May 13, 2013

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

Paula Diehr, *University of Washington*
Stephen Thielke, *University of Washington*
Anne Newman, *University of Pittsburgh*

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

Paula H. Diehr,[1,2] Stephen M. Thielke,[3,4] Anne B. Newman,[5] Calvin Hirsch,[6] and Russell Tracy[7]

[1]Department of Biostatistics,
[2]Department of Health Services,
[3]Department of Psychiatry and Behavioral Sciences,
[4]Geriatrics Research, Education, and Clinical Center, Puget Sound VA Medical Center,
University of Washington, Seattle.
[5]Department of Epidemiology and Center for Aging and Population Research, University of Pittsburgh, Pennsylvania.
[6]Departments of Internal Medicine and Public Health Sciences, University of California, Davis Medical Center,
Sacramento, CA.
[7]Department of Biochemistry, University of Vermont College of Medicine, Burlington.

Address correspondence to Paula Diehr, PhD, Department of Biostatistics, University of Washington, 1959 NE Pacific Ave, Seattle,
WA 98195. Email: pdiehr@uw.edu

Background. The health of older adults declines over time, but there are many ways of measuring health. It is unclear whether all health measures decline at the same rate or whether some aspects of health are 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, activities of daily living, instrumental activities of daily living, 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, whereas health measured by activities of daily living, instrumental activities of daily living, and gait speed declined the most. These rankings were independent of age, sex, mortality patterns, and the method of standardization.

Conclusions. 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, activities of daily living, and instrumental activities of daily living. Public health interventions to address problems with gait speed, activities of daily living, and instrumental activities of daily living may help older adults to remain healthier in all dimensions.

Key Words: Aging—Cognition—ADL—IADL—Gait.

THE average health of older adults declines with age, usually more steeply near the time of death (1–3). But there are many different aspects of health, which may decline on different schedules (4,5). The Wonderful One-Hoss Shay, which "fell to pieces all at once," has been called an attractive metaphor for aging but one that is inadequate because humans experience differential wear out (6). The rate at which different aspects of health change has not been well studied. To address this issue, we calculated the change over time in 13 health variables and studied which variables declined faster and slower. The variables were all standardized to be on the same scale.

Table 1 gives the full names and abbreviations of the 13 variables used here, which are common measures of health used in aging research. These variables encompass multiple domains of health, such as functional health, mental health, cognition, quality of life, overall health and function, and "freedom" from confinement.

Our goal was to compare the 5-year change in the 13 standardized variables, to determine which declined fastest

and which remained relatively stable. We evaluated several hypotheses about relative decline over time:

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. We also expected the rankings of change to be the same under alternate methods of standardization.

METHODS

Data

Data came from the Cardiovascular Health Study, a population-based longitudinal study of risk factors for heart disease and stroke in 5,888 adults aged 65 and older at baseline (11). 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 5,201 participants, recruited in about 1990, had up to 10 annual clinic examinations. A second cohort of 687 African Americans, from three of the original study communities, was enrolled in about 1993 and had up to seven annual examinations. Follow-up is ongoing for mortality.

All data, from 1990 to 1999, referred to here as the reference data set, were used to create the standardized variables (see "Standardization of Health Variables" section). For the decline analyses, we used years 1991–1996 for Cohort 1 and years 1994–1999 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 1 year after baseline (referred to as Year 1 in this study) and had at least one observation on each variable. Except for one sensitivity analyses, all analyses include both cohorts.

Standardization of Health Variables

A major methodological 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, whereas activities of daily living (ADL) are scored from 0 to 6. How would a 10-point decline in the 3MSE be compared with 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 ten 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 participant has died and are usually treated as missing instead.

To deal with these difficulties, we standardized each of the 13 variables to a 100-point scale, using self-rated health as the standard. The self-rated health item asked each individual if her/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 "percent 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 percent of persons at that value who were EVGG, calculated in the reference data set. The third column in Table 1 gives examples of the standardized values 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 in the reference data set 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 = 18).

