10-22-2012

Decline in Health for Older Adults: 5-Year Change in 13 Key Measures of Standardized Health

Paula H. Diehr  
*University of Washington, pdiehr@u.washington.edu*

Stephen M. Thielke  
*University of Washington - Seattle Campus, sthielke@u.washington.edu*

Anne B. Newman  
*University of Pittsburgh, anewman+@pitt.edu*

Calvin H. Hirsch  
*University of California, Davis, chhirsch@ucdavis.edu*

Russell Tracy  
*University of Vermont, russell.tracy@uvm.edu*

Suggested Citation

http://biostats.bepress.com/uwbiostat/paper385

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.  
Copyright © 2011 by the authors
Decline in health for older adults:

5-year change in 13 key measures of standardized health.

Diehr PH¹,², Thielke SM³,⁴, Newman AB⁵, Hirsch C,⁶ Tracy R⁷

From the University of Washington departments of (1) Biostatistics, (2) Health Services, (3) Psychiatry and Behavioral Sciences, Seattle, Washington, USA; (4) Geriatrics Research, Education, and Clinical Center, Puget Sound VA Medical Center; (5) Department of Epidemiology and Center for Aging and Population Research, University of Pittsburgh, Pittsburgh, PA 15213, USA; (7) University of Vermont College of Medicine, Burlington, VT

Corresponding Author Paula Diehr, pdiehr@uw.edu, box 34922, University of Washington, Seattle, WA, 98195

Key words: aging, hospitalization, bed days, cognition, extremity strength, feelings about life as a whole, satisfaction with the purpose of life, self-rated health, depression, digit symbol substitution test, grip strength, ADLs, IADLs, and gait speed

Word count: abstract 246/250; MS. 4534; 4 tables, 2 figures

Running Title: Change in 13 measures of health
Decline in health for older adults:

5-year change in 13 key measures of standardized health.

Abstract

Introduction
The health of older adults declines over time, but there are many ways of measuring health. We examined whether all measures declined at the same rate, or whether some aspects of health were less sensitive to aging than others.

Methods
We compared the decline in 13 measures of physical, mental, and functional health from the Cardiovascular Health Study: hospitalization, bed days, cognition, extremity strength, feelings about life as a whole, satisfaction with the purpose of life, self-rated health, depression, digit symbol substitution test, grip strength, ADLs, IADLs, and gait speed. Each measure was standardized against self-rated health. We compared the 5-year change to see which of the 13 measures declined the fastest and the slowest.

Results
The 5-year change in standardized health varied from a decline of 12 points (out of 100) for hospitalization to a decline of 17 points for gait speed. In most comparisons, standardized health from hospitalization and bed days declined the least while health measured by ADLs, IADLs, and gait speed declined the most. These rankings were independent of age, sex, mortality patterns, and the method of standardization.

Discussion
All of the health variables declined, on average, with advancing age, but at significantly different rates. Standardized measures of mental health, cognition, quality of life and hospital utilization did not decline as fast as gait speed, ADLs, and IADLs. Public health interventions to address problems with gait speed, ADLs, and IADLs may help older adults to remain healthier in all dimensions.
1.0 Introduction

On average, the health of older adults declines with age, usually more steeply near to the time of death. But there are many different aspects of health, which may decline on different schedules. Donald Kennedy described this issue as follows:

“Oliver Wendell Holmes provided one metaphor for the perfect life-span in his poem "The Deacon's Masterpiece Or, the Wonderful One-Hoss Shay: A Logical Story." Built of carefully selected parts that the builder thought would wear out but not break down, it lasted exactly a hundred years in good condition. Then, the Wonderful One-Hoss Shay collapsed into a mound of dust, going to pieces "...all at once, and nothing first--just as bubbles do when they burst." The shay's life cycle would be an attractive metaphor for us humans if the span were long enough. Alas, those of us at a Certain Age are all too acutely conscious of differential wear-out. [emphasis PD]. As Roth et al. point out in exploring the similarities between aging in humans and rhesus monkeys, there is a canonical sequence: presbyopia, cataracts, loss of motor activity, decline in memory performance. It would be nice if these things happened all at once instead of sequentially--as long as it wasn't too soon!”

The goal of “squaring out the mortality curve” involves sustaining health across multiple domains until a time close to death. It may be possible to select or tailor interventions likely to improve the domains that are most susceptible to decline, with the goal that a person’s health would more nearly fall to pieces all at once, thus sustaining individuals’ functional life expectancy. If consistent patterns of decline are observed across populations, then decline in certain domains could identify aging adults who are earlier in the decline process, and could be potentially valuable targets for interventions.

In this paper we compared the 5-year change in 13 health variables that encompassed multiple domains of health. These domains include:

1-Functional Health: (which was measured by) gait speed; self-rated extremity strength; measured grip strength; activities of daily living (ADLs).

2-Mental Health: Center for Epidemiologic Studies Short Depression score
3-Cognition: modified mini mental state examination (3MSE); digit symbol substitution test (DSST)

4-Quality of life: feelings about life as a whole; satisfaction with the purpose of life (10-point scale).

5-Over-all health and function: self-rated health; instrumental activities of daily living (IADLs).

6-“Freedom”: not being hospitalized in the previous year; not being confined to bed because of illness or injury. (Category 6 was created only after initial findings were available – we had originally classified hospitalization and bed days as measures of Functional Health).

