April, 1995

Including deaths when measuring health status over time

Paula Diehr, *University of Washington*

MEDICAL CARE
Vol. 33, No. 4, pp AS164–AS172, Supplement

92

Including Deaths When Measuring Health Status Over Time

PAULA DIEHR, PHD,*,† DONALD PATRICK, PHD,† SUSAN HEDRICK, PHD,†,‡
MARGARET ROTHMAN, PHD,§
DAVID GREMBOWSKI, PHD,† TRIVELLORE E. RAGHUNATHAN, PHD*
AND SHIRLEY BERESFORD PHD¶

Measuring health status over time is problematic when some subjects die, because death does not have a defined value on most health status measures. This situation is different from the usual missing data problem because the health status of the dead is, in a sense, known. We examined eight strategies for incorporating deaths into such analyses using three health status measures taken from two data sets, after which we used computer simulation to explore more fully the effect of deaths. The strategies differed in the amount of influence given to the deaths, varying from none (deaths were discarded) to complete (mortality itself was the health measure). The strategies that gave less influence to deaths tended to show more favorable changes in health over time, and therefore, tended to favor the group that had more deaths. The strategies that were more influenced by death showed more negative changes over time and favored the group with fewer deaths. The choice of strategy should depend on the goals of an intervention. For health promotion studies, we recommend recoding the health variables to estimate the probability that a person will be healthy in 2 years (or in some other period that can be estimated from the data). Key words: longitudinal; cohort; health status; missing data. (Med Care 1995;33:AS164–AS172)

Health status measured over time is needed for understanding the long-term impact of health-related interventions. Unfortunately, when subjects are followed over time, some of them die and follow-up measures (other than mortality) cannot be obtained. This is particularly a problem for studies of older or very sick populations, and when longer-term interventions are being evaluated.

Some health status measures are specifically designed to incorporate information about death. Sullivan[1] described two indexes based on a life table model that integrates morbidity with mortality: the expectation of life free of disability and the expectation of disability. Based on this approach, health-adjusted life expectancy can be calculated by combining a life table stationary population with cross-sectional population estimates of self-reported health

*From the Department of (1)Biostatics, University of Washington, Seattle, WA.

†From the Department of Health Sciences, University of Washington, Seattle, WA.

‡From the HSR&D Field Program, Veterans Administration Medical Center, Seattle, WA.

§From the Battelle Centers for Public Health Research and Evaluation, Arlington, VA.

¶From the Department of Epidemiology, University of Washington, Seattle, WA.

Supported in part by the Health Care Financing Administration under Cooperative Agreement #95-C-99161, and by the Department of Veterans Affairs Health Services Research and Development Service, Projects #SDR 85-07 and SDR85-071.

Address correspondence to Paula Diehr, PhD, Department of Biostatistics, University of Washington, Seattle, WA 98195.

status used in the calculation of a health status index.[2] Other measures based on adjustments of life expectancy have been developed and applied widely to different populations.[3] Utility or preference-based measures of health status incorporate values into the adjustment of life expectancy, resulting in indexes that combine duration of survival as modified by health status. Patrick and Erickson[4] reviewed these measures and their uses in health policy and health care evaluation.

Many health status measures, however, do not explicitly incorporate mortality. In these cases, researchers often report only the "complete" cases, discarding the deaths. However, this limits the evaluation to subjects in better health. Below, we consider alternative approaches for assigning numeric values to deaths so they can be included in health status analyses.

Accounting for deaths in longitudinal health surveys can be viewed as a problem of missing data.[5-11] However, most approaches assume that the missing data mechanism is "ignorable"[12]; that is, the "missingness" of a health status measure does not depend on the actual health status of the subject. This assumption is unlikely to hold in most of the situations of interest, because death is usually highly related to health.

In this paper, we propose eight strategies for incorporating deaths in the analysis of health status measures. We then compare these different strategies empirically using three health measures from two different studies. We also report the results of computer simulations conducted to permit more general conclusions about the effect of deaths on such analyses.

