Accounting for deaths in longitudinal studies using the SF-36: the performance of the Physical Component Scale of the Short Form 36-Item Health Survey and the PCTD

Paula Diehr, University of Washington
Accounting for Deaths in Longitudinal Studies Using the SF-36
The Performance of the Physical Component Scale of the Short Form
36-Item Health Survey and the PCTD

Paula Diehr, PhD,*† Donald L. Patrick, PhD, MPH,† Mary B. McDonell, MS,§ and Stephan D. Fihn, MD, MPH†‡§

Background. Commonly used measures such as the Physical Component Scale of the Short Form 36-item health survey (PCS) are undefined at death, limiting longitudinal analyses to survivors, a healthier cohort that cannot be identified prospectively, and that might have had little change in health. One proposed approach is to transform the PCS into the Physical Component Transformed, with Deaths included (PCTD), which is the probability of being healthy 1 year later and for which deaths logically have a value of zero. Data missing for other reasons than death have not been considered.

Objective. To examine the performance of the PCTD, to determine the influence of including deaths, the additional effects of imputing missing values and adjusting for covariates, and the calibration of the PCTD in different populations.

Methods. We imputed missing values of the PCTD, calling the new variable the PCTDI. We compared the distributions of the PCS, PCTD, and PCTDI cross-sectionally and over time. In 3 different populations, we determined whether the PCTD accurately predicted the probability of being healthy 1 year later.

Results. The patients who died did not have extreme values on the PCTD. The experience of the cohort was best described by the PCTDI. The calibration of the PCTD was surprisingly good in all the populations examined. Results were similar for the physical function index.

Conclusion. The PCTDI is an improvement over the PCS, in which patients who had died have no influence, and over the PCTD, where they might have too much influence. We recommend the PCTDI for longitudinal analyses of physical health when deaths occur, for primary or secondary analysis.

Key words: Death; health-related quality of life; QALY; longitudinal; health status; SF-36. (Med Care 2003;41:1065–1073)
complete cases, discarding or reporting separately the data for people who died. This approach limits the analysis to survivors, who began or remained in better health, and could miss important changes over time. Paradoxically, a comparison group with higher mortality might have an advantage under such an analysis, because more of its sickest members would be removed from the analysis.

Diehr et al. proposed a monotonic transformation of a health measure that has no value for death to a different variable for which death has a logical value. The method was illustrated using the SF-36 physical component score (PCS), although the necessary transformations were provided for all of the usual subscales, and for the single-item self-rated health. The approach is to transform the PCS to the estimated probability of being healthy in 1 year, and then set the patients who died to zero, because they have no probability of being healthy. We refer to this transformed measure as the PCTD, for Physical Component, Transformed, with Deaths included. In the earlier paper, the PCTD was found to have about the same power as the PCS in comparing 2 groups cross-sectionally, and to have more power in detecting change over time. Because it included deaths, results for the PCTD could be generalized to the entire initial cohort, not just to the survivors. The area under the curve over time (AUC) could also be interpreted as years of healthy life.

The PCTD transformation equation was derived using data from the Ambulatory Care Quality Improvement Project (ACQUIP) study of veterans, which included mainly older, chronically ill men. Its performance has not been examined further, in the original setting or elsewhere.

**Methods**

**Physical Component Score**

The SF-36 is a widely used health status measure developed for the Medical Outcomes Study (MOS). Results are usually reported in 8 subscales and 2 summary scores, with a mean of 50 and a standard deviation of 10 in the original MOS reference population. Diehr et al. used logistic regression to estimate the probability of being “healthy” (a variable that has a value of 1 if the person is healthy 1 year later, and 0 otherwise) as a function of the baseline score for each component of the SF-36. That paper presented the regression coefficients needed to transform all subscales for 3 choices of the definition of “healthy”: being alive, or in excellent/very good/good health, or in the top 75% of Ware’s reference population 1 year later. Additional details were shown for the PCS, and we extend that example into the current paper. The original example defined “healthy” as being in the top 75% of the reference distribution. Now, we define “healthy” as being in excellent, very good, or good health, because that definition was found in the earlier paper to have somewhat better statistical properties. The logistic regression equation estimated in the previous paper was: \( \text{logit (healthy)} = -3.77 + 0.1089 \times \text{PCS} \). We used the resulting equation to transform all values of PCS into PCT, as follows:

\[
PCT = \frac{\exp(-3.77 + 0.1089 \times \text{PCS})}{1 + \exp(-3.77 + 0.1089 \times \text{PCS})} \times 100
\]