We evaluated several hypotheses. (1) Different measures of health will decline at different rates. (2) Functional Health will decline fastest. (3) Decline will differ by age and sex, with women and younger persons declining the least because of their lower mortality. (4) The rankings of decline among the variables will be independent of age and sex. (5) Change over time will be different for self-rated versus objectively observed items. (6) Decline within a domain will be more similar than decline across domains. (7) We also expected that the rankings of change would be the same under an alternate method of standardization.

2.0 Methods

2.1 Data

Data came from the Cardiovascular Health Study (CHS), a population-based longitudinal study of risk factors for heart disease and stroke in 5888 adults aged 65 and older at baseline. Participants were recruited from a random sample of Medicare eligible persons in four U.S. communities, and extensive data were collected during annual clinic visits and telephone calls. The original cohort of 5201 participants,
recruited in about 1990, had up to ten annual clinic examinations. A second cohort of 687 African Americans, from 3 of the original study communities, were enrolled in about 1993 and had up to seven annual examinations. Follow-up is on-going for mortality.

Table 1 gives the abbreviations and full names of the 13 variables used here, which are common measures of health used in aging research. All data, from 1990 to 1999, were used to create the standardized variables (see section 2.2). For analysis we used years 1991 to 1996 from cohort 1, and years 1994 to 1999 for cohort 2. The baseline year was excluded to decrease effects of selection bias and regression to the mean after enrollment. The study involves the 5,688 persons who were alive one year after baseline (referred to as year 1 in the current study) and had at least one observation on each variable.

[Table 1 about here]

2.2 Standardization

A major methodologic challenge in comparing change across different variables is that they are not measured on the same scale. For example, the 3MSE is scored from 0 to 100, while ADL is scored from 0 to 6. How would a ten-point decline in the 3MSE correspond to a new ADL difficulty? Further, many of these variables are on ordinal scales, meaning that the difference between two levels does not have a consistent interpretation – a decline of 10 3MSE points may have different interpretation if the person changes from 100 to 90 versus from 70 to 60. Finally, the measures are not defined after the subject has died, and are usually treated as missing instead.

To deal with these difficulties, we standardized each of the 13 variables on a 100-point scale, using self-rated health as a reference. The self-rated health item asked each individual if his health was Excellent, Very Good, Good, Fair, or Poor. (This variable is referred to from here...
on as EVGGFP). We standardized the variables by transforming them all to the “% probability of being healthy”, where “healthy” is defined as EVGGFP being excellent, very good, or good (EVGG), rather than fair or poor. That is, we replaced each original value with the % of persons at that value who were EVGG. The third column in Table 1 gives examples of the standardization for each variable. For hospitalization in the previous year, having no hospitalizations was coded as 76, and having one or more hospitalization was coded as 55; these values were used because 76% of the persons who were not hospitalized reported their health as EVGG, but only 55% of those who had been hospitalized one or more times were EVGG. The second row shows that 76% of persons with no days in bed in the previous 2 weeks were EVGG (standardized score = 76), but only 18% of those who were in bed the entire 14 days were EVGG (standardized score = 14).

The entire dataset (all persons, all years) were used for standardization. EVGG, a binary variable, was set to 1 for Excellent, Very Good, or Good and 0 for Fair or Poor. The transformed values were estimated by a logistic regression of EVGG on the logarithm of the variable of interest. (We added 1 before taking the logarithm because for some measures 0 was a valid value. For the 3MMS we used the logarithm of 101-3MMS because that variable was negatively skewed). The estimated probabilities (multiplied by 100) were used as the standardized values for each variable. Note that the estimates (say, for ADL at a particular time) depend only on the person’s ADL value at that time, not on their EVGGFP). The standardization values did not depend on the study year.

The resulting standardized variables are all on the same scale (representing the % of persons expected to be EVGG). Standardized health has the property of being on an interval/ratio scale, so that a change of a certain number of points has the same interpretation at
every initial level. And finally, because we may assume that dead persons are not EVGG, deaths can appropriately be coded as 0 on the standardized scale. These standardizations (aka transformations) have been described elsewhere for the SF-36 and EVGGFP, ADLs, bed days, blocks walked, BMI, depression, EVGGFP, hospitalization, IADLs, 3MSE, blood pressure, and gait speed. We chose EVGGFP as the standard because it had been used elsewhere. We could have standardized the health variables to some other measure of health, such as ADL difficulties. The only requirement is that the standard variable be monotonically related to all of the other variables.

Standardized health can be interpreted in several ways. Standardized ADL, for example, would be strictly interpreted as the probability that a person with a particular number of ADL difficulties would be in EVGG health in the large dataset used for estimation. But it can be more loosely thought of as “EVGGFP-standardized ADL” or “standardized health from ADL”. One disadvantage of the standardization approach is that EVGGFP itself can not be standardized in this way (it would take on only the values of 0 or 100). EVGGFP was instead transformed to the estimated probability of being EVGG one year later. [12]

A different standardization method was examined briefly, which is described below.