Methods

Health Status Variable Definitions

Three standard measures of health status were used. Each was recoded so that a higher value always would correspond to better health, and 100 was the highest score. The EVGFP measure is based on the usual self-rating of health as one of the following: excellent (100), very good (80), good (60), fair (40), or poor (20). The HSIP measure is based on the Sickness Impact Profile (SIP),[13] which was designed to detect moderate to severe levels of dysfunction due to sickness. The SIP measure takes on values of 0, for no impact, to a maximum of 84 in our data set. We recoded SIP scores as HSIP = 100—SIP, so that "no impact" is scored as 100. The lowest observed HSIP score was 16. The Quality of Well-being Scale (QWB),[5] is preference-weighted, with 100 indicating perfect health and 0 indicating death. The lowest nondeath score in our data was 38.

Data Sets Used

The first data set, "Day Care," comes from a study of 895 veterans who had a potential risk of needing nursing home care and who were randomized to receive either adult day health care or usual care.[15] The SIP and EVGFP measures were taken at baseline (T0) and 12 months later (T1). Approximately 25% of the subjects were omitted because they were alive at T1 but had no SIP score, often because they were too sick to participate in the lengthy interview. In the second study, 2524 volunteer health maintenance organization members older than 64 years of age were randomized to receive either special clinic visits and classes aimed at health promotion and disease prevention (HP/DP) or usual care.[16] The QWB and EVGFP measures were taken at T0 and 2 years later (T1). Only 6% had to be omitted. The goal of both studies was to determine whether the treatment slowed decline in health status.

Eight Strategies for Handling Deaths

Each strategy has been assigned a "rank" denoting the amount of influence given to

TABLE 1. Eight Sets of Codes for EVGFP

Strategy	Dead	Poor	Fair	Good	Very Good	Excellent
HP/DP						
1. Remove deaths	NA	20	40	60	80	100
2. Remove dead	NA	20	40	60	80	100
3. Healthy (yes/no)	0	0	0	100	100	100
4. Better/same/worse[a]						
5. Death arbitrary	0	20	40	60	80	100
6. Prob (healthy)[b]	0	19.0	34.9	75.9	92.8	95.8
7. Prob alive[b]	0	75.9	91.9	96.5	98.7	98.3
8. Alive (yes/no)	0	100	100	100	100	100
Day Care						
6. Prob (healthy)[b]	0	20.3	27.9	55.9	52.9	48.2
7. Prob (alive)[b]	0	71.4	82.1	84.3	80.4	74.1

EVGFP, Excellent, very good, good, fair, poor; HP/DP, health promotion and disease prevention study; Day Care, Adult Day Health Care project; NA, not applicable.

[a]For Better/same/worse, improvement was coded +1, same 0, decline -1.

[b]For SIP, the estimated probabilities in percents are as follows: prob(alive) = 89.04 - 0.227* Sickness Impact Profile (SIP), and Prob (healthy) = 53.89 - 0.5875*SIP. For Quality of Well Being (QWB), the estimated probabilities in percents are as follows: Prob (alive) = 37.1 + 1.497 QWB - 0.00916 QWB^2, and Prob (healthy) = -139 – 5.036 QWB + 0.02746 QWB^2. Equations were derived from regression of SIP (QWB) on Alive (yes/no) or on Healthy (yes/no) as defined in the table.

death, ranging from no influence (discard the deaths) to full influence (use mortality itself as the health status measure). We first describe the eight methods organized by the type of adjustment that is made for death, rather than list them in rank order. We will describe the use of EVGFP in some detail and briefly discuss the other variables. Table 1 shows the codes assigned under each strategy.

One strategy is to remove the deaths (Strategy 1). This strategy uses all people at T0 but omits the deaths at T1. The sample sizes at T0 and T1 therefore are different. It is used implicitly by investigators who present the average health at baseline and at follow-up. It also is known as "available case analysis." A related strategy is to remove the dead (Strategy 2), which omits the people who died from both the T0 and the T1 samples, leaving a sample who were alive at both times. This is known as "complete case analysis."

Another common approach is to recode the data as better, same, or worse (Strategy 4). Each person is assigned +1 if his or her health is better at T1 than at T0, 0 if it is the same, and -1 if it is worse (including deaths). For EVGFP, any change of categories is classified as better or worse. For HSIP, a change of less than five points is considered the same, and more than that is coded as +1 or -1, depending on the direction. (There is documentation that a three-point change in the SIP is clinically important in some settings.[17]) For the QWB, a change of less than two points is considered the same.[14]

Another way to include deaths is to recode them as healthy yes/no (Strategy 3). In this method, each person's T0 and T1 scores are coded as 100 if they are healthy (in excellent, very good, or good health) and 0 if their health is otherwise (including deaths). For HSIP, a score of 90 to 100 was considered healthy (and given a score of 100). For QWB, 60 to 100 was considered healthy. These thresholds were chosen to yield approximately the same number of people in good health as there was in the EVGFP measure. A variant of this approach, which is included only for completeness, is alive yes/no, in which subjects are coded 100 if they are alive and 0 if dead (Strategy 8).

