

University of Washington

From the Selected Works of Paula Diehr

June, 1999

Survival versus years of healthy life; which is more powerful as a study outcome?

Paula Diehr, *University of Washington*

Survival Versus Years of Healthy Life: Which Is More Powerful as a Study Outcome?

Paula Diehr, PhD, Donald L. Patrick, PhD, MSPH,

Gregory L. Burke, MD, and Jeff Williamson, MD

Department of Biostatistics (P.D.) and Health Services (P.D., D.L.P.), University of Washington, Seattle, Washington, USA; and Department of Public Health Science, School of Medicine (G.L.B.) and the J. Paul Sticht Center on Aging (J.W.), Wake Forest University, Winston-Salem, North Carolina, USA

ABSTRACT: Studies of interventions that are intended to improve patients' health are often evaluated with survival as the primary outcome, even when a measure adjusted for quality of survival, such as years of healthy life (YHL), would seem more appropriate. The purpose of this article is to determine whether studies based on survival are more or less powerful than studies based on YHL in clinical trials where either measure might be appropriate. We used data from the Cardiovascular Health Study (CHS) to estimate the sample size that would be needed in studies of 156 different health conditions, for the two outcome measures. The median sample size for a 5-year study was 687 if survival was the endpoint and 484 for YHL. YHL usually required lower sample sizes than survival, although survival was more powerful for some health conditions. We also found that lengthy studies, and studies with many follow-up measures per person, did not have appreciably higher power than less intensive studies. We conclude that clinical investigations in which the goal is to improve health may often be performed more efficiently with YHL rather than survival as the primary outcome measure. Such studies can be short in duration, with relatively few measures per person of health status. *Control Clin Trials* 1999;20:267-279 © Elsevier Science Inc. 1999

KEY WORDS: *Aged, healthy life expectancy, QALY, clinical trials, survival, health status, power*

INTRODUCTION

Investigators often design clinical trials and observational cohort studies using survival as the primary outcome, even when the major goal of the intervention in question is to improve patients' health. Years of healthy life (YHL) would seem a more appropriate outcome measure for such studies, because it incorporates both health and mortality [1]; however, except in some preliminary investigations [2], it has seldom been used as the primary outcome.

The best primary outcome measure for a particular study is the one that fits the study's goals. If preventing mortality is the major goal, then survival is

*Address reprint request to Dr. P. Diehr, Department of Biostatistics Box 357232, University of Washington, Seattle, WA 98195. E-mail: pdiehr@biostat.washington.edu.
Received April 30, 1998; accepted November 23, 1998.*

most appropriate; if improving health is the major goal, then both survival and years of healthy life could be reasonable outcome measures. Given the choice of two appropriate outcome measures, investigators usually prefer the outcome that permits a study with lower sample size, or with greater power for the same sample size. In this study, we aim to compare YHL to survival in a variety of situations in which either measure might be appropriate, to find out which outcome measure is more powerful.

Suppose we were designing a clinical trial of an intervention to prevent or cure a health condition such as angina. We might first locate existing data from a study that measured health status in persons with and without the condition. We could use these data to estimate how large a difference in outcome we could expect in the two groups, if the treatment were successful. We could then use this expected difference to calculate the necessary sample size for the new study to achieve 80% power. In what follows, we show that such an angina study would require about 300 subjects per group if survival were the endpoint, but only 140 per group if we used YHL. We might prefer to use YHL as the primary outcome measure for the new angina study, because it would allow us to conduct an adequately powered study with fewer patients.

We used this approach to compare the power of survival and years of healthy life for 156 different "health conditions," such as angina, in a large cohort of older adults. We considered the 156 health conditions to be a sample from the population of all health conditions that might be studied. Our goal was to determine, for each health condition, whether survival or YHL was the more powerful outcome measure. Because the health conditions include a wide variety of patient descriptors and problems, we hope to provide general recommendations for studies of older adults.

We calculated the required sample sizes for each outcome measure with each health condition. We also examined whether 5 years of follow-up were more powerful than 2.5 years, and how much power we could gain from obtaining additional follow-up measurements per person.