The standard, EVGG, a binary variable, was set to 1 for *excellent*, *very good*, or *good* and 0 for *fair* or *poor*. The standardized 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 3MSE, we used the logarithm of 101-3MSE because 3MSE 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 a person's ADL value at that time not on their EVGGFP at that time. Any changes or differences that occur in mean standardized ADL are due only to changes or differences in the distribution of ADL difficulties.

The resulting standardized variables are all on the same scale (representing the percentage 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 (12,13), ADLs, bed days, blocks walked, body mass index, depression, EVGGFP, hospitalization, IADLs, Modified Mini-Mental State Examination (3MSE), blood pressure, gait speed (14), and quality of life (15). 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.

Table 1. Definitions and Standardized Values for 13 Health-Related Variables

Measures	Label	Examples of Standardized Values* (% probability of being in excellent, very good, or good health)
Hospitalization (1 y)	HOSP	No HOSP last year = 76%; Yes = 55%
Bed days due to illness or injury (last 14 d)	BED	0 = 76%, 1 = 61%, 2 = 52%, 5 = 35%, 8 = 27%, 10 = 23%, and 14 = 18%
Cognition (3MSE, 0–100 [7])	COG	0 = 28%, 20 = 33%, 40 = 43%, 60 = 49%, 80 = 63%, 90 = 74%, and 95 = 81
Extremity strength (problems of lifting, reaching, gripping, coded 0–3, sum is 0–9)	EXSTR	No limitations = 85%,1 = 68%, 2 = 57%, 3 = 49%, 5 = 37%, 7 = 30%, and 9 = 24%
Feeling about life as a whole (8)	FLW	Delighted = 90%, pleased = 80%, mostly satisfied = 69%, mostly dissatisfied = 58%, unhappy = 48%, and terrible = 40%
Grip strength-dominant hand (measured)	GRIP	0 = 23%, 5 = 52%, 10 = 64%, 29 = 74%, 40 = 82%, and 60 = 86%
Satisfaction with the purpose of life (1–10)	SPL	Extremely satisfied: 1 = 82%, 2 = 81%, 3 = 76%, 4 = 71%, 6 = 62%, and 8 = 56%; extremely dissatisfied: 10 = 50%,
Depression (CESD [9])	DEP	0 = 92%, 2 = 85%, 5 = 80%, 10 = 63%, 15 = 48%, 20 = 35%, and 30 = 17%
Self-rated health (EVGGFP)†	EVG	E = 95%, VG = 90%, G = 80%, F = 30%, and P = 15%
Digit symbol substitution test (# correct [10])	DSST	10 = 50%, 20 = 67%, 40 = 80%, 60 = 86%, 80 = 89%, and 90 = 90%
Number of difficulties with ADL—walking, transferring, eating, dressing, bathing, or toileting	ADL	0 difficulties = 81%, 1 = 57%, 2 = 42%, 3 = 34%, 4 = 29%, 5 = 26%, and 6 = 24%
Number of difficulties with IADL—heavy or light housework, shopping, meal preparation, money management, or telephoning	IADL	0 difficulties = 84%, 1 = 61%, 2 = 46%, 3 = 37%, 4 = 32%, 5 = 29%, and 6 = 28%
Gait speed (number of seconds to walk 15 feet)	GAIT	2 = 95%, 4 = 86%, 6 = 75%, 10 = 54%, and 50 = 4%

Notes: CESD = Center for Epidemiologic Studies Depression scale; E = excellent; VG = very good; G = good; F = fair; P = poor.
  *Dead is always coded as 0.
  †EVGGFP is standardized as the probability of being healthy 1 year later (12).

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 in the reference data set with a particular number of ADL difficulties would be in EVGG health. 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 cannot be standardized in this way (it could take on only the values of 0 or 100). EVGGFP was instead transformed to the estimated probability of being EVGG 1 year later, using values derived elsewhere (12).

For an individual, standardized "X" may be quite different for different X's. For example, for a person with "good" ADL (no ADL difficulties) but "bad" (slow) gait speed, standardized ADL would be better than standardized gait. If a person's ADL changed very little, but her/his gait speed slowed over time, then the change in standardized ADL would be smaller than the change in standardized gait.