Because death usually is considered even worse than poor health, another reasonable approach is that in which death is recoded to an arbitrary extreme value (Strategy 5). Deaths were given a value of 0 for all variables, representing the worst health status. This is an appropriate value for the QWB but is arbitrary for the other measures.

Using an approach similar to that used in the Medical Outcomes Study (Rogers WH, Ware JE Jr. Personal communication, 1994), we recoded the data according to the percent probability of living 2 more years (Strategy 7). This approach recodes the known values to approximate the probability of being alive at a particular time in the future, which can be estimated from available data. "Dead" then is given the value of 0. In the HP/DP study, 98.3% of those in excellent health at T0 were still alive at T1. As shown in Table 1, we assigned the value of 98.3 to patients whose health was excellent. Similarly, we recoded poor health as 75.9. The probability of being alive at T1 was calculated by regression for HSIP and QWB, predicting dead/alive at T1 as a function of the HSIP or QWB at T0, and using the resulting equation (shown in Table 1) to provide the estimated probabilities. For Day Care, the probability of living 1 year was used. It may be preferable to use a shorter time period, because this will be more relevant to current health status.

Because Strategy 7 emphasizes death rather than health, we also considered the percent probability of being healthy in 2 years (Strategy 6). The percent probability of being in excellent, very good, or good health at T1 was used. For example, in Table 1, excellent was recoded as 95.8, and poor as 19.0. As above, death was coded as 0, and the estimated probabilities for HSIP and QWB were determined by regression of "healthy at T1" (based on EVGFP) on HSIP or QWB, respectively.

Analysis

For Strategies 2, 3, and 5-8, we used a one-sample t-test of whether the mean change (T1 minus T0) differed significantly from zero. We also compared the treatment to the control group by using an unpaired t-test on the change scores. The value for Strategy 4 (better, same, or worse; coded as +1, 0, or -1) was treated as a change score. For Strategy 1, there is no standard person-based analysis, because some but not all of the people for whom T0 data were available also were used at T1. For the sake of completeness, we performed two comparable t-tests. For change, we assumed (incorrectly) that the T0 and T1 scores were independent and performed a two-sample t-test. For treatment versus control, we used the means and standard errors computed for the treatment and control group change scores in a two-sample t-test. The first analysis is probably conservative because it ignores correlations between an individual's measures at T0 and at T1.

Simulation

The relative performance of the eight strategies might be sensitive to special characteristics of the two data sets. To increase the generalizability of the findings, we also performed a computer simulation of the EVGFP results. We created a new treatment group and a new control group by sampling at random, with replacement, from the combined set of treatment and control subjects. The null hypothesis was thus true, because treatment and control subjects were drawn from the same population. We then calculated new values for Strategies 6 and 7, assigned the label of "treatment" to the group with more deaths, and calculated the same t statistics as above. This was repeated 1000 times, and the average t statistic was calculated for each strategy. In an additional simulation, we doubled the number of deaths in the populations before creating the treatment and control samples.

TABLE 2. Description of Two Studies

	Day Care		HP/DP	
	Treatment	Control	Treatment	Control
N	317	305	1179	1195
Mean age	71.5	72.5	72.6	72.7
% Female	4.1	5.3	60.6	62.1
% Died[a]	23.8	20.7	4.8	2.7[b]
At T0				
% Healthy (E, V, G)	30.9	34.8	82.4	84.8
Mean HSIP	66.2	65.3	NA	NA
Mean QWB	NA	NA	70.5	70.1

Day Care, Adult Day Health Care project; HP/DP, health promotion and disease prevention study; E, excellent; V, very good; G, good; NA, not applicable; HSIP, health status based on Sickness Impact Profile; QWB, Quality of Well-Being Scale.

[a]Percentage that died in 1 year for day care, and in 2 years for HP/DP.

[b]The treatment and control groups had significantly different mortality in the HP/DP study.