METHODS

Data

We used data from the Cardiovascular Health Study (CHS), a population-based longitudinal study of 5201 adults ≥ 65 years of age designed to identify factors related to the occurrence of coronary heart disease and stroke [3, 4]. We recruited CHS subjects from a random sample of the HCFA Medicare eligibility lists in four communities in the United States. Persons eligible to participate were not institutionalized at baseline and were expected to remain in the area for the next 3 years. We excluded persons who needed to use wheelchairs at home or were receiving hospice treatment, radiation therapy, or chemotherapy for cancer at baseline. At baseline, and every 6 months thereafter, we asked subjects to rate their health status as excellent, very good, good, fair, or poor. This scale is a commonly used simple measure of health status [1]. Health data for the year 0.0 (baseline) came from a home interview; data from years 1.0, 3.0, 4.0, and 5.0 from a mailed survey; and data from years 0.5, 1.5, 2.5, 3.5, and 4.5 from a telephone interview. We did not ascertain health status

at year 2.0. To ensure that participants were reasonably similar at baseline, we considered only the 1909 persons who reported their health as “good” at baseline. We shall describe methods for a study with 5-year follow-up. The methods for 2.5-year follow-up are similar.

Health Outcome Measures

The primary goal was to compare the power of survival to YHL as a study outcome. We considered two survival outcomes, 5-year survival and 5-year survival without stroke, myocardial infarction (MI), or death (referred to as “bad events”) [4]. Persons without an event were censored at 5 years.

We considered two health status measures, both based on self-reported health status (EVGF): “Would you say that your health is excellent, very good, good, fair, or poor?” EVGF, a simple but well-known measure that has received detailed study [5, 6], has been found to be predictive of future health events in many studies [7]. Because we were examining health status over time, we added a sixth health state, death.

We recoded EVGF in two ways so that we could interpret its mean value. The first method recoded EVGF as “healthy (yes/no),” with “excellent,” “very good,” and “good” coded as 100 and “fair,” “poor,” and “death” coded as zero. The mean of “healthy” is thus the percent of persons who were healthy at that time. A second measure, $\text{prob}(\text{healthy})$, uses values suggested elsewhere [8], in which “excellent” is coded as 96, “very good” as 93, “good” as 76, “fair” as 35, “poor” as 19, and “death” as zero. These weights are approximately the percent probability that the person will be “healthy” 2 years later.

We interpolated missing values of health status linearly whenever a person had a valid value before and after the missing value. Because death has a valid value of zero, only 3.8% of the subjects had missing YHL after interpolation. We excluded those subjects.

Survival and YHL

Survival in the first 5 years of the study and time without a bad event (MI, stroke, or death) are straightforward measures. About 87% of the subjects were alive at 5 years, and 79% did not have a bad event in 5 years.

YHL values were calculated as follows. Suppose that health status is measured at time zero and at K later times, covering the period from time t_0 to t_K . If the mean health status at each time is plotted against time, the area under that curve is the average person-years of healthy life that the population experienced. In our notation, at times t_0, t_1, \dots, t_K , a person’s measured health status values are h_0, h_1, \dots, h_K . If the measurements occur at regular intervals, and $t_0 = 0$, the simplest estimate of the area under the curve (YHL) comes from a trapezoidal or “connect-the-dots” strategy [9], with:

$$\text{YHL} = \frac{(h_0 + h_K + 2 \sum_{j=1}^{K-1} h_j) t_K}{2K} \quad (1)$$

For example, this study followed the patients for 5 years, with health status

measured every 6 months; thus, $K = 10$, $t_K = 5$, and the estimated YHL is $[h_0 + h_{10} + 2(h_1 + h_2 + \dots + h_9)]/4$, with a maximum of 5 years of healthy life. Subjects averaged 3.74 and 3.48 years of healthy life based on “healthy (yes/no)” and “prob(healthy),” respectively.

The survival and the YHL measures are positively correlated. The estimated correlation between the two measures of survival and YHL are 0.76, and 0.96, respectively, the correlations between the survival measures and the YHL measures range from 0.49 to 0.63.