To examine the robustness of the standardization method, we also standardized the data based on the probability of having no ADL difficulties instead of the probability of being in EVGG health. Further, we examined standardization by age, replacing each value with the mean age in the reference data set of persons with that value.

Death was assigned the mean age of persons who were dead in all years (mean age was 82.6). More information is in Supplementary Appendix 1 and in a more detailed online technical report (16). Key analyses were repeated using these differently standardized variables.

Outcome Measure

The outcome measure was standardized health at Year 6 minus the value at Year 1, referred to here for clarity as the slope (the change per 5 years), and was calculated separately for each person for each variable. (We could instead have calculated the slopes using all 6 years of data but did not do so because differential nonlinearity over time among the variables might have been mistaken for differential change from Year 1 to Year 6.) We further adjusted the standardized variables so they would all start, on average, at the same point, to make it easier to compare the slopes.

Missing Data

Missing data were imputed, separately for each variable, by linear interpolation of the person's own standardized data over time (15–17). Because death has a value (zero), everyone who died before 2005 (the end of mortality *f/u*, when these data were compiled) had complete imputed data after interpolation. Any data still missing at the end of the sequence were imputed as the mean of

the last available observation and 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 (5,688 persons × 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 was measured only 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.

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 with 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 age groups. Another analysis was limited to persons still alive at Year 6, allowing age and sex comparisons to be interpreted independent of mortality. As a sensitivity analysis, the primary analysis was repeated using the differently standardized health variables. We also 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.

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 bottommost line is for GAIT, which declined fastest. Although it is difficult to distinguish among the remaining lines, Figure 1
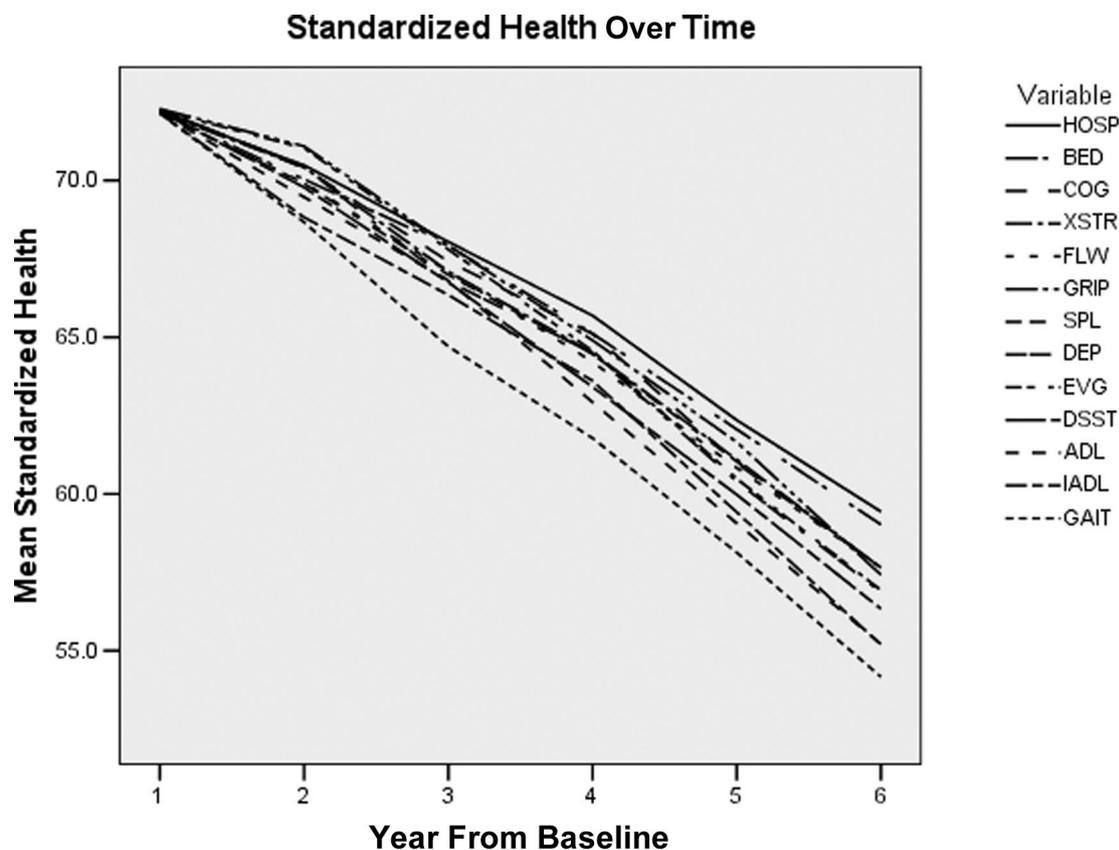

Figure 1. Standardized health over time.

does indicate that all the trajectories had reasonably linear decline, on average, across the 6 years.