2.3 Outcome Measure

Our goal was to compare the 5-year change in the 13 standardized variables, to determine which declined fastest, and which remained relatively stable. The outcome measure was standardized health at year 6 minus the value at year 1, referred to here as the slope, and was calculated separately for each person for each variable. We further adjusted the standardized variables so they would start, on average, at the same point, to make it easier to compare the slopes.
2.4 Missing Data

Missing data were imputed, separately for each variable, by linear interpolation of the person’s own standardized data over time.\textsuperscript{15} \textsuperscript{16} Because death has a value, everyone who died before 2005 (the end of mortality f/u when these data were compiled) thus had complete imputed data after interpolation. Any data still missing at the end of the sequence, for persons still alive in 2005, were imputed as the mean of the last available observation and the value for standardized EVGGFP at that time. (EVGGFP was collected more often and for a longer time than the other variables, and so was the most complete of the variables). The amount of missing data varied, but was generally small. Consider ADL, which could be reported either by telephone, by mail, or at a clinic visit. Of the 34,128 observations used in this analysis (5688 persons x 6 annual values), 84% were observed, 7% were not observed because of death, 7% were missing and imputed by interpolation, and 2% were missing and extrapolated as the mean of the last available ADL value and EVGGFP (both on the standardized scale). For GAIT, which could only be measured in the clinic, 79% were observed, 7% were not observed because of death, 9% were missing and imputed by interpolation, and 5% were missing and imputed by extrapolation.

2.5 Analysis

To examine the 5-year change in standardized health we tested whether the average slopes over time (year 6 minus year 1) were significantly different from one another, using paired t-tests and a Bonferroni correction for multiple comparisons (78 tests in all). The primary analysis included all persons. Additional analyses were performed within six age and sex groupings, because decline is likely related to age and sex; however, we expected that the ordering of the slopes among measures would be substantially the same in all groups. Another
analysis was limited to persons still alive at year 6, allowing age/sex comparisons to be interpreted independent of mortality. We performed one person-level analysis to determine the number of persons whose health was better, the same, or worse at year 6 than at year 1, on each variable. Better was arbitrarily defined as an improvement of 5 or more points on the standardized scale, and worse was defined as a decline of 5 or more points.

3. Results

Figure 1 shows average standardized health over time, from year 1 to year 6, for each of the 13 variables. Mean health in year 1 is 77.4 for all variables, because 77.4% of persons were EVGG at year 1. The topmost two lines are for HOSP and BED, which had the smallest slopes and thus the least decline of all the standardized variables. The bottom-most line is for GAIT, which declined fastest. Although it is difficult to distinguish among the remaining lines, the figure does indicate that all the trajectories had reasonably linear decline, on average, in the 5 years.

[Figure 1 about here]

Table 2 lists the average standardized health for each variable in each year. There is substantial variability at year 6, indicating different slopes for different variables. The tabled variables are ordered so that the topmost variable (HOSP) had the least change and the bottom-most (GAIT) had the most change. The final columns of the table present the mean slope (year 6 minus year 1) and its standard deviation (s.d.). For example, mean standardized HOSP declined from 77.4 to 65.1 (slope = -12.3 points) while GAIT dropped from 77.4 to 60.2 (slope = -17.2 points). Note that the s.d. for EVG is the largest, perhaps because it was standardized in a different way. The last line shows the difference between the slope for HOSP and the slope for GAIT. There is a 5 point difference between the highest and lowest slopes.
Figure 2 shows 50.0% confidence intervals for each slope. The low level of confidence was chosen to account approximately for paired comparisons and multiple comparisons (see appendix). In most cases, if two error bars do not overlap, then those variables have significantly different slopes. Four (sets of) variables were significantly different from all the others: (1) HOSP, (2) BED, (3) ADL and IADL; and (4) GAIT. The remaining variables had similar slopes to one another. This figure does not completely represent the results from the 78 paired t-tests, which is available in the appendix.

To address whether the ordering of the slopes was independent of age and sex, Table 3 shows the average slopes in 6 age and sex subsets. The main purpose is to determine whether the rankings of decline for the different variables are independent of age and sex; that is, whether the slopes are in descending order within each age/sex grouping (within each column). It can be seen that this is approximately the case. HOSP and BED have the smallest slopes in each column, while ADL, IADL, and GAIT usually have the largest slopes. The rankings of the slopes are thus fairly stable, meaning that the rankings were independent of age and sex. The one exception is for EVG, whose rank was quite variable, perhaps because it was standardized differently from the other variables.

The bottom line in Table 3 shows the difference in the slopes for HOSP and GAIT, which increased somewhat with age. That difference was slightly larger for women than for men at each age. This may be misleading, however, because the columns have different death rates, and columns with more deaths will show more decline. To better address this issue, Table 4 presents
the slopes for the subgroup who survived at least to year 6. As expected, the ordering is the same as in Table 3, verifying that the rankings of decline in the different variables were independent of the deaths. Table 4 was intended to show decline as a function of age and sex, without the complication of survival. As expected, the slopes became steeper with age. (The only exceptions were HOSP for men, and EVG for both sexes, where the relationship with age was not monotonic). There was no consistent gender pattern. The table’s bottom line shows the difference in the slope for HOSP and GAIT. This difference increased with age, and was larger for women than for men. These differences were not tested formally because they were not the main interest of this paper.

[Table 4 about here]

All of the analyses showed the mean, or population-level decline. As a supplemental analysis, we calculated the percent of persons who improved by 5 or more points (“better”), declined by 5 or more points (“worse”) or the remainder who were called “same”. We found that 10 to 22% of the persons improved, depending on the measure, that 25% to 51% stayed the same, and that 28% to 53% got worse (data not shown). Thus, in contrast to the negative population trend, only half or less of the sample had worse health at the end of 5 years, and some even improved.