For example, a person with a PCS of 50 would have a PCT of 84, representing an 84% chance of being healthy 1 year later. PCT is the transformation of the observed PCS into the estimated probability of being healthy in 1 year. It is a monotonic transformation in that people with a high value of PCS will also have a high value of PCT and low values will correspond to low values. When the patients who had died are assigned a value of zero, the resulting variable is called the PCTD. We considered the following variants of the PCS: \( \text{PCT} = \text{PCS} \), transformed to the probability of being healthy 1 year later, using the logistic regression equation; \( \text{PCTI} = \text{PCT} \), with missing values imputed; \( \text{PCTD} = \text{PCT} \), with deaths set to zero; and \( \text{PCTDI} = \text{PCTD} \), with missing values imputed.

The PCT is used instead of the PCS when we want to compare the performance of the PCS with the other variables, because the PCT is on the same scale as the PCTD and PCTDI. The method used to impute the missing values not resulting from death is given in Appendix 1.

**Data**

Data were taken from 3 sources. Most data came from ACQUIP, a group-randomized trial to determine whether the outcomes of health care for veterans are improved when primary care providers have access to systematic assessments of their
patients’ general and condition-specific health and function. In the control group, SF-36 measures were obtained at year 0, year 1, and year 2. In the treatment group, which needed information to feed back to providers, measures were obtained at years \(-0.5, -0.25, 0, 0.5, 1.0, 1.5, \) and 2.0. We will refer to several subsets of the ACQUIP data. The original subset (n = 9894), comprising measures at years 0 and 1 in the control group and \(-0.5 \) and +0.5 in the treatment group, was used in the earlier paper to estimate the PCTD equations. We now define the baseline subset as the year 0 treatment and control data (n = 11,174). The cross-sectional subset is the year 2.0 treatment and control data (n = 5847). The longitudinal subset (n = 6640) is data from years 0, 0.5, 1.0, 1.5, and 2.0 in the treatment group only. The dying subset is the longitudinal data for treatment group members who died before year 2.5 (n = 582), and the living subset is its complement. The calibration subset was the year 1.0 and year 2.0 data for both groups (n = 6196). The latter subsets overlap very little with the original subset and therefore provide substantially independent information. In the baseline subset, 97% were men and mean age was 65. There was substantial loss to follow up, which is addressed later.

Baseline ACQUIP data included self-report of 23 baseline health problems: arthritis, coronary artery disease, cancer, congestive heart failure, chest pain, chronic obstructive pulmonary disease, depression, diabetes, drug abuse, heartburn, hypertension, HIV-AIDS, kidney or liver problems, myocardial infarction, osteoporosis, pneumonia, enlarged prostate, posttraumatic stress disorder, seizures, stroke, ulcers, or thyroid problems. These variables were used as covariates to adjust the PCS and PCTD for baseline health conditions, as explained subsequently in this article. Deaths were thought to be completely ascertained because the Department of Veterans Affairs provides substantial death benefits after the report of death.

We also examined the public use dataset from the MOS, in which the PCS was originally validated. We used the 1557 subjects who had known values of sex, age, PCS information at baseline (variable HLS30521), and self-rated health 1 year later (HLS30534). In this group, only 41% were men and mean age was 56. Data from a randomized trial of footwear for 400 diabetics, involving both veterans and HMO enrollees, were also used. In that study, 78% were men and the mean age was 62.

### Adjustment for Baseline Health Problems

Because health measures are usually analyzed in conjunction with other covariates, we regressed each PCS variable (PCS, PCTD, and so on) on age and the 23 baseline health problems noted previously. A person’s “adjusted value” is the residual from the regression plus the grand mean. This approach is analogous to comparing the mean health of 2 groups at some future time controlling for baseline characteristics; such an analysis is essentially a t test comparing the average residuals in the 2 groups. The grand mean was added to give the adjusted variables the same mean as the unadjusted variables.

### Analysis

The goals of this study were to examine the cross-sectional and longitudinal effects of including death, adjusting for baseline characteristics, and imputing missing values. We also examined whether the transformation equation developed in the original VA subset provided appropriate estimates of the probability of being healthy 1 year later in other populations. The data subsets for each analysis were described in the data section previously in this article.