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 variables are ordered so that the topmost variable (HOSP) had the least change and the bottommost (GAIT) had the most change. The final columns of the table present the mean slope (Year 6 − Year 1) and its standard deviation. For example, mean standardized HOSP declined from 77.4 to 65.1 (slope = −12.2 points), whereas GAIT dropped from 77.4 to 60.1 (slope = −17.2 points). Note that the standard deviation for EVG is the largest, perhaps because it was standardized in a different way from the other variables. The last line shows the difference between the slope for HOSP and the slope for GAIT. There is a 5-point difference (approximately) between the highest and lowest slopes. The difference between HOSP and GAIT is referred to in Tables 3 and 4.

Figure 2 shows 50% confidence intervals for the slopes. The low level of confidence was chosen to account approximately for paired comparisons and multiple comparisons (see Supplementary Appendix 2). In most cases, if two error bars do not overlap, then those variables have significantly different slopes at the 0.05 level (using a paired two-tailed t test, adjusted for multiple comparisons). Four (sets of) variables were significantly different from all the others: (a) HOSP, (b) BED, (c) ADL and IADL, and (d) GAIT. The remaining variables had similar slopes to one another. Figure 2 does not perfectly represent the results from the 78 paired t tests, which are available in Supplementary Appendix 2.

To address whether the ordering of the slopes was independent of age and sex, Table 3 shows the average slopes in six 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 and 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, whereas 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 of Table 3 shows the difference in the slopes of HOSP and GAIT. That difference became somewhat larger at older ages and 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 have more decline for that reason alone. 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 slopes of 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 article.

All of the analyses showed the mean- or population-level decline. As a supplemental analysis, we calculated the percentage of persons whose standardized health improved by 5 or more points (*better*), declined by 5 or more points (*worse*),

Table 2. Mean Standardized Health Over Time for 13 Measures of Health

	Years							
	1	2	3	4	5	6	Slope (Year 6 − Year 1)	*SD* (Slope)
HOSP	77.4	75.8	73.3	71.0	68.0	65.1	−12.3	26.0
BED	77.4	75.3	73.3	70.6	67.7	64.7	−12.7	26.5
COG	77.4	75.6	72.8	70.5	66.6	63.5	−13.9	26.7
XSTR	77.4	75.7	72.4	69.8	66.4	63.1	−14.3	27.6
FLW	77.4	75.2	72.3	69.4	66.1	63.0	−14.4	26.9
GRIP	77.4	75.9	72.9	70.0	66.8	62.8	−14.6	26.8
SPL	77.4	75.1	72.1	69.4	66.0	62.6	−14.8	27.0
DEP	77.4	75.6	71.9	69.6	65.6	62.1	−15.3	27.6
EVG	77.4	76.2	73.0	69.7	65.7	62.1	−15.3	31.3
DSST	77.4	75.1	72.1	68.8	65.5	62.0	−15.4	27.3
ADL	77.4	74.8	72.3	68.6	64.9	61.2	−16.2	27.5
IADL	77.4	74.2	71.8	69.1	65.1	61.0	−16.4	28.1
GAIT	77.4	74.2	70.4	67.6	64.0	60.2	−17.2	27.7
Mean	77.4	75.3	72.4	69.6	66.0	62.6	−14.8	
HOSP − GAIT							5.0	

Notes: ADL = activities of daily living; BED = bed days; COG = cognition; DEP = depression; DSST = digit symbol substitution test; EVG = self-rated health; FLW = feeling about life as a whole; GAIT = gait speed; GRIP = grip strength-dominant hand (measured); HOSP = hospitalization; IADL = instrumental activities of daily living; SPL = satisfaction with the purpose of life; XSTR = extremity strength.