We also explored a different method of standardization, which dichotomized each variable as to whether the value was “healthy” or “not healthy”, using thresholds shown elsewhere. That method ranked the slopes similarly to the method used in this paper, but was sensitive to the thresholds used to define healthy. For example, depending on how we defined a healthy gait speed, GAIT either had the most extreme slope or fell somewhere in the middle of the variables.
4.0 Summary and discussion

4.1 Summary

Table 2 and Figure 1 give the main results of this study. All variables declined over time, on average. Slopes were similar, but there were significant differences. For the entire sample, the 5-year decline in standardized health varied from a decline of 12 points for hospitalization to a decline of 17 points for gait speed. In the older subgroups, decline was greater and the spread among the slopes was larger. In nearly all comparisons, standardized health based on hospitalization and bed days declined the least while standardized ADL, IADL, and gait speed declined the most. The statistical significance of the differences between variables can be determined approximately by comparing the error bars in Figure 2, or more completely in the appendix. These rankings were independent of age and sex. For survivors, decline became greater with age, but the relationship with gender was mixed.

4.2 Were the hypotheses confirmed?

Some, but not all, of the hypotheses were confirmed.

(1) There was statistically significant variation among the slopes, as expected.

(2) We hypothesized that Functional Health would decline fastest. This was true for ADL and GAIT, but less so for grip strength and extremity strength. Hospitalization and bed days, which we originally classified as measures of Functional Health, actually declined the least of all variables. This hypothesis was not confirmed.

(3) Decline became steeper with age, as hypothesized, but women did not tend to have less decline than men, once mortality was accounted for.

(4) Rankings of the slopes were consistent within age and sex groupings.
(5) Whether the variable was self-reported or objectively assessed was unrelated to the rankings of the slopes.

(6) We expected variables that measured the same dimension of health to have similar performance. The declines of the two quality of life variables (FLW and SPL) were quite similar, as were the two measures of strength (XSTR and GRIP). However the slopes for the two cognition variables (COG and DSST) were not very similar, nor were the variables labeled as Functional Health. The hypothesis was not supported, but it is possible that the difference within the cognition or the functional health measures were not clinically significant.

(7) As expected, the alternative standardization method yielded similar rankings to the method used in this paper.

4.2 Features of variables with low and high decline.

HOSP and BED (measures of freedom), declined the least, which is encouraging from the perspective of health maintenance. Even in a population with declining health, most persons were still out and about, and did not increase the use of hospital-based care over time as much as might have been suggested by declines in their Functional Health. One technical issue is that the prevalence and incidence of hospitalization or bed days was low (data not shown); therefore it was relatively uncommon for a person to get better or to get worse on these measures, suggesting that floor and ceiling problems restricted the amount of change over time.

GAIT, ADL, and IADL declined the most. Gait speed is a major component of the Fried frailty index, and is considered by some as the “sixth vital sign”, because it is a robust outcome measure and a powerful predictor of functional decline, risk of development of frailty, and the risk of mortality. The similarities between ADL and IADL suggest that we should have classified IADL as functional health, but we did not move IADL from its original category
when we found this. ADLs are essential for independent human functioning, whereas IADLs are more discretionary activities related to domestic and community independence. ADL and IADL were sequential items on the questionnaire, and were asked in a similar format, which may explain some of their commonality. Gait speed and ADL and IADL difficulties are easy to measure, and might be used to monitor health changes for older adults. The specific health problems that affected these three measures could be investigated if an unexpected decline was observed.

4.4 Did standardization affect the results?

Standardization had the desirable features of putting all variables on the same interpretable integer/ratio scale while also accounting for death. The standardized values were the same at every study year, and depended only on the person’s (say, ADL) score at that time, not on their EVGGFP. The slopes had similar ranks under a different method of standardization, suggesting that the results are robust to the method of standardization.

Standardization has some similarities to item response theory, which equates individual items based on the expected response of a person with a given underlying “latent health” status. We instead equated variables according to expected self-rated health. For example, from Table 1, having 2 bed days, having a 3MSE score of 60, feeling unhappy about life as a whole, being extremely unsatisfied with the purpose of life, or having a CESD score of 15 can be “equated” because they all correspond to a standardized score of about 50 (only about half the persons with those values were expected to be in excellent, very good, or good health). An item response analysis would not have accounted for death, and was not necessary for our purposes.

4.5 Did mortality affect the results?
Including a value for death has the appeal of allowing every person to contribute to every year, and depends only on the assumption that the dead have no chance of being in EVGG health. Data that were missing just before death were imputed using the information of impending death, which might have down-graded some of the imputed values from their true (but unknown) values. Some of the decline in Figure 1 was due to mortality rather than specifically to worse health on a particular health dimension, and the smaller slopes shown in Table 4 affirm that there was less decline if the decedents were removed. However, deaths could not have had any effect on the relative ordering of the slopes, because exactly the same persons (and the same deaths) were included for each variable. The ordering of the slopes was substantially the same in Table 3 and Table 4, even though Table 4 represented a healthier subset of those in Table 3 (survivors). Thus inclusion of death did not affect the rankings.

4.6 Previous Literature

We are not aware of published research that compares changes over time on multiple dimensions of health with all variables on the same standardized scale. One related publication, based on earlier data from the Cardiovascular Health Study, examined change over time for many of the variables included here, but each variable was reported on its original scale. That paper examined the effects of the aging versus the dying processes on each variable, but did not compare changes across the health variables.