In the cross-sectional analysis, we computed the usual PCS and the PCTD with and without adjusting for baseline characteristics and health problems. We constructed histograms of these variables to inspect their distributions, and to assess whether including the dead produced a distorted picture of the group’s health. In the longitudinal analyses, we examined 5 versions of the variable: PCS (the original), PCT, PCTI, PCTD, and PCTDI. We plotted the mean of each variable over time to demonstrate where the measures differed. Some results were reported separately for the dying subset.

To examine the generalizability of the published regression equations to different datasets, we transformed the baseline PCS to PCTD, which should be an estimate of the percent healthy 1 year later. We compared the estimated to the observed percent healthy, and fit a regression line between the observed and predicted values. A perfectly calibrated estimate would result in a 45° regression line passing through the origin. For this exercise, we used data from the MOS, the footwear trial, and the ACQUIP calibration subset.
Findings

The variables are defined in Table 1, which shows whether each variable is transformed, has deaths set to zero, or has missing values imputed. Table 1 shows how many of the 6640 longitudinal subjects had a known value at year 2.0. It then shows the percent with known values at year 2, separately by whether they were dead or alive 6 months after the end of the study (by year 2.5). For example, PCS was known at year 2.0 for only 3551 of the original 6640, comprising 9.7% of the 582 “dying” and 57.6% of the 6058 “living.”

Cross-sectional Comparisons. In the AC-QUIP cross-sectional subset, the distribution of the usual PCS was somewhat bell-shaped, with a mode at 27 and a mean of 32.6 (not shown). The low mean value is typical of veterans who receive health care at the Veterans Affairs Hospital.7 The adjusted PCS variable had a similar distribution (not shown). Figure 1 shows the distribution of the PCTD, which is somewhat trimodal. There is a mode for persons with an approximately 90% chance of being healthy 1 year later, a second mode with approximately 20% probability, and a third mode of those with no probability of future health because they were dead. The patients who had died are thus not outliers, but they might still have too much influence on the mean, because the PCTD is never missing for dead persons. PCTD at year 2 was known for 85% of the dying but only 58% of the living. Thus, there are “too many” deaths in Figure 1, which could lead to deaths being overly influential. This suggests that the missing data not resulting from death should be imputed in some way.

Figure 2 shows the adjusted PCTD, in which the distribution is surprisingly “bell-shaped” compared with Figure 1. People who had died (denoted by lighter bars) had low adjusted values, but there were many living persons with equally low adjusted values (that is, low residuals from the regression). Thus, although the unadjusted PCTD had a somewhat peculiar distribution, analyses that controlled for baseline health conditions would yield normally distributed residuals. The persons who had died did not have extreme adjusted PCTD values relative to the rest of the distribution. As expected, a few persons with large residuals had adjusted values outside the range of 0 to 100. The distribution of the PCTDI was similar to Figure 2 but with a larger number of living subjects (not shown).

Longitudinal Results. In the longitudinal subset, only 2310 persons had PCS data at all 5 times. Of the remaining 4330, all had baseline data but only 1241 had data at year 2. Those with complete data had better physical health on average than those with incomplete data (not shown). A complete case analysis of the PCS would thus be based on a smaller and healthier subset of the cohort. An available case analysis could also be misleading because the sicker veterans would be fully represented at year 0 but seriously underrepresented at year 2.

The average PCS over time for persons still alive 6 months after the end of the study declined slightly over time from 34 at year 0 to 32 at year 2 (not shown). The dying group decreased only a little more, from approximately 28 to 24. The surprising lack of change in a group known to be dying is in part the result of missing data and deaths; of the 582 in the dying group, only 59 had PCS responses at year 2. Figure 3 shows the plot of mean PCT, PCTI, PCTD, and PCTDI over time (using all available data) for only the dying. PCT, the top-most solid line in Figure 3, is just the transformation of the PCS with no adjustment for missing values and deaths, and so also showed little change over time. The PCTI, with imputed missing values, was similar. For PCTD, there was a steeper decline over time, because the value of zero was assigned for times after death. PCTDI, with imputed missing values, showed a steep decline over time, but less steep than for PCTD.