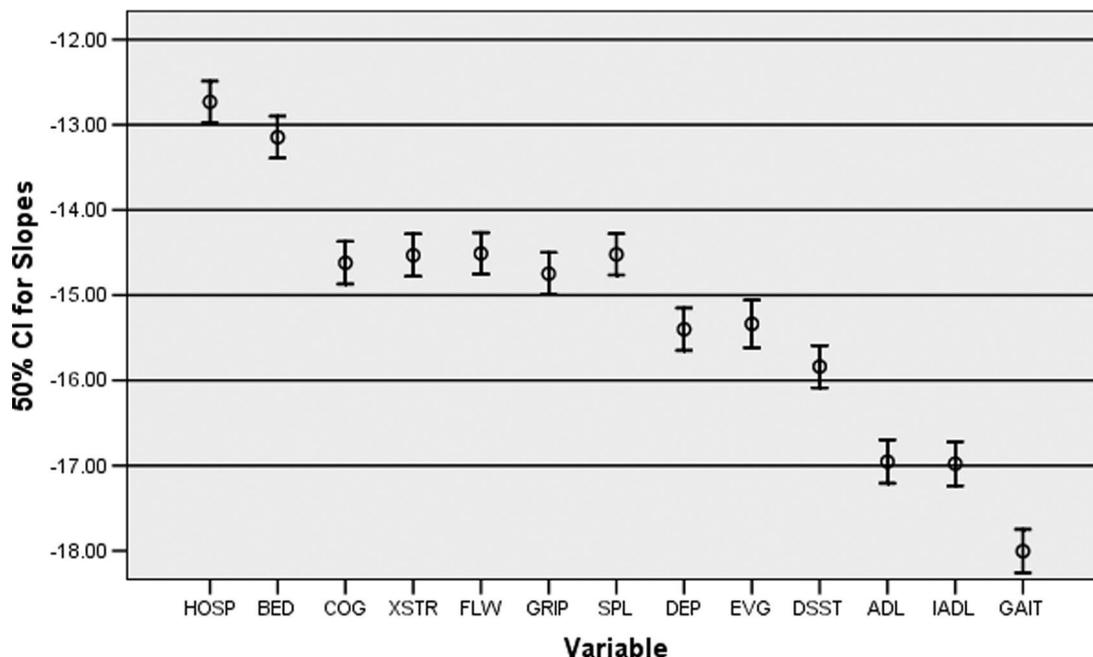## 50% Confidence Bars for Slopes of Standardized Health

Figure 2. Fifty percent confidence intervals for slopes of standardized health.

Table 3. Slopes* of Standardized Health, by Age and Sex

Sex	Female			Male		
Age, Year 1 (y)	66–70	71–75	76–99	66–70	71–75	76–99
N	1,211	1,020	1,056	745	802	854
HOSP	−5.0	−7.4	−15.9	−9.6	−13.6	−25.1
BED	−5.5	−6.9	−16.9	−9.5	−13.5	−26.6
COG	−5.4	−9.3	−19.7	−9.7	−14.8	−27.4
XSTR	−6.4	−8.6	−18.6	−11.0	−15.4	−29.1
FLW	−6.6	−8.7	−18.9	−11.1	−15.4	−28.7
GRIP	−6.5	−9.4	−18.6	−11.1	−16.2	−28.7
SPL	−6.4	−9.4	−20.2	−11.2	−15.9	−28.9
DEP	−7.0	−9.8	−19.7	−12.0	−17.4	−29.2
EVG	−10.0	−11.4	−17.3	−11.8	−15.7	−27.8
DSST	−7.6	−10.8	−20.2	−11.3	−16.5	−28.7
ADL	−8.3	−11.2	−21.5	−11.5	−16.8	−30.1
IADL	−8.5	−11.5	−21.2	−11.9	−17.3	−30.5
GAIT	−8.7	−12.7	−22.6	−12.5	−18.7	−31.1
Mean	−7.1	−9.8	−19.3	−11.1	−15.9	−28.6
HOSP − GAIT	3.7	5.3	6.6	2.9	5.1	6.0