4.7 Limitations

This study was primarily observational and hypothesis-generating, and findings need to be replicated. Tables 3 and 4 replicated somewhat the main analysis in Table 2, indicating that the rankings were robust. Data were not presented on their original scales, but this is available
elsewhere for most of these variables. The findings for EVGGFP may be biased because it was standardized in a different way from the other variables.

4.8 Discussion

Trends were similar for all variables, but there were statistically significant differences in the slopes. The differences may not be clinically significant, but they grew larger and presumably more clinically significant with age. The 5-year decline varied from 12 points (HOSP) to 17 points (GAIT), while the difference among slopes was about 5 points. The effect of 5 years of aging was thus more important than which dimension of health was being considered.

Looking at multiple domains of health simultaneously may yield a more nuanced picture of changes in health during aging. Much of the research on changes in health during aging concentrates on single measures, especially on gait speed and difficulties with IADLs and ADLs, which had the most decline of the 13 variables. The trends over time in the other dimensions give a less pessimistic view of aging. Further, 10 to 21% of persons improved their health in 5 years, depending on the measure, while only half or less got worse. Unlike the one-hoss shay, living systems can adapt and repair themselves, and advanced age does not preclude such positive developments.

If the goal of public health is to help older adults to square out the mortality curve and “fall to pieces all at once”, this goal might be furthered by re-allocating the relative amount of public health resources devoted to maintaining health across the various dimensions, with more attention to health problems that affect gait speed, ADLs, and IADLs. For example, there could be greater emphasis on exercise programs for walking speed, or occupational therapy interventions for IADL and ADL impairments, or more generally, ways to limit the development
of frailty. This emphasis could unfortunately suggest less public health attention to improving mental health and quality of life, but it is possible that these domains sustain themselves, or that interventions to improve functional health would improve the other dimensions as well. Individual-level interventions would, of course, emphasize the problem areas for that individual, perhaps based on more frequent ascertainment of the relevant domains.

4.9 Conclusions

Older adults did not, on average, “fall to pieces all at once”, but rather the measures of mental health and quality of life deteriorated more slowly than did functional health. Improvement in physical function measures might be the most reasonable target for public health interventions for older adults, and GAIT may be the most sensitive indicator of age-related decline in older adults.

Further research is needed to validate these findings. Different measures of health from different datasets would be of interest. Other research could investigate whether these differences among variables are clinically important for prognosis or decision-making, for instance in advising individuals and families about advance care planning, starting or foregoing treatments, the need for assistance in activities of daily living, or transitions in living situation. The time horizons at which different changes become relevant also merit attention, and whether the relatively small declines seen in younger persons may be ignored. Specific hypotheses, based on these findings, can be tested more efficiently in future research because there will be fewer “multiple comparisons” to account for.
Table 1. Definitions of “healthy” based on 16 health-related variables (Dead=0)

<table>
<thead>
<tr>
<th>Label</th>
<th>Measure</th>
<th>Examples of Standardization * (% Probability of being EVGG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOSP</td>
<td>Hospitalization (1 yr)</td>
<td>No Hosp last year = 76%; Yes = 55%</td>
</tr>
<tr>
<td>BED</td>
<td>Bed Days due to illness or injury (last 14 days)</td>
<td>0 = 76%; 1 = 61%; 2 = 52%; 5 = 35%; 8=27%; 10 = 23%; 14 = 18%</td>
</tr>
<tr>
<td>COG</td>
<td>Cognition (3MSE, 0-100)</td>
<td>0 = 28%; 20 = 33%; 40=43%; 60 = 49%; 80=63%; 90=74%; 95 = 81</td>
</tr>
<tr>
<td>EXSTR</td>
<td>Extremity Strength (problems of lifting, reaching, gripping coded 0-3% sum is 0-9)</td>
<td>No limitations= 85%; 1 = 68%; 2 = 57%; 3 = 49%; 5 = 37%; 7 = 30%; 9 = 24%</td>
</tr>
<tr>
<td>FLW</td>
<td>Feeling about Life as a Whole</td>
<td>Delighted=90%; pleased=80%; mostly satisfied=69%; mostly dissatisfied=58%; unhappy=48%; terrible=40%</td>
</tr>
<tr>
<td>SPL</td>
<td>Satisfaction with the Purpose of Life (1 to 10)</td>
<td>Extremely satisfied (1)=82%; 2 = 81%; 3 = 76%; 4 = 71%; 6 = 62%; 8 = 56%; extremely dissatisfied (10) = 50%</td>
</tr>
<tr>
<td>EVG **</td>
<td>Self-rated Health (EVGFFP)</td>
<td>E = 95%; VG=90%; G=80%; F = 30%; P=15%</td>
</tr>
<tr>
<td>DEP</td>
<td>Depression (CESD)</td>
<td>0 = 92%; 2 = 85%; 5 = 80%; 10 = 63%; 15 = 48%; 20 = 35%; 30 = 17%</td>
</tr>
<tr>
<td>DSST</td>
<td>Digit Symbol Substitution Test (# correct)</td>
<td>10=50%; 20 = 67%; 40 = 80%; 60 = 86%; 80=89%; 90=90%</td>
</tr>
<tr>
<td>GRIP</td>
<td>Grip strength-dominant hand (measured)</td>
<td>0 = 23%; 5 = 52%;10 = 64%; 29 = 74%; 40 = 82%; 60 = 86%</td>
</tr>
<tr>
<td>ADL</td>
<td># of difficulties with Activities of Daily Living - walking, transferring, eating, dressing, bathing, or toileting)</td>
<td>0 difficulties = 81%; 1 = 57%; 2= 42%; 3=34%; 4 = 29%; 5=26%; 6=24%</td>
</tr>
<tr>
<td>IADL</td>
<td># of difficulties with Instrumental Activities of Daily Living—heavy or light housework, shopping, meal preparation, money management, or telephoning)</td>
<td>0 difficulties = 84%; 1 = 61%; 2 = 46%; 3 = 37%; 4 = 32%; 5 = 29%; 6 = 28%</td>
</tr>
<tr>
<td>GAIT</td>
<td>Gait speed (# of Seconds to walk 15 feet)</td>
<td>2 = 95%; 4 = 86%; 6 = 75%; 10 = 54%, 50 = 4%</td>
</tr>
</tbody>
</table>