Health Conditions

We ascertained a large number of health conditions (such as angina) at baseline for each CHS participant. The data include self-reported demographics, medical history, and health behaviors, as well as a variety of clinical and laboratory measures. There are many detailed risk factors for cardiovascular disease. We show a few of the conditions in Table 2. Although the emphasis of CHS is to find correlates of cardiovascular disease, we collected many general demographic, medical, social, and behavioral variables, such as education, asthma, depression, cognitive status, and social support.

About half of the health condition variables were initially dichotomous. We dichotomized the remaining variables as being high (above the mean) or low (below the mean). For example, we dichotomized age as high (≥ 73) or low (< 73). Of the original 260 conditions, we retained only those that were significantly correlated with at least one of the survival or YHL measures calculated from 2.5 years of follow-up. We further eliminated variables that were calculated from baseline health status, or that were perfectly correlated with other variables. Finally, we eliminated variables in which the high or low category had fewer than 50 persons.

This process left 156 health conditions for analysis. Some, such as patient age or gender, are not usually thought of as health conditions. Here, however, we used such variables to represent health conditions associated with health or mortality differences that are “similar to” differences between persons above and below age 73, or between men and women.

We studied all 156 conditions together, as well as several subsets of conditions. Two of us (J.W. and D.L.P.) independently categorized the 153 health conditions into those likely to result in death versus those primarily associated with a change in health status short of death. We also classified health conditions as being self-reported versus diagnosed in clinics or laboratories, and we looked separately at well-known cardiovascular risk factors.

Required Sample Size

In the familiar sample size calculation, the number of cases needed in each group for a study with 80% power is:

$$N = \frac{(1.96 + 0.84)^2 (2 s^2)}{d^2} \quad (2)$$

where d is the expected difference in outcome in the two groups and s is the

sample standard deviation of the outcome measure. We calculated the required sample size for each health condition with each health measure.

Investigators usually use survival analysis to describe time-to-event. In preliminary analyses, we found that the square root of the log-rank statistic from survival analysis was almost perfectly correlated ($r = 0.99$) with a t -statistic in which the dependent variable was “dead or alive at year 5.” For this reason, we chose “dead or alive at 5 years” and “bad event (yes/no)” as the two survival measures in our study, and we used the method of Eq. (2) to calculate the required sample size.

Analysis

We calculated the required sample size for various health events and outcome measures, and we present the median and 75th percentile of the required sample size overall and within groups of health conditions for both 2.5-year and 5-year studies. Although our primary aim has been descriptive, we also performed Wilcoxon rank-sum tests to determine whether one measure required significantly larger sample sizes than another. Because the independence assumption for the test is probably not met, because some of the health conditions are highly correlated, we use the results of this test descriptively.

FINDINGS

We first present the angina example mentioned earlier in some detail, to familiarize the reader with the type of calculations involved. We then show the results for all 156 health conditions, and for subsets of conditions. We show results for 2.5- and 5-year studies, and for one, two, five, and ten follow-up measures.

Angina Example

Suppose we are designing a 5-year randomized controlled trial of older adults to determine the efficacy of a new drug that was expected to cure angina, using the CHS data to plan the size of the new study. We restrict the analysis to subjects whose health was “good” at baseline, so that the subjects will be comparable at baseline. Two possible outcomes are survival and YHL. We assume that the subjects without angina at baseline are similar to the treatment group (if the drug works) and that those with angina are similar to the control group.

The first column of Table 1 shows that, on average, 89.3% of the 1528 persons without angina are alive 5 years later, compared with 81.9% for the 381 with angina, a difference of 7.4 percentage points in survival. This is the difference that we could expect in the new trial if the drug works as intended. The standard deviation of “dead/alive at 5 years” is 32.7. If we use Eq. (2), a study with 80% power to determine differences in 5-year survival between patients with and without angina would require 307 patients per group.

We next repeat the sample size calculation using YHL as the endpoint. Figure 1 shows a plot of mean health status [based on $\text{prob}(\text{healthy})$] over time for persons with and without angina. The areas under the two curves are 3.27

Table 1 Sample Size Calculation for Angina Example Survival Versus YHL as the Outcome

Angina	Outcome Measure	
	Survival ^a	YHL ^b
No		
Mean	89.3%	3.58
N	1528	1528
Yes		
Mean	81.9%	3.27
N	381	381
d^c	7.4	0.30
s^d	32.7	0.91
N^e	307	139

^aFive-year required sample size for “survival.”