Our main goal was to describe the average health of the entire longitudinal cohort over time, which is shown in Figure 4. The 4 lines are identical at year 0, because everyone had a known baseline value. The lines diverge considerably over time, with PCT and PCTI showing only 2% to 3% points change over time, whereas PCTD decreased 7 points and PCTDI decreased 5 points. For PCT and PCTI, the deaths had no influence at year 2 (ie, the dead were missing on these measures). For PCTD, 85% of the dying had a value at year 2.0, versus only 58% of the living, suggesting that the deaths had too much influence (Table 1). The PCTDI data were 97% complete for the living and 99.6% complete for the dying, suggesting that the deaths had the appropriate amount of influence using the PCTDI.

We also performed these analyses transforming the Physical Function Index (PFI), one of the components of the PCS, to the analogous PFTD. Because results were very similar to those for the
PCS, we summarize them here. The histogram of the PFI at 2 years was fairly flat as compared with the bell-shaped distribution of the PCS. Figure 1 for PFTD had only one mode, at the low end, rather than 3 modes. Figure 2 was similar for the PCS and PFI. The PFI dropped approximately 4 points in 2 years for the living and 8 points for the dying. Figures 3 and 4 were very similar to those for the PCS. The cross-sectional distributions of the unadjusted PFI variables were thus a little different from those for the PCS, but the distribution of the adjusted variables and the longitudinal results were very similar.

**Generalizability to Other Populations.** The PCTD is the estimated probability of being healthy 1 year later using regression coefficients estimated from the original subset of the ACQUIP data. We wondered if the published prediction equations would perform well in other settings, such as the ACQUIP calibration subset, the original MOS data, or data from the footwear trial. Figure 5 shows results from the MOS data. The X axis is the predicted percent healthy at year 1 estimated from the PCS at year 0, that is, the PCTD at year 0. The Y axis is the observed percent healthy at year 1 (the percentage in excellent, very good, or good health). Each point represents a group of people with similar PCTD at year 0. The darker line is a least-squares fit to those points. The dotted line is the 45° line of perfect calibration and is very close to the fitted line. In the ACQUIP calibration data and in the footwear trial data, the 2 lines were virtually indistinguishable (graphs not shown). The calibration was much better than we had expected, suggesting that the published transformation coefficients might be usable in other populations, even though they were created from Veterans Affairs data. Summary statistics such as the AUC should have about the same interpretation across studies. Calibration of the PFI was virtually perfect for the ACQUIP calibration subset and fairly good for the footwear trial data (not shown). However, estimates were consistently approximately 7 points low (too pessimistic) in the MOS subset.

**Discussion**

This article examined the performance of a method to transform a health variable that has no value for death to a new variable that has such a value, using the PCS measure from the SF-36 as an example. We compared 4 measures: the PCS/ PCT, PCTI, PCTD, and PCTDI. The PCS is used frequently, but because it has no value for death, analyses based on it will exclude the sickest subjects. Imputation of the missing values, as in the PCTI, did not improve this situation. In both cases, average physical health was overestimated and decline over time was underestimated. The PCTD included all data for persons who died (zeros assigned after death), but had relatively less data for the living because some had missing values. This resulted in an overestimate of decline over time. The PCTDI included almost all subjects and therefore was most appropriate. We recommend the use of the PCTDI, which requires that missing data be imputed in some way. We used a person-specific regression estimate for imputation (see Appendix 1), but other approaches that make use of the individual’s known longitudinal data would probably have a similar effect.

The cross-sectional distributions of the PCTD and PCTDI were closer to uniform than to normal, but the distribution would rarely be a problem in analysis; the central limit theorem guarantees normality of the necessary test statistics, even with