*Dead is always coded as 0.

**EVGG is standardized as probability of being healthy 1 year later.
Table 2 Mean Standardized Health by Year (N=5688)

Mean Standardized Health over Time for 13 Measures of Health

<table>
<thead>
<tr>
<th>YEAR</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>(y6 - y1)</th>
<th>Slope (slope)</th>
<th>S.D. (slope)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOSP</td>
<td>77.4</td>
<td>75.8</td>
<td>73.3</td>
<td>71.0</td>
<td>68.0</td>
<td>65.1</td>
<td>-12.3</td>
<td>26.0</td>
<td></td>
</tr>
<tr>
<td>BED</td>
<td>77.4</td>
<td>75.3</td>
<td>73.3</td>
<td>70.6</td>
<td>67.7</td>
<td>64.7</td>
<td>-12.7</td>
<td>26.5</td>
<td></td>
</tr>
<tr>
<td>COG</td>
<td>77.4</td>
<td>75.6</td>
<td>72.8</td>
<td>70.5</td>
<td>66.6</td>
<td>63.5</td>
<td>-13.9</td>
<td>26.7</td>
<td></td>
</tr>
<tr>
<td>XSTR</td>
<td>77.4</td>
<td>75.7</td>
<td>72.4</td>
<td>69.8</td>
<td>66.4</td>
<td>63.1</td>
<td>-14.3</td>
<td>27.6</td>
<td></td>
</tr>
<tr>
<td>FLW</td>
<td>77.4</td>
<td>75.2</td>
<td>72.3</td>
<td>69.4</td>
<td>66.1</td>
<td>63.0</td>
<td>-14.4</td>
<td>26.9</td>
<td></td>
</tr>
<tr>
<td>GRIP</td>
<td>77.4</td>
<td>75.9</td>
<td>72.9</td>
<td>70.0</td>
<td>66.8</td>
<td>62.8</td>
<td>-14.6</td>
<td>26.8</td>
<td></td>
</tr>
<tr>
<td>SPL</td>
<td>77.4</td>
<td>75.1</td>
<td>72.1</td>
<td>69.4</td>
<td>66.0</td>
<td>62.6</td>
<td>-14.8</td>
<td>27.0</td>
<td></td>
</tr>
<tr>
<td>DEP</td>
<td>77.4</td>
<td>75.6</td>
<td>71.9</td>
<td>69.6</td>
<td>65.6</td>
<td>62.1</td>
<td>-15.3</td>
<td>27.6</td>
<td></td>
</tr>
<tr>
<td>EVG</td>
<td>77.4</td>
<td>76.2</td>
<td>73.0</td>
<td>69.7</td>
<td>65.7</td>
<td>62.1</td>
<td>-15.3</td>
<td>31.3</td>
<td></td>
</tr>
<tr>
<td>DSST</td>
<td>77.4</td>
<td>75.1</td>
<td>72.1</td>
<td>68.8</td>
<td>65.5</td>
<td>62.0</td>
<td>-15.4</td>
<td>27.3</td>
<td></td>
</tr>
<tr>
<td>ADL</td>
<td>77.4</td>
<td>74.8</td>
<td>72.3</td>
<td>68.6</td>
<td>64.9</td>
<td>61.2</td>
<td>-16.2</td>
<td>27.5</td>
<td></td>
</tr>
<tr>
<td>IADL</td>
<td>77.4</td>
<td>74.2</td>
<td>71.8</td>
<td>69.1</td>
<td>65.1</td>
<td>61.0</td>
<td>-16.4</td>
<td>28.1</td>
<td></td>
</tr>
<tr>
<td>GAIT</td>
<td>77.4</td>
<td>74.2</td>
<td>70.4</td>
<td>67.6</td>
<td>64.0</td>
<td>60.2</td>
<td>-17.2</td>
<td>27.7</td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>77.4</td>
<td>75.3</td>
<td>72.4</td>
<td>69.6</td>
<td>66.0</td>
<td>62.6</td>
<td>-14.8</td>
<td>27.7</td>
<td></td>
</tr>
<tr>
<td>HOSP MINUS GAIT</td>
<td>77.4</td>
<td>75.3</td>
<td>72.4</td>
<td>69.6</td>
<td>66.0</td>
<td>62.6</td>
<td>-14.8</td>