^bFive-year required sample size based on “prob(healthy).”

^cDifference in means.

^dStandard deviation of measure for both groups together.

^e $N = (1.96 + 0.84)^2 (2) (s^2)/d^2 =$ required sample size for 80% power.

and 3.58 years of healthy life, respectively. The second column of Table 1 shows that we could conduct the angina study with only 139 persons per group if YHL were the endpoint.

Other things being equal, an investigator would prefer YHL as an outcome measure for studies of angina, or in other situations in which the two groups

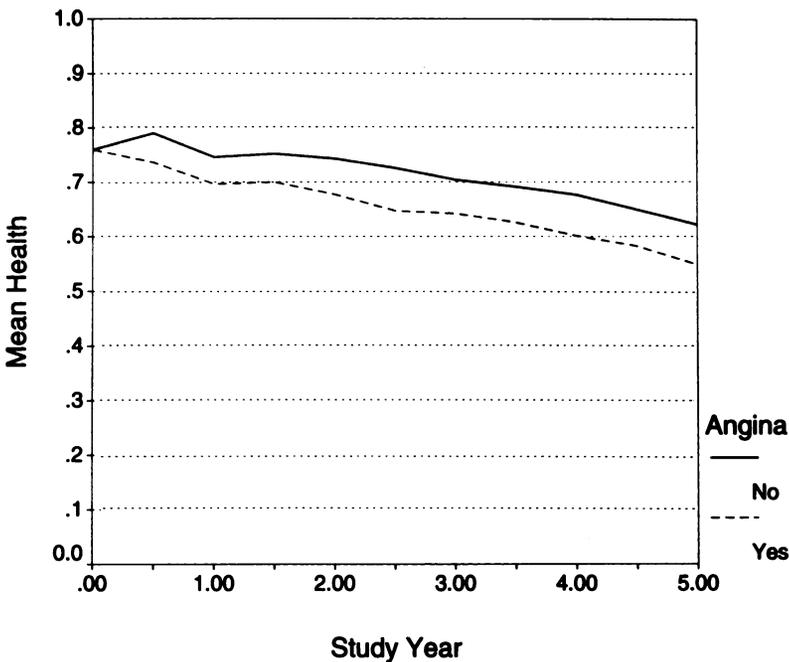


Figure 1 Health over time by BL angina status. YHL: no angina, 3.58 years; angina, 3.27 years.

Table 2 Description and Required Sample Size for Selected Health Conditions

Variable	% High	N	Required Sample Size	
			Survival	YHL
Minimum ankle : arm ratio	55.6	1873	361	233
Age at baseline	44.5	1909	118	123
Cigarettes smoked per day	43.1	1851	739	3139
Angina status at baseline	20.0	1909	307	139
Any diuretic at baseline	26.6	1908	700	588
Aortic regurgitation	20.3	1898	200	237
Arthritis	53.6	1887	6774	591
Atrial fibrillation : self-reported	5.3	1417	73	155
Dizziness in the past year	24.0	1900	7928	257
Heart rate (30 seconds)	50.1	1905	3024	515
Benzodiazepines at baseline	8.5	1908	12,194	131
Bioelectrical impedance reactance	48.1	1882	326	624
Bioelectrical impedance resistance	46.8	1880	638	1082
Bleed or bruise easily	35.0	1897	4898	2022
Blocks walked per week	30.3	1890	3447	328
Brachial pressure for AAI	46.3	1887	394	536
Stoppage of breathing during sleep	8.9	1633	189	172
Short of breath when hurrying	38.8	1902	1323	151

of interest were expected to differ by about as much as persons with angina differ from those without it. To determine whether these findings generalize to other situations, we repeated these calculations for the remaining health conditions. We consider the other conditions to be a sample from the population of all conditions that might be studied in an RCT.

Analysis of All Health Conditions

For brevity, we forbear listing all 156 health conditions, but we show in Table 2 the conditions whose variable names begin with "A" or "B," along with the percentage of persons who were in the higher category and the number of persons involved. Table 2 illustrates that the number of persons per condition could vary. Also, note that some variables are related, such as bioelectric impedance reactance and resistance. The health conditions are not independent, and we are using them in a descriptive manner only.