Notes: ADL = activities of daily living; BED = bed days; COG = cognition; DEP = depression; DSST = digit symbol substitution test; EVG = self-rated health; FLW = feeling about life as a whole; GAIT = gait speed; GRIP = grip strength-dominant hand (measured); HOSP = hospitalization; IADL = instrumental activities of daily living; SPL = satisfaction with the purpose of life; XSTR = extremity strength.
*Slope is defined as standardized health in Year 6 minus health in Year 1.

or the remainder who were called "*same*." We found that 10%–22% of the persons improved, depending on the measure, 25%–51% stayed the same, and 28%–53% got worse (data not shown). Thus, in contrast to the uniformly negative population trends, only half or less of the sample had worse health at the end of 5 years and up to a quarter even improved.

Sensitivity Analyses

Several sensitivity analyses were performed for Table 2, which are described in more detail in Supplementary Appendix 1 and elsewhere (16). The variables HOSP, BED, and COG had ranks 1, 2, and 3 in Table 2. These three variables were in the top three whether we used

Table 4. Slopes* of Standardized Health for Persons Alive in Year 6 (*N* = 4,786)

Sex	Female			Male		
Age, Year 1 (y)	66–70	71–75	76–99	66–70	71–75	76–99
N	1,135	9,33	834	659	660	565
HOSP	−0.7	−1.8	−2.3	−1.8	−1.8	−3.0
BED	−1.2	−1.3	−3.3	−1.6	−1.3	−4.4
COG	−0.9	−3.8	−7.7	−1.7	−2.9	−7.8
XSTR	−2.1	−3.2	−6.3	−2.8	−2.9	−6.9
FLW	−2.4	−3.1	−5.7	−3.2	−3.4	−6.8
GRIP	−2.3	−4.0	−6.5	−2.7	−3.7	−6.9
SPL	−2.1	−3.8	−7.1	−3.2	−3.9	−6.9
DEP	−2.8	−4.5	−7.3	−3.8	−5.3	−7.6
EVG	−5.9	−6.4	−5.8	−4.2	−4.3	−8.9
DSST	−3.1	−5.3	−8.7	−3.2	−4.7	−10.1
ADL	−4.0	−5.9	−9.5	−3.5	−4.7	−9.3
IADL	−4.2	−6.2	−9.5	−3.7	−5.0	−10.0
GAIT	−4.3	−7.3	−11.4	−4.1	−6.4	−10.8
Mean	−2.8	−4.4	−7.0	−3.0	−3.9	−7.7
HOSP − GAIT	3.6	5.5	9.1	2.3	4.6	7.8

Notes: ADL = activities of daily living; BED = bed days; COG = cognition; DEP = depression; DSST = digit symbol substitution test; EVG = self-rated health; FLW = feeling about life as a whole; GAIT = gait speed; GRIP = grip strength-dominant hand (measured); HOSP = hospitalization; IADL = instrumental activities of daily living; SPL = satisfaction with the purpose of life; XSTR = extremity strength.

*Slope is defined as standardized health in Year 6 minus health in Year 1.

EVGGFP, ADL, or age to standardize the variables. If only the first cohort of persons was used, followed 8 years instead of 5, the same three variables were always in the top three. The variables ADL, IADL, and GAIT had ranks 11, 12, and 13 in Table 2. In the sensitivity analyses, these three variables were always in the bottom four (declined the most) but DSST and GRIP were each in the bottom once. The sensitivity analysis thus showed that the rankings of the variables were robust to different standardization and different data sets but that small differences in order did occur.

SUMMARY AND DISCUSSION

Summary

Table 2 and Figure 2 give the main results of this study. On average, all variables declined over time. Slopes were similar, but there were significant differences. For the entire sample, the 5-year decline in standardized health varied from a decline of 12.276 points for hospitalization to a decline of 17.2491 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, whereas 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 Supplementary Appendix 2. These rankings were independent of age and sex. For survivors, decline was greater in the older groups, but the relationship with gender was mixed. Sensitivity analyses found that using

a different variable as the standard did not substantially change the highest and lowest rankings from those shown in Table 2.