<td>5.0</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Mean Slopes of Standardized Health by age and sex  (N=5688)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Female</th>
<th></th>
<th></th>
<th></th>
<th>Male</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age year 1</td>
<td>66-70</td>
<td>71-75</td>
<td>76-99</td>
<td>66-70</td>
<td>71-75</td>
<td>76-99</td>
<td>66-70</td>
<td>71-75</td>
</tr>
<tr>
<td>N</td>
<td>1211</td>
<td>1020</td>
<td>1056</td>
<td>745</td>
<td>802</td>
<td>854</td>
<td>1211</td>
<td>1020</td>
</tr>
<tr>
<td>HOSP</td>
<td>-5.0</td>
<td>-7.4</td>
<td>-15.9</td>
<td>-9.6</td>
<td>-13.6</td>
<td>-25.1</td>
<td>-5.0</td>
<td>-7.4</td>
</tr>
<tr>
<td>BED</td>
<td>-5.5</td>
<td>-6.9</td>
<td>-16.9</td>
<td>-9.5</td>
<td>-13.5</td>
<td>-26.6</td>
<td>-5.5</td>
<td>-6.9</td>
</tr>
<tr>
<td>COG</td>
<td>-5.4</td>
<td>-9.3</td>
<td>-19.7</td>
<td>-9.7</td>
<td>-14.8</td>
<td>-27.4</td>
<td>-5.4</td>
<td>-9.3</td>
</tr>
<tr>
<td>XSTR</td>
<td>-6.4</td>
<td>-8.6</td>
<td>-18.6</td>
<td>-11.0</td>
<td>-15.4</td>
<td>-29.1</td>
<td>-6.4</td>
<td>-8.6</td>
</tr>
<tr>
<td>FLW</td>
<td>-6.6</td>
<td>-8.7</td>
<td>-18.9</td>
<td>-11.1</td>
<td>-15.4</td>
<td>-28.7</td>
<td>-6.6</td>
<td>-8.7</td>
</tr>
<tr>
<td>GRIP</td>
<td>-6.5</td>
<td>-9.4</td>
<td>-18.6</td>
<td>-11.1</td>
<td>-16.2</td>
<td>-28.7</td>
<td>-6.5</td>
<td>-9.4</td>
</tr>
<tr>
<td>SPL</td>
<td>-6.4</td>
<td>-9.4</td>
<td>-20.2</td>
<td>-11.2</td>
<td>-15.9</td>
<td>-28.9</td>
<td>-6.4</td>
<td>-9.4</td>
</tr>
<tr>
<td>DEP</td>
<td>-7.0</td>
<td>-9.8</td>
<td>-19.7</td>
<td>-12.0</td>
<td>-17.4</td>
<td>-29.2</td>
<td>-7.0</td>
<td>-9.8</td>
</tr>
<tr>
<td>EVG</td>
<td>-10.0</td>
<td>-11.4</td>
<td>-17.3</td>
<td>-11.8</td>
<td>-15.7</td>
<td>-27.8</td>
<td>-10.0</td>
<td>-11.4</td>
</tr>
<tr>
<td>DSST</td>
<td>-7.6</td>
<td>-10.8</td>
<td>-20.2</td>
<td>-11.3</td>
<td>-16.5</td>
<td>-28.7</td>
<td>-7.6</td>
<td>-10.8</td>
</tr>
<tr>
<td>ADL</td>
<td>-8.3</td>
<td>-11.2</td>
<td>-21.5</td>
<td>-11.5</td>
<td>-16.8</td>
<td>-30.1</td>
<td>-8.3</td>
<td>-11.2</td>
</tr>
<tr>
<td>IADL</td>
<td>-8.5</td>
<td>-11.5</td>
<td>-21.2</td>
<td>-11.9</td>
<td>-17.3</td>
<td>-30.5</td>
<td>-8.5</td>
<td>-11.5</td>
</tr>
<tr>
<td>GAIT</td>
<td>-8.7</td>
<td>-12.7</td>
<td>-22.6</td>
<td>-12.5</td>
<td>-18.7</td>
<td>-31.1</td>
<td>-8.7</td>
<td>-12.7</td>
</tr>
<tr>
<td>HOSP MINUS GAIT</td>
<td>3.7</td>
<td>5.3</td>
<td>6.6</td>
<td>2.9</td>
<td>5.1</td>
<td>6.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Slope is defined as standardized health in Year 6 minus health in Year 1.
Table 4  Mean slopes of Standardized Health by age and sex, for survivors only.