We also show in Table 2 the calculated required sample sizes. For example, as seen in Table 1 angina has required sample sizes of 307 for survival and 139 for YHL. A study comparing persons with and without atrial fibrillation, however, would have more power if survival were the endpoint. Of the health conditions shown in Table 2, 11 favored YHL and 7 favored survival.

Survival Versus YHL

Some of the sample sizes are extremely large, making the mean sample size a misleading statistic. Therefore, Table 3 shows the median of the required sample sizes for the 156 health conditions, for the four outcome measures. The

Table 3 Comparison of Four Outcome Measures on Median Required Sample Size

Study Length (years)	Statistic	Survival	Bad Event	YHL	
				Healthy(Yes/No)	Prob(Healthy)
5	Median	483	464	348	356
2.5		960	843	532	433
5	75th percentile	1864	1714	731	693
2.5		2389	4617	1155	1142

Wilcoxon signed-rank test results: for both 5 years and 2.5 years, prob(healthy) is significantly better than the other three measures, and healthy(yes/no) is significantly better than "survival" and "bad event."

first line of the table gives results for a 5-year study. The median required sample size for survival was 483, and for survival without a bad event (MI, stroke, or death) the median required sample size was 464. The two YHL measures had lower median required sample sizes than survival, 348 and 356, respectively. The second line of Table 3 shows the required sample sizes if the study lasts only 2.5 years, instead of 5. The YHL measures again have smaller required sample sizes than the survival measures. The 75th percentiles of the sample sizes, shown on lines 3 and 4, demonstrate a similar pattern. The Wilcoxon rank-sum test shows that YHL based on prob(healthy) required significantly smaller sample sizes than the other three measures for both study lengths. YHL based on healthy(yes/no) required significantly smaller sizes than the two survival measures for both study lengths.

The required sample sizes are larger for a 2.5-year study than for the 5-year study, but they are not twice as large, as we might have expected for a study lasting half as long. The differences by study length are always statistically significant by the Wilcoxon rank-sum test.

Condition-Specific Results

We also looked at specific health conditions in which the required sample sizes of survival and YHL were quite different. For 22 health conditions, survival required half the sample size as YHL, based on prob(healthy); YHL required half the size for 54 variables. There was no obvious rule as to which health conditions favored which outcomes.

We next compared the necessary sample sizes within the groups of health conditions prespecified by the investigators (Table 4). When raters 1 and 2 expected survival to be the better outcome measure (lines 1 and 3), all four of the measures had low required sample sizes, and prob(healthy) was often significantly better than healthy(yes/no). This pattern also held true for the laboratory and clinical findings, and for the cardiovascular risk factors.

For the health conditions in which we expected YHL to perform better (lines 2 and 4), and for self-reported health conditions (line 6), larger sample sizes were usually required as compared with the previously mentioned subgroups. There were large differences among the outcome measures. The two YHL measures were always significantly better than the two survival measures, with median sample sizes about half as large. In addition, for a 2.5-year study (not shown), survival was significantly better than bad event, and prob(healthy) was significantly better than healthy(yes/no).

Table 4 Median Sample Size in Subsets of Health Conditions (5 Years)^a

Rater Subset	Survival	Bad Event	YHL		N	Significant Differences
			Healthy(yes/no)	Prob(Healthy)		
1 Survival better	235	185	160	171	44	b
YHL Better	759	1063	446	421	48	c
2 Survival better	389	315	384	372	48	d
YHL better	643	608	263	295	68	c
3 Exam or tests	378	318	416	372	70	b
Self-reported	693	672	272	320	86	c
4 Cardiovascular	211	135	240	208	18	b

^aHealth conditions were divided into several groups by investigators. See text. Wilcoxon rank-sum results: ^bprob(healthy) significantly better than Healthy(yes/no); ^ctwo YHL measures significantly better than two survival measures; ^d“bad event” significantly better than “survival.”

Table 5 Median Sample Size by Number of Follow-Up Measures (K) (5-Year Study)

K	YHL Based on:	
	Healthy(Yes/No)	Prob(Healthy)
1	474	373
2	456	444
5	358	357
10	347	355

Wilcoxon signed-rank sum test results: for both measures, ten follow-ups are significantly better than one or two.