Were the Hypotheses Confirmed?

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

1. As expected, there was statistically significant variation among the slopes. But eventually the lines in Figure 1 must come together, when all have died.
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. This hypothesis was only partially confirmed.
3. As hypothesized, decline became steeper with age, but women did not tend to have less decline compared with men once differential mortality was accounted for.
4. Rankings of the slopes were consistent within age and sex groupings.
5. Rankings of the slopes were unrelated to whether the variable was self-reported or objectively assessed.
6. As expected, the alternate standardization methods yielded similar rankings to the method used in this article. But GRIP and DSST sometimes showed more change than in the main analysis. Similar sensitivity analyses are recommended for further research.

Features of Variables With Low and High Decline

HOSP and BED 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 expected from the declines in their functional health. One technical issue is that the prevalence and incidence of hospitalization or bed days was low ([18]); 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 ([19]) and is considered by some as the "sixth vital sign" because it is both a robust outcome measure and a powerful predictor of functional decline, risk of development of frailty, and the risk of mortality ([20]). ADL and IADL had similar rankings. ADLs are essential for independent human functioning, whereas IADLs are more discretionary activities related to domestic and community independence ([21]). 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, ADL, and IADL difficulties are easy to measure and seem to be sensitive ways to monitor population health changes for older adults.

Did Standardization Affect the Results?

Standardization had the desirable features of putting all variables on the same interpretable ratio scale while also accounting for death. The standardized values of (say) ADL depended only on a person's (say) ADL score at that time, not on his actual EVGGFP at that time. The 13 slopes had similar rankings under several different methods of standardization, suggesting that the results are reasonably robust to the standardization method. Nevertheless, there were a few difference. Therefore, we recommend similar sensitivity analyses to those used here, to ensure that the most important findings are robust to the type 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 ([21]). Standardization effectively equated variables according to expected self-rated health. For example, from Table 1, having two 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 Center for Epidemiologic Studies Depression 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.

Did Mortality Affect the Results?

Including a value for death (zero, in this case) has the appeal of allowing every person to contribute to every year, and it requires only 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 downrated some of the imputed values from their true (but unknown) values. Most of the decline in Figure 1 was due to mortality rather than specifically to worse health on a particular health dimension. For comparison, "standardized death" would assign the grand mean EVGG to all living persons and 0 to dead. The comparable change would be from 77.4 to 65.7, for a slope of −11.7 points. That is a lower bound for the slopes in Table 2, and the slopes for HOSP and BED are only slightly steeper than could be accounted for by death alone. The slopes in Table 4 were smaller than those in Table 3, which affirms that there was less decline if the decedents were removed. However, deaths could not have affected 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 Tables 3 and 4, even though Table 4 represented a healthier subset of those in Table 3 (the survivors). Thus, inclusion of death did not affect the rankings.

Previous Literature

We are not aware of published research that compared changes over time on multiple dimensions of health with all variables on the same standardized scale. Two related studies, based on earlier data from the Cardiovascular Health Study, examined change over time for many of the variables used here, but each variable was reported on its original scale ([3],[22]). Another recent study looked at trends in ADL, IADL, self-rated health, and grip strength by age ([23]). Those variables were recoded as "z scores," but were not specifically compared (and death was not accounted for). The z scores were all measures of different quantities and so were not standardized health in the sense used here. Those articles did not compare slopes of different health variables.

Limitations

This study was primarily observational and hypothesis-generating, and findings need to be replicated. Tables 3 and 4 and the sensitivity analyses 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 ([3],[22]). The findings for EVGGFP may be biased because it was standardized in a different way from the other variables. We discussed only the highest and lowest ranked variables, for purposes of brevity, but changes in the other variables are also of interest.

Discussion

Trends were similar for all variables, but there were statistically significant differences in the slopes. The differences may not appear to be clinically significant, but they grew larger and presumably more clinically significant with age.

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 gait speed and difficulties with IADLs and ADLs, which were the most sensitive to aging of the 13 variables. But the trends over time for the other dimensions give a less pessimistic view of aging. Further, 10%–21% of persons improved their health in 5 years, depending on the measure, whereas only half or fewer got worse. Unlike the one-hoss shay, living systems can adapt and repair themselves and advanced age does not preclude such positive developments (18). Future research can expand upon these person-level findings.