Findings

Demographics

Table 2 describes the two data sets used in the study. The Day Care study included frail veterans whose mean age was 72. Only 4.7% were female, and 22.2% had died before the 1-year follow-up. In contrast, the HP/DP subjects were volunteer seniors enrolled in a health maintenance organization. Only 3.7% died in 2 years. Sixty percent of the subjects were female. In the Day Care study, only 33% rated themselves as healthy (in excellent, very good, or good health) at T0; approximately 84% of the HP/DP subjects rated themselves as healthy at T0. Thus, although both of these studies are randomized trials of interventions designed to slow decline in health status, they are quite different. In both studies, even though the treatment and control groups were very similar, there were more deaths in the treatment group. This difference was statistically significant in the HP/DP study.

A positive value indicates that treatment was better than control. Two questions must be considered. First, did health change over time? Second, was that change in health significantly different in the treatment and control groups?

Change Over Time

Table 3 shows a summary of the change-score t statistics, separately for the treatment group and the control group. They are in rank order, where a higher rank means that death is more influential. For example, for Strategy 1 (remove deaths) for the Day Care study, the t statistic based on EVGFP for the treatment group was +3.4, denoting a significant improvement over time in health status. For Strategy 8, however, the t value of -9.9 denotes a significant decline in health. There is a strong ordering effect, in seven of the eight columns, with Strategies 1 and 2 showing more positive effects and Strategies 7 and 8 showing more negative changes over time. At the bottom of each column are correlations between the t statistic and the rank. All but one of these correlations are below -.9, indicating a strong linear trend, although the ordering is not perfect.

TABLE 3. T-Statistics[a] for Change over Time in Both Groups

	Day Care		HP/DP	
	EVGFP	HSIP	EVGFP	QWB
T-Statistics for the treatment group				
Rank strategy				
1. Remove deaths	3.4	1.6	-1.7	-0.9
2. Remove dead	3.3	0.7	-4.4	-2.0
3. Healthy (yes/no)	0.4	-0.4	-4.6	-3.4
4. Better/same/worse[b]	-2.8	-4.3	-6.6	-3.7
5. Death arbitrary	-2.6	-8.3	-7.8	-7.4[c]
6. Prob (healthy)	-3.7	-7.9	-6.6	-7.6
7. Prob (alive)	-9.6	-9.8	-8.1	-8.0
8. Alive (yes/no)	-9.9	-9.9	-7.7	-7.7
Correlation with rank	-0.97	-0.96	-0.89	-0.94
T-Statistics for the control group				
Rank strategy				
1. Remove deaths	2.1	0.7	-5.7	-0.2
2. Remove dead	2.0	0.4	-9.8	-0.9
3. Healthy (yes/no)	0.1	-0.6	-6.9	-1.0
4. Better/same/worse[b]	-2.4	-4.1	-11.1	-2.0
5. Death arbitrary	-3.0	-7.4	-11.4	-4.8[c]
6. Prob (healthy)	-3.2	-7.1	-8.0	-4.2
7. Prob (alive)	-8.5	-8.8	-7.4	-5.7
8. Alive (yes/no)	-8.9	-8.9	-5.7	-5.7
Correlation with rank	-0.96	-0.96	0.11	-0.96

Day Care, Adult Day Health Care project; HP/DP, health promotion and disease prevention study; EVGFP, excellent, very good, good, fair, poor; HSIP, health status based on Sickness Impact Profile; QWB, Quality of Well-Being Scale.

[a]A positive value means that health improved over time.

[b]Compare proportion better to proportion worse.

[c]For the QWB, 0 is the preference-weighted value for deaths, and is not arbitrary.

Comparison of Change in Treatment and Control Groups

Table 4 shows t statistics for the comparison of the treatment and control groups. Again, the correlations with rank are all considerable, with Strategies 1 and 2 showing more positive effects for the treatment group (more positive t statistics). In column 3, the t statistics actually ranged from significantly positive to significantly negative. In columns 1 and 2, the results would not have been statistically significant in any case.