Number of Follow-Up Measures (K)

We also studied the relationship of number of follow-up measures to the required sample size. Table 5 shows the required sample sizes for the two YHL measures using one, two, five, or ten follow-up measures over 5 years. Median required sample size decreased only slightly as the number of follow-up measures increased, especially for YHL based on prob(healthy). For healthy(yes/no), ten follow-ups were significantly better than the other three choices. For prob(healthy), ten follow-ups were significantly better than one or two.

SUMMARY AND DISCUSSION

We shall summarize and assess findings on the following topics: survival versus YHL; variants of survival and YHL; length of study; and number of follow-up measures.

Survival Versus YHL

Table 3 shows that YHL measures generally have a lower required sample size than survival measures. If we compare survival and prob(healthy), YHL is better than survival for about two thirds of the health conditions; YHL is twice as good as survival for 54 conditions, whereas survival is twice as good for only 22. There are some conditions for which survival is better than YHL, but they are in the minority. YHL was substantially better than survival for conditions that the raters classified as likely to have more effect on health status than on mortality, and about the same when the condition was expected to be related more strongly to survival. This means that a study whose goal is to improve health will usually not lose power by using YHL instead of survival as the primary endpoint, and will very likely improve power.

Variants of Survival and YHL

Survival without a "bad event" was usually similar in power to survival alone. The only significant differences appeared in the subsets (Table 4), and the direction of the difference was not consistent.

The two YHL measures had similar median performances but healthy(yes/no) was often significantly less powerful than prob(healthy) by the Wilcoxon

rank-sum test. This apparent discrepancy occurred because the distribution of sample sizes based on healthy(yes/no) has a longer right tail than that for prob(healthy), as shown in the 75th percentiles in Table 3.

YHL based on healthy(yes/no) is easy to interpret, as it is the number of person-years spent in excellent, very good, or good health. YHL based on prob(healthy) is somewhat harder to interpret. If the codes used were each person's exact probability of being healthy in 2 years rather than a crude overall value assigned without regard to age and gender, we would interpret this YHL as years of healthy life in the future starting 2 years after baseline, rather than starting at baseline. This interpretation is difficult to understand. On the other hand, prob(healthy) has better face validity than healthy(yes/no) in that it assigns different health status to subjects in, say, fair health and death. The interpretation for the measure is probably adequate. Prob(healthy) was the best outcome measure of the four examined.

Length of Study

Table 3 shows that the required sample size of the two survival measures is somewhat higher in a 2.5-year study than in a 5-year study. In analyses comparing 2.5 to 5 years (not shown), we found that the 2.5-year study required at least twice as many subjects 44–47% of the time when survival was the outcome measure, but only 31–35% of the time when YHL was the outcome measure. Thus, the length of the study is less important for YHL than for survival.

Table 3 also shows that, on average, a 5-year study with survival as the outcome requires about the same sample size as a 2.5-year study using YHL as the outcome. Choosing a different outcome measure can permit a shorter study with the same power. Even for the same outcome measure, these data may provide an argument for shorter studies when long-term follow-up is not the primary concern. For YHL calculated from prob(healthy), for example, the median study with 80% power would require 356 subjects per group to be followed for 5 years (1780 person-years), compared with 433 subjects followed for 2.5 years (1083 person-years). If recruitment is not a problem, a 2.5-year study of 433 subjects could be much less expensive than a 5-year study of 356 subjects. YHL studies can and should be short.

Number of Follow-Up Measures

Table 5 shows clearly that for YHL the improvement in sample size achieved by making additional measurements per person is small. Diehr et al. showed that additional follow-up measurements per person provide very little improvement in power if the correlations among measures are high [9]. Here, correlations ranged from about 0.4 for measures 5 years apart to 0.7 for measures 6 months apart, explaining in part why additional measures provided little additional information or power. In most situations, the fact that the number of health status measurements per person need not be large could permit substantial savings as compared with a study making more interim measures. There are other reasons, such as maintaining contact, for surveying subjects more frequently, but power is not among them.

Which Measure for Which Health Conditions?

