1-29-2004

Incorporating Death into Health-Related Variables in Longitudinal Studies

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

Laura Lee Johnson
National Cancer Institute, lauraleehome@gmail.com

Donald L. Patrick
University of Washington, donald@u.washington.edu

Bruce Psaty
University of Washington, psaty@u.washington.edu

Introduction

Research on aging explores how health-related variables such as disability change over time, for persons with different characteristics. One of the best ways to demonstrate this change is a graph, such as a plot of the amount of disability over time. Regression analyses and tests of hypotheses would flow naturally from such graphs. Unfortunately, if people die during the study, their longitudinal health variable is "missing" after death. The usual analytic approach is to perform a complete case analysis or an available data analysis, both of which restrict study to the healthiest subgroup of the population (survivors), and may give an incorrect picture of the trajectory of health. Graphs that omit the deaths have been used to "prove" that younger people age faster than older,[1] and that the dying process has little effect on physical health.[2]

Several approaches have been suggested for the ubiquitous measure "how is your health?: excellent, very good, good, fair, or poor" (EVGGFP).[3,4] When this measure is used longitudinally, one can **add a category for the dead**. The proportion of people in each health state over time can then be shown, using stacked bar graphs.[1,5] EVGGFP can also be **dichotomized** into healthy/not healthy, with healthy defined as excellent, very good, or good and not healthy including fair, poor, or dead. This transformation permitted plotting the trajectories of health over time before and after sentinel events such as stroke, MI, and death.[1,5] Finally, the variable can be transformed to the **probability of being healthy one year later**, conditional on the current value, with deaths logically set to zero because there is no possibility of being healthy in the future.[3] Here, we use these approaches to transform 11 different health-related variables, listed in Table 1, into new variables that account for death. A definition of "healthy" was proposed for each variable. The variables were also classified as to whether they are primarily measures of health status (physical or mental), function (physical or mental), health-related mobility, or are primarily risk factors. All of these variables except BMI are likely to be acutely affected by stroke. We address whether it makes sense to include death as part of these variables, and examine properties of the transformed variables by comparing the effect of stroke on each variable. A related approach by Bozzette at all, that considers many definitions of being healthy simultaneously, is not addressed specifically here. [6]

Methods

Data and Context: the Cardiovascular Health Study.

The Cardiovascular Health Study (CHS) is a population-based longitudinal study of 5,888 adults aged 65 and older at baseline. [7] Subjects were recruited from a random sample of the Medicare eligibility lists in four U.S. counties, and extensive baseline data were collected for all subjects. After baseline, subjects had an annual clinic visit and provided additional information by mail and telephone. Two cohorts were followed, one with 9 years of follow-up (n=5201) and the second (all African American, n=687) with 6 years of follow-up. Data collection began in about 1990, and follow-up for longitudinal variables was virtually complete for all surviving subjects in 1999. [8] At baseline the mean age was 73 (range 65 to 100), 58% were women, and 84% were white. Morbidity and mortality outcomes were identified through patient or family member self-report, review of hospital and physician records and death certificates, and adjudicated by a physician review panel.[8] As of the year 2000, 658 CHS participants had suffered an incident stroke. We used their information in the 3 years before and after the stroke (about 6 measures per person) to illustrate the effect of stroke on each variable, accounting for death.

The CHS study has longitudinal data for a large number of older adults, with 9 years of follow-up, and very little loss to follow-up. While few data were missing, about 30% of participants died during the first 9 years of follow-up. We examine 11 longitudinal health variables that were used elsewhere. [9] These include measures of function, behaviors, clinical variables or risk factors, and also have a mix of categorical and continuous distributions. We imputed values for the missing data, except those missing because of death or loss to follow-up, as explained elsewhere.[5 9] For brevity, the fine detail of how variables were defined and computed is left to the referenced papers.

Analysis

First, we categorized each continuous variable (Y) into 5 categories (plus death) and created stacked bar graphs over time. Next, we chose a definition of "healthy" for each variable, and plotted the percent who were healthy over time (with dead coded as not healthy). Finally, we estimated the probability of being healthy one year later as a function of Y using logistic regression, defining healthy as being healthy on the variable of interest 1 year later, or alternatively as being in excellent, very good, or good health 1 year later. We used the transformed variables to illustrate change before and after stroke in the 11 variables of interest. A comparison group was constructed by assigning each CHS participant a random date, and treating that as the date of a "comparison event" if he/she was still alive at that date. Mean age for the stroke and comparison groups were 75 and 77 respectively, and the percentages male were 42% and 40%. Analysis is primarily descriptive, based on graphs and tables.

Findings.

The three approaches (add a category, dichotomize, probability of being healthy) have been described elsewhere.[3] We illustrate them in detail for activities of daily living (ADL), which takes on values 0-6, as defined in Table 1. Similar information is provided for the other 10 variables. In the following we abbreviate "ADL Difficulties" as "ADLs".

[Table 1 about here]

Approach 1. Add a category for death to the longitudinal variable

ADL: Figure 1 shows a stacked bar graph of ADL before and after a stroke. The percent healthy (with no ADL difficulties) is quite flat in the three years before the stroke, drops by 25-30 points at the time of the stroke, and continues to decline after the stroke. This drop is caused in part by the deaths and in part by more people having ADLs. A qualitatively different picture emerges in Figure 2, which does not include the deaths. This figure suggests a drop of only 10-20 points at the time of the stroke, followed by a small decline, understating the devