF-12 scores decreases and uncertainty increases and enhanced imputation models are worth considering. We can see this phenomenon through two measures: the mean variance of prediction from the imputation models and the standard errors of the mean imputed scores. The mean variance of prediction increases almost monotonously as the number of missing SF-12 items escalates. Although the increase of standard errors could be partially affected by the fact that the number of patients with a substantial number of missing SF-12 items is fewer than those with smaller numbers of missing items, its increase is consistent with the increase of mean variance of prediction, which is a “sample-size-free” estimate of the variance of predicted value from imputation models.

Results from simulation analyses provided strong support for our algorithm. By using those who have complete SF-12 items and other predictors, we were able to link and compare the imputed SF-12 summary scores to the observed values to validate the proposed imputation algorithm. By testing the differences between the observed and the imputed SF-12 scores using the identified cut point, we found that the multi-pattern regression based imputation algorithm produced good estimates of the SF-12 scores using either available SF-12 items only or the available SF-12 items plus patient demographics, smoking status, and the medical comorbidities. However, with most of the SF-12 items already present in the model, the enhanced models fail to bring the imputed scores significantly closer to their observed values and did just about the same as the simple models. In fact, the correlation coefficients between the observed and the imputed PCS and MCS scores from the enhanced model were about the same as that from the simple model. In contrast,

when the number of missing SF-12 items was six or more, the correlation coefficients between the observed and the imputed SF-12 scores derived from the enhanced models were significantly improved compared with that derived from simple models. This is because when more SF-12 items were missing, the predictive power of the imputation models that only use the few available SF-12 items leave a large amount of unexplained variation in SF-12 scores. Under such circumstances, the enhanced model with additional patient information is able to significantly improve the accuracy of the imputed values.

To further evaluate the algorithm, the same validation analysis was repeated using the 1998 follow-up data. The validation results based on the 1998 data once again indicated that additional patient information was needed to improve the quality of the imputed SF-12 scores when the number of missing SF-12 items was six or greater. However, when the number of missing SF-12 items was less than six, the majority of the SF-12 items available still carried most of the information about the SF-12 scores. In such situations, the additional patient information can make only a trivial contribution to improve the accuracy.

Since more missing SF-12 items is associated with more uncertainty, in circumstances with more missing, one might benefit from using data from outside the scale. The additional information provided by patient demographics, smoking status, and medical comorbidity accounts for the enhanced model's ability to account for variation in SF-12 scores when most of the items were missing. Although it appears that the use of data exogenous to the original scale makes the relationship between PCS or MCS and the outside scale measures more complicated, this linkage between SF-12 measures to the likely predictors during imputation have in fact tightened the real relationship between the quality of life measures and the potential predictors.

When the number of missing items is six or more, even with the enhanced models, the quality of the imputed scores is not the same as that when missing is less than six. Therefore imputed scores should be carefully evaluated for patients with a large number of missing items, particularly if the missings are more concentrated on physical or mental items. Also, in order to implement the enhanced models, one needs to collect the additional measures beyond the SF-12 scale. Although extra effort may be needed to collect the information, the value and trade-off for being able to use enhanced models are likely to be worthwhile in certain circumstances.

Given large variations in observed measures of health related quality of life among patients and the complex scoring algorithm of PCS and MCS, naïve methods, such as mean imputation, will not be powerful enough to

generate accurate imputations. Even for simple summated scales, naïve mean imputation will likely lead to bias since it ignores the variation within an item and the difference from patient to patient. For complicated methods, one could, for example, consider using a MI approach to introduce sampling variations to the imputed values. The obvious limitation of single imputation methods is that it cannot reflect sampling variability under one model for nonresponse or uncertainty about the correct model for nonresponse. MI imputes multiple PCS and MCS scores for a given patient. If the multiple imputed values are derived from one model, one can combine them to create an inference that validly reflects sampling variability. If the multiple imputed values are from more than one model, uncertainty about the correct model is displayed by the variation in valid inference across the models (Little and Rubin 1987). For the particular question of imputing SF-12 PCS and MCS scores using our data, the point and variance estimates are similar between MI and multi-pattern regression-based imputation, which has a level of complexity similar to other single imputation methods. The differences of point estimates are trivial ranging from  $-.32$  to  $.02$ , which are  $.62$  to  $.04$  percent of the point estimate from MI. The differences in variance estimates are larger than that of the means but not great and the estimated variances from multi-pattern imputation are smaller than that from MI. This is probably because multi-pattern imputation imputes a single value for each missing and does not reflect the uncertainty about the predictions of the unknown missing values. However, most of the MI algorithms are based on the assumption of multivariate normal distribution and most of the SF-12 items and the outside scales measured used in the enhanced models are either binary or nominal categorical measures. Furthermore, more steps are involved to generate the multiple imputed values and then to combine the results to make an inference. We aimed to create an effective SF-12 imputation method for general SF-12 users and the algorithm proposed in the paper uses a straightforward regression approach and yields satisfying results.

There are several limitations in these analyses that should be noted. First, although cardiovascular disease, pulmonary disease, diabetes, and back pain are prevalent chronic conditions in the general adult population, the definition of chronic disease we used in the study is selective, and it does not cover the universe of chronic disease. Secondly, because of the complexity of multi-pattern imputation, no closed-form solutions of standard errors of the mean imputed values and the difference of the correlation coefficients were available. The bootstrapped standard errors were used for statistical testing. Although we used a large number of replicates in an attempt to increase



the accuracy, the precision of the estimated standard errors are limited by the bootstrapping algorithm and the software used. Lastly, the statistical testing for the differences of correlation coefficients was based on asymptotic distribution and it may not work well with a finite sample.

In summary, the proposed imputation algorithm for SF-12 physical and mental component summary scores can help with the prevalent problem of missing data for some of the SF-12 items. This algorithm estimates the physical and mental scores that would be missing in using the standard SF-12 algorithm. The approach used in the paper can also be applied to other multi-item health scores.

## ACKNOWLEDGEMENT

This research was supported by grants from Agency for Health Care Research and Quality (AHRQ) (U01 HS09951-01), American Association of Health Plans (AAHP), and Robert Wood Johnson Foundation (RWJ). Drs. Liu and Hays were also supported in part by the UCLA/DREW Project EXPORT, National Institutes of Health, National Center on Minority Health & Health Disparities (P20-MD00148-01), and the UCLA Center for Health Improvement in Minority Elders/Resource Centers for Minority Aging Research, National Institutes of Health, National Institute of Aging (AG-02-004).

The authors wish to acknowledge the technical support from Victor Gonzalez for manuscript preparation.

## REFERENCE

- Damberg, C. L., and L. Bloomfield. 1997. *Physician Value Check Survey Final Report*. San Francisco: Pacific Business Group on Health.
- Efron, B. 1979. "Bootstrapping Methods: Another Look at the Jackknife." *Annals of Statistics* 7: 1-26.
- Efron, B., and R. Tibshirani. 1986. "Bootstrap Measures for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy." *Statistical Science* 1: 54-77.
- Goldstein, R. 1996. "Sed10: Patterns of Missing Data." *Stata Technical Bulletin* 32: 12-13. (Reprinted in *Stata Technical Bulletin Reprints* 6, 115).
- . 1996. "Sed10.1: Patterns of Missing Data, Update." *Stata Technical Bulletin* 33: 2. (Reprinted in *Stata Technical Bulletin Reprints* 6, 115-116).
- Kahn, K. L., H. H. Liu, J. L. Adams, W. P. Chen, D. Tisnado, D. M. Carlisle, R. D. Hays, C. M. Mangione, and C. L. Damberg. 2003. "Methodological Challenges

- Associated with Longitudinal Studies Regarding Quality of Care and Health Status.” *Health Services Research* 38 (6): 1579–98.
- Little, R. J. A., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc.
- National Committee for Quality Assurance (NCQA). 1998–2004. *HEDIS®*, Volume 6: *Specifications for the Medicare Health Outcomes Survey*. Washington, DC: NCQA Publication.
- Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, Inc.
- . 1996. “Multiple Imputation after 18+ Years.” *Journal of the American Statistical Association* 91: 473–89.
- Stata Corp. 1999. *Stata Statistical Software: Release 6.0*. College Station, TX: Stata Corporation.
- Ware, J. E. Jr., M. Kosinski, and S. D. Keller. 1995. *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales*. Boston: The Health Institute, New England Medical Center.
- Ware, J. E. Jr., M. Kosinski, D. M. Turner-Bowker, and B. Gandek. 2002. *How to Score Version 2 of the SF-12® Health Survey (With a Supplement Documenting Version 1)*. Lincoln, RI: Quality Metric Incorporated.
- Ware, J. E., K. K. Snow, M. Kosinski, and B. Gandek. 1993. *SF-36 Health Survey Manual and Interpretation Guide*. Boston: New England Medical Center, The Health Institute.