### Table 1. Characteristics of Transformations of PCS (longitudinal subset; available information at year 2.0)

<table>
<thead>
<tr>
<th>Version</th>
<th>Transform</th>
<th>Deaths = 0</th>
<th>Impute Missing</th>
<th>No.* (of 6640)</th>
<th>Dying † (% of 582)</th>
<th>Living ‡ (% of 6058)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCS</td>
<td>X</td>
<td></td>
<td></td>
<td>3551</td>
<td>9.7</td>
<td>57.6</td>
</tr>
<tr>
<td>PCT</td>
<td>X</td>
<td></td>
<td></td>
<td>3551</td>
<td>9.7</td>
<td>57.6</td>
</tr>
<tr>
<td>PCTI</td>
<td>X</td>
<td>X</td>
<td></td>
<td>6416</td>
<td>95.0</td>
<td>96.8</td>
</tr>
<tr>
<td>PCTD</td>
<td>X</td>
<td>X</td>
<td></td>
<td>3985</td>
<td>84.7</td>
<td>57.6</td>
</tr>
<tr>
<td>PCTDI</td>
<td>X</td>
<td>X</td>
<td></td>
<td>6443</td>
<td>99.7</td>
<td>96.8</td>
</tr>
</tbody>
</table>

*Number (of the original 6640) who had a known value at year 2.0.
†Percent of the 582 “dying” persons who had a known value at year 2.0.
‡Percent of the 6058 “living” persons who had a known value at year 2.0.
surprisingly small sample sizes. When the PCTD was adjusted for baseline health problems, the distribution of the residuals was close to normal and there were no distributional concerns.

The transformation method is not restricted to the PCS. The earlier paper provided transformation equations for all of the subscales of the SF-36 and also for self-rated health. Because the trans-
formation is monotonic, it does not correct the recently noted problem that the maximum value of the PCS can be attained only by persons with very poor mental health. If both the PCS and the MCS were transformed to the PCTD and Mental Component Transformed, with Deaths included...
(MCTD), respectively, the latter 2 would no longer be uncorrelated, because dead persons would have a value of zero on both the PCTD and MCTD.

We were pleasantly surprised at how well the transformation equations worked in 3 other settings for the PCS. This suggests that it might be appropriate to use the published transformation equations without developing new equations in other populations. It also means that the interpretation of the PCTD as the probability of being healthy 1 year later, or of the area under the curve as years of healthy life, might be appropriate in other settings. The transformation equations for the PFI were less successful in the MOS data, suggesting that the relationship of the PFI to self-rated health might be somewhat different in those populations. Additional research is needed.

The transformation approach assigns an interpretable value to every level of the original scale, including death. The resulting variable is on a ratio scale, and its use (especially with imputation) permits almost the entire original cohort to be included in the analysis. Additional detail about the rationale for the PCTD, its derivation, its calculation, and its use in analysis were presented in the earlier paper.2

We speculate that most generic health and function measures could be transformed in this way to include an interpretable value for death. Further work is needed to determine whether condition-specific health measures, health behaviors, or clinical data can sensibly be transformed in this way.

**Limitations**

The study data were primarily from veterans, who were quite different in health from the general population. However, the calibration results for the other datasets were extremely encouraging. Experience with these transformations in more populations is needed. Other approaches for missing data might have yielded slightly different results. Our goal was to examine whether deaths were too influential if missing values were not imputed rather than to compare methods of imputation.

**Conclusion**

For longitudinal studies involving the PCS in which some subjects die, we recommend reporting both the PCS and the PCTDI to examine the influence of deaths on the study conclusions. The
PCTDI might also be useful in cross-sectional studies or studies without deaths, because of its understandable interpretation as the probability of being healthy 1 year later. The results might hold for other subscales of the SF-36. The transformation approach should be considered for additional variables.

References


Appendix 1

Imputation of Missing Data

Imputation of missing data cannot be shown to be accurate or even reasonable, but the method chosen should at least reflect the nature of the data and the study goals. One analysis of longitudinal data for older adults found estimates based on other people’s data to be “too healthy,” and that it was better to impute a person’s missing data from his own earlier or later nonmissing values on that same variable. Taking this approach, we fit a separate regression of PCS, PCT, and PCTD on time for each person, using the regression estimate to impute any missing value. For example, for a person with a PCTD of 50 at baseline and 40 at year 1, the regression equation would be $PCTD = 50 + 10 \times \text{time}$, and the missing values at 0.5, 1.5, and 2.5 years would be estimated as 45, 35, and 30, respectively. We did not impute values for persons with only one known observation. A regression based on only 2 points is likely to provide poor estimates, and we generally prefer not to impute values beyond the range of the known data. Most controversial, we used regression imputation even after death. Doing otherwise would have omitted some of the people who died from the comparisons involving the PCTI, which was our major comparison of interest. (It would not have had much substantive effect because the PCT and PCTI results were very similar.)