Survival measures did not perform significantly better than the YHL measures for any of the subgroups of health conditions that we defined in advance. All measures had similar performance for conditions in which survival was expected to be better, and the YHL measures were far superior to the survival measures for the remaining health conditions. Especially for the shorter, 2.5-year study, using YHL based on $\text{prob}(\text{healthy})$ seems to be the wisest choice unless the investigator is confident that the health condition affects only mortality.

Limitations

The findings may be specific to the health conditions that were available for study. We believe that the health conditions were varied enough to represent many comparisons of interest in future clinical trials. Because the purpose of the CHS was to measure cardiovascular risk factors, this available set of health conditions may have included more serious outcomes; if so, then YHL would be an even more useful outcome measure among a less serious set of health conditions. If relevant pilot data are available, however, investigators should base design choices on those data instead of on our general findings.

We used only a very simple health status indicator. EVGFP may not perform as well as a condition-specific measure, and it does not distinguish among the various dimensions of health status. A more sophisticated or appropriate measure of YHL would have probably led to favoring YHL over survival even more strongly.

Another potential concern is that this study used data from older adults who were relatively healthy at baseline and may not generalize to all situations of interest, such as comparing two groups of very sick children. Data for persons <65 years is likely to be more favorable to YHL than we have found here, because death would be even less common. It is also possible that changes in health among sick children are similar to changes in health for older adults. This issue requires further study.

We did not adjust for baseline differences between the two health condition groups, except to require that all subjects have "good" health at baseline. A more careful sample size calculation or analysis could control for age, gender, and other important baseline health conditions. For this reason, the required sample sizes that we calculated here are probably smaller than what a particular study actually needs. The major goal of this study has been to compare YHL to survival measures. In this comparison, each person is essentially his own control, meaning that biases in required sample size due to not controlling for all baseline characteristics will tend to cancel out. We did not control for more baseline factors, because we preferred to use such potential covariates as "health conditions" for this exercise. The comparisons of survival with YHL should be substantially correct, even if the estimates of the required sample sizes are low.

CONCLUSIONS

In designing a study, one should choose an outcome measure that fits its goals. If one wants to compare health in two groups, perhaps after an intervention, then using YHL as the primary endpoint is likely to result in a study

more efficient than one using survival. Investigators may have avoided basing studies on YHL in the past because its characteristics were unknown; one goal of this study has been to make these characteristics better understood. Other past reservations might have included concern about the difficulty of obtaining multiple measurements of health status over time. We have shown that trials with YHL as an outcome can often be shorter than those based on survival, and that the number of follow-up measurements need not be large. In the future, investigators should consider using YHL as the primary outcome measurement.

This work was supported by contracts N01-HC-85079 through N01-HC85086 from the National Heart, Lung, and Blood Institute.

REFERENCES

1. Patrick DL, Erickson P. *Health Status and Health Policy: Allocating Resources to Health Care*. New York: Oxford University Press; 1993.
2. Rawles J, Light J, Watt M. Quality of life in the first 100 days after suspected acute myocardial infarction—a suitable trial endpoint? *J Epidemiol Commun Health* 1992; 46:612–616.
3. Fried L, Near O, Enright P, et al. The cardiovascular health study: Design and rationale. *Ann Epidemiol* 1991;1:250–276.
4. Ives G, Fitzpatrick A, Bild D, et al. Surveillance and ascertainment of cardiovascular events: The cardiovascular health study. *Ann Epidemiol* 1995;5:278–285.
5. Krause NM, Jay GM. What do global self-rated health items measure? *Med Care* 1994;32:930–942.
6. Gold M, Franks P, Erickson P. Assessing the health of the nation: The predictive validity of a preference-based measure and self-rated health. *Med Care* 1996;34: 163–177.
7. Idler EL, Benyamini Y. Self-rated health and mortality: A review of twenty-seven community studies. *J Health Soc Behav* 1997;38:21–37.
8. Diehr P, Patrick D, Hedrick S, et al. Including deaths when measuring health status over time. *Med Care* 1995;33:AS164–AS172.
9. Diehr PK, Patrick DL, Psaty B. Required sample size and power for clinical trials that measure years of healthy life. *Stat Med* 1997;16:1211–1223.