Simulation Results

Strategies that gave less emphasis to death favored the treatment group, and those that gave more emphasis favored the control groups. To be certain that this was because the treatment group had more deaths, we conducted a simulation study in

TABLE 4. T-Statistics[a] for Treatment Group versus Control Group

	Day Care		HP/DP	
	EVGFP	HSIP	EVGFP	QWB
Rank strategy				
1. Remove deaths	1.0	0.6	2.8	-0.5
2. Remove dead	0.9	0.2	3.8	-0.8
3. Healthy (yes/no)	0.2	0.2	1.6	-1.8
4. Better/same/worse[b]	-0.2	0.0	3.0	-1.2
5. Death arbitrary	0.3	-0.5	2.1	-2.6[c]
6. Prob (healthy)	-0.2	-0.5	0.7	-3.0
7. Prob (alive)	-0.9	-0.9	-1.9	-2.9
8. Alive (yes/no)	-0.8	-0.8	-2.7	-2.7
Correlation with rank	-0.93	-0.97	-0.88	-0.9

Day Care, Adult Day Health Care project; HP/DP, health promotion and disease prevention study; EVGFP, excellent, very good, good, fair, poor; HSIP, health status based on Sickness Impact Profile; QWB, Quality of Well-Being Scale.

[a] A positive value means that treatment was better than control.

[b] Z-statistic for trend was used.

[c] For the QWB, 0 is the preference-weighted value for deaths, and is not arbitrary. Note that the t-statistics for strategies 6, 7, and 8 are similar to that for strategy 5 (the true QWB).

which the two groups were drawn from identical populations, but the group with more deaths was labeled as the treatment group. The ordering of the average t statistics was very similar to that shown in Table 4, with correlations well below -.9. Therefore, even when the null hypothesis was true, the ordering effect for the eight strategies persisted. When the number of deaths in the population was doubled, the t statistics for Strategy 1 became more positive, and for Strategies 3 to 8, more negative. (Strategy 2 discards the dead subjects and therefore did not change). Therefore, the differences among the strategies is due primarily to deaths and not other features of the data.

Summary and Discussion

Eight strategies were examined for analyzing health data in which some T1 measurements were missing because death occurred. There was a strong ordering in the results, with the lower-ranked strategies finding more positive changes over time and favoring the group with more deaths (the treatment group), whereas strategies with higher ranks found negative changes over time and favored the group with fewer deaths.

The rank assigned to each strategy was defined by the relative importance of death, as compared with poor and excellent health, in the coding scheme. As can be seen in Table 1, in Strategies 1 and 2, deaths are simply discarded and therefore are given no importance. In Strategies 3 and 4, deaths are combined with other categories; therefore death is treated as one of several poor outcomes. In Strategies 5 and 6, death is given its own low code (0), but poor health has a code almost as low. Death therefore is treated as the worst outcome but not the only bad outcome. In Strategy 7, death is the worst outcome and poor health is much better than death (though not as good as excellent health). The codes for

Strategy 7 are not very different from those of Strategy 8, which gives all living subjects the same code.

Given these findings, which method should be used? We might consider the QWB as a standard, because it contains a value for death. In Tables 3 and 4, Strategy 5 for the QWB is the "true" QWB statistic. The *t* statistics for the true QWB were quite similar to those of Strategies 6 through 8, suggesting that these methods might be used when the QWB score is not available.

It is best to choose a strategy that agrees with the goals of the intervention. If the goal is to increase survival, then Strategy 8 (alive yes/no) is clearly preferable. If survival is most important but health of the survivors is also of concern, then Strategy 7 might be chosen. If health and death are more nearly equal, then Strategies 5 and 6 should be considered. If death is only one of many bad outcomes, one should choose Strategy 3 or 4. Finally, if death can be ignored, Strategy 2 might be chosen. Death, for instance, could be ignored in a study of a young and healthy population with a demonstrably benign experimental treatment, if there were few deaths and all were due to accidents or trauma. It is difficult to think of a situation in which Strategy 1 would be appropriate.

For both the HP/DP and the Day Care studies, the primary goal was to improve health, but death was an important outcome. For these programs, Strategy 6, the probability of being healthy, seemed best because it stresses both health and survival and because it seemed to agree with the true QWB rankings mentioned above. Strategy 5 seemed less desirable because of the arbitrary nature of assigning a 0 to the deaths, although its effectiveness was very similar to that of Strategy 6.

The probability of being healthy was estimated separately in the two studies. Table 1 shows that the scores for poor health were remarkably similar for the Day Care and HP/DP studies, but that the scores for higher levels of health were rather different. For a new study, probabilities could be calculated from that study's own data if the sample sizes were sufficient. Otherwise the values of Table 1 could be used, perhaps using the HP/DP weights if the new study population were relatively healthy and the Day Care weights if the study were to involve a less healthy population.