Conclusions

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

Further research is needed to validate these findings, which are limited to the variables we had available. Different measures of health and different data sets 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 ADLs, 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.

References

1. McLaughlin SJ, Jette AM, Connell CM. An examination of healthy aging across a conceptual continuum: prevalence estimates, demographic patterns, and validity. *J Gerontol A Biol Sci Med Sci*. 2012;67:783–789.

2. McLaughlin SJ, Connell CM, Heeringa SG, Li LW, Roberts JS. Successful aging in the United States: prevalence estimates from a national sample of older adults. *J Gerontol B Psychol Sci Soc Sci*. 2010;65B:216–226.

3. Diehr P, Williamson J, Burke GL, Psaty BM. The aging and dying processes and the health of older adults. *J Clin Epidemiol*. 2002;55:269–278.

4. Yashin AI, Arbeev KG, Kulminski A, Akushevich I, Akushevich L, Ukraintseva SV. Health decline, aging and mortality: how are they related? *Biogerontology*. 2007;8:291–302.

5. Hsu HC, Jones BL. Multiple trajectories of successful aging of older and younger cohorts. *Gerontologist*. 2012;52:843–856.

6. Kennedy D. Longevity, quality, and the one-hoss shay. *Science*. 3 September 2004;305:1369. doi:10.1126/science.305.5689.1369

7. Teng EL, Shui HC. The Modified Mini-Mental State (3MS) examination. *J Clin Psychiatry*. 1987;48:314–318.

8. Neugarten BL, Havighurst RJ, Tobin SS. The measurement of life satisfaction. *J Gerontol*. 1961;16:134–143.

9. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med*. 1994;10:77–84.

10. Wechsler D. *Wechsler adult intelligence scale–Revised manual*. New York: Psychological Corporation; 1981.

11. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1:263–276.

12. Diehr P, Patrick DL, Spertus J, Kiefe CI, McDonell M, Fihn SD. Transforming self-rated health and the SF-36 scales to include death and improve interpretability. *Med Care*. 2001;39:670–680.

13. 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. *Med Care*. 2003;41:1065–1073.

14. Diehr P, Johnson LL, Patrick DL, Psaty B. Methods for incorporating death into health-related variables in longitudinal studies. *J Clin Epidemiol*. 2005;58:1115–1124.

15. Diehr P, Lafferty WE, Patrick DL, Downey L, Devlin SM, Standish LJ. Quality of life at the end of life. *Health Qual Life Outcomes*. 2007;5:51.

16. Diehr, PH. Methods for dealing with death and missing data, and for standardizing different health variables in longitudinal datasets: the Cardiovascular Health Study (February 2013). *UW Biostatistics Working Paper Series*. Working Paper 390. Web site: http://biostats.bepress.com/uwbiostat/paper390

17. Engels JM, Diehr P. Imputation of missing longitudinal data: a comparison of methods. *J Clin Epidemiol*. 2003;56:968–976.

18. Thielke SM, Diehr P. Transitions among health states using 12 measures of successful aging in men and women: results from the Cardiovascular Health Study. *J Aging Res*. 2012; 2012. Article ID: 243263. doi:10.1155/2012/243263. Web site: www.hindawi.com/journals/jar/2012/243263/

19. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56:M146–M156.

20. Lusardi ML. Is walking speed a vital sign? Absolutely! *Top Geriatr Rehabil*. 2012;28:67–76.

21. McHorney CA, Cohen AS. Equating health status measures with item response theory: illustrations with functional status items. *Med Care*. 2000;38(9 Suppl):II43–II59.

22. Diehr P, Derleth A, Newman AB, Cai L. The number of sick persons in a cohort. *Res Aging*. 2007; 29:555–575 (PMID: 17411436).

23. Leopold L, Engelhartdt H. Education and physical health trajectories in old age. Evidence from the Survey of Health, Ageing and Retirement in Europe (SHARE). *Int J Public Health*. 2013;58:23–31.