Slopes* of Standardized Health for Persons Alive in Year 6 (N=4786)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age year 1</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>66-70</td>
<td>71-75</td>
<td>76-99</td>
</tr>
<tr>
<td>HOSP</td>
<td>1135</td>
<td>933</td>
<td>834</td>
</tr>
<tr>
<td>BED</td>
<td>-0.7</td>
<td>-1.8</td>
<td>-2.3</td>
</tr>
<tr>
<td>COG</td>
<td>-1.2</td>
<td>-1.3</td>
<td>-3.3</td>
</tr>
<tr>
<td>XSTR</td>
<td>-0.9</td>
<td>-3.8</td>
<td>-7.7</td>
</tr>
<tr>
<td>FLW</td>
<td>-2.1</td>
<td>-3.2</td>
<td>-6.3</td>
</tr>
<tr>
<td>GRIP</td>
<td>-2.4</td>
<td>-3.1</td>
<td>-5.7</td>
</tr>
<tr>
<td>SPL</td>
<td>-2.3</td>
<td>-4.0</td>
<td>-6.5</td>
</tr>
<tr>
<td>DEP</td>
<td>-2.1</td>
<td>-3.8</td>
<td>-7.1</td>
</tr>
<tr>
<td>EVG</td>
<td>-2.8</td>
<td>-4.5</td>
<td>-7.3</td>
</tr>
<tr>
<td>DSST</td>
<td>-5.9</td>
<td>-6.4</td>
<td>-5.8</td>
</tr>
<tr>
<td>ADL</td>
<td>-3.1</td>
<td>-5.3</td>
<td>-8.7</td>
</tr>
<tr>
<td>IADL</td>
<td>-4.0</td>
<td>-5.9</td>
<td>-9.5</td>
</tr>
<tr>
<td>GAIT</td>
<td>-4.2</td>
<td>-6.2</td>
<td>-9.5</td>
</tr>
<tr>
<td>MEAN</td>
<td>-5.9</td>
<td>-7.3</td>
<td>-11.4</td>
</tr>
</tbody>
</table>

* Slope is defined as standardized health in Year 6 minus health in Year 1.
Figure 1  Standardized Health over time

Standardized Health over Time

Variable
- HOSP
- BED
- COG
- XSTR
- FLW
- GRIP
- SPL
- DEP
- EVG
- DSST
- ADL
- IADL
- GAIT

Mean Standardized Health

YEAR FROM BASELINE

10_TABLES_GRAPHS_FOR_PAPER_15.sps 6-29-2012
Figure 2  
50% confidence intervals for slopes

50% Confidence Bars for Slopes of Standardized Health

Variable

See Appendix for exact comparisons of variables.
Appendix 1

Test for differences in slopes

The paired t statistic to compare two slopes measured on the same person is

\[
t_{\text{paired}} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{(s_1^2 + s_2^2)(1-r_{12}) / N}}
\]

\{1\},

where \(y_1\) and \(y_2\) represents two slopes, \(s_1\) and \(s_2\) the standard deviations of the slopes, and \(r_{12}\) represents the correlations between the slopes. We conducted 78 paired t-tests, one for each pair of slopes, and used the Bonferroni method to account for multiple comparisons, multiplying each p-value by 78. The great majority of results were statistically significant (the adjusted p-value was \(< .05\)). Results were as follows:

- **HOSP**: significantly different from all other variables
- **BED**: significantly different from all
- **COG**: significantly different from all but XSTR, FLW
- **XSTR**: all but COG, FLW, SPL, EVG
- **FLW**: all but COG, XSTR, EVG
- **SPL**: all but XSTR, EVG, DEP, DSST
- **Evg**: all but XSTR, FLW, SPL, DEP, DSST, GRIP, ADL
- **DEP**: all but SPL, EVG, DSST, GRIP
- **DSST**: all but SPL, EVG, DEP
- **GRIP**: all but EVG, DEP, DSST
- **ADL**: all but EVG, GRIP, IADL
- **IADL**: all but ADL
- **GAIT**: significantly different from all.
In general, the slopes of variables with similar rankings were not significantly different from one another, but that was not always the case. Note that SPL and FLW were significantly different, despite being so close in value. This is because the two variables were highly correlated ($r_{12} = .95$ if deaths are included). The high correlation makes the denominator in equation {1} small, resulting in a high t-statistic. The high correlation is likely due to the similar content of the two items, and also to the fact that SPL and FLW were asked in the same part of the questionnaires, and so likely had similar response and missingness patterns. ADL was significantly different from DEP and DSST, even though it is not different from EVG and GRIP, which have higher and lower slopes, respectively. This is because the t statistic is not a measure only of the mean difference, but also of variances and correlations, as shown in equation {1}. EVG was not significantly different from many of the other variables, perhaps because it was the standard used for the other variables.

Figure 2 shows 50% confidence intervals for the slopes of the 13 health variables. The traditional 95% confidence intervals would have provided an approximate test for significant differences if the slopes were independent, but they were not. Each slope was calculated for the same 5,688 persons, and so a paired analysis was required. In our data the t statistic for the paired test was typically about 2.9 times as large as the t statistic for the unpaired test (data not shown). If 1.96 is the critical value for t-unpaired, we should use $1.96/2.9 = 0.675$ to represent the paired test. In the table of normal probabilities, the area below .675 is about .75, meaning that the 1-tailed alpha is $1 - .75 = .25$, and the 2-tailed alpha is about .50. Thus we could approximate a paired t test by doing an unpaired test with alpha = .50. We showed 50% confidence intervals in Figure 2 to account approximately for the pairedness, assuming that all
pairs had the same standard deviations and correlations, even though that was not the case. The exact results are shown above, in this appendix.
Acknowledgments

The research reported in this article was supported by contracts HHSN268201200036C, N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, and grant HL080295 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at [http://www.chs-nhlbi.org/pi.htm](http://www.chs-nhlbi.org/pi.htm).
References


5 Hsu H-C, Jones BL. Multiple Trajectories of Successful Aging of Older and Younger Cohorts The Gerontologist 8 March 2012. (online)


Diehr P, Patrick DL, McDonell MB, Fihn SD. 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. Medical Care 2003; 41:1065-1073. (PMID: 12972846)