This study is not definitive, because many other strategies could have been studied. We restricted the strategies to those that provided unequivocal results (e.g., positive scores favor the treatment group) and that controlled for T0 health status, although the latter is not always needed in a randomized trial. We used *t*-tests to make results comparable for all strategies, but we might have used trend tests or the McNemar test for Strategy 4 (better, same, or worse). Chi-square and Mann–Whitney tests are reported elsewhere.[18] We might have used the QWB as a standard for recoding EVGFP; however, this can be done only in a study that measures both variables. Other data sets might have provided different results. However, the two we used were quite different, yet agreed very well. The simulation allowed us to examine the effect of more deaths on the strategies, expanding our results somewhat beyond the two data sets available.

A large number of the Day Care subjects were eliminated from this paper because they were alive but too sick to provide complete data at T1. Such people might have been assigned a category between "poor" and "death" according to strategies similar to those described in this paper.

In summary, we found that Strategies 1 and 2, the most commonly used strategies, resulted in a favorable bias toward the study with more deaths. Strategies 2 through 8 might be appropriate, depending on the relative importance of death and poor health in the intervention being studied. For the studies described here, we prefer Strategy 6, in which original data are recoded to the probability of being healthy at T1, and the assignment of a value of 0 to deaths, because this measure agrees with the goals of the study and agreed with the QWB measure when it was available.

References

1. Sullivan DF. A single index of morbidity and mortality. Health Services and Mental Health Administration Health Reports 1971;86:347.

2. Wilkins R, Adams OB. Health expectancy in Canada, late 1970s: Demographic, regional, and social dimensions. Am J Pub Health 1983;73:1073.

3. Robine JM, Mathers CED, Bucquet D. Distinguishing health expectancies and health-adjusted life expectancies from quality-adjusted life years. Am J Public Health 1993;83:797.

4. Patrick DL, Erickson P. Health Status and Health Policy: Quality of Life in Health Care Evaluation and Resource Allocation. New York: Oxford University Press, 1993.

5. Stanish WM, Gillings DB, Koch GG. An application of multivariate ratio methods for the analysis of a longitudinal clinical trial with missing data. Biometrics 1978;34:305.

6. Stram DO, Wei LJ, Ware JH. Analysis of repeated ordered categorical outcomes with possibly missing observations and time dependent covariates. Journal of the American Statistical Association 1988;83:631.

7. Laird NM. Missing data in longitudinal studies. Statistics in Medicine 1988;7:305.

8. Brown CH. Protecting against nonrandomly missing data in longitudinal studies. Biometrics 1990;46:143.

9. Woolson R, Clarke W, Leeper J. Missing data in longitudinal studies. Statistical Models for Longitudinal Studies 1991:277.

10. Zwinderman AH. Statistical analysis of longitudinal quality of life data with missing measurements. Quality of Life Research 1992;1:219.

11. Dawson JD, Lagakos SW. Analyzing laboratory marker changes in AIDS clinical trials. J Acquir Immune Defic Syndr 1991;4:667.

12. Rubin DB. Inference and missing data. Biometrika 1976;63:581.

13. Bergner M, Bobbitt RA, Carter W, et al. The sickness impact profile: Development and final revision of a health status measure. Med Care 1981;19:787.

14. Kaplan RM, Anderson J, Wu A, et al. The quality of well-being scale: Applications in AIDS, cystic fibrosis, and arthritis. Med Care 1989;27(Suppl):S27.

15. Rothman ML, Diehr P, Hedrick S, et al. Effects of contract adult day health care on health outcomes and satisfaction with care. Med Care 1993;31:SS75.

16. Durham M, Beresford S, Diehr P, et al. Participation of higher users in a randomized trial of Medicare reimbursement for preventive services. Gerontologist 1991;31:603.

17. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: Statistics and Strategies for evaluation. Controlled Clin Trials 1981;12:142S.

18. Diehr P, Patrick D, Grembowski D, et al. Accounting for deaths when measuring health status over time. In: Proceedings of the 1993 Public Health Conference on Records and Statistics. Rockville, MD: U.S. Department of Health and Human Services; 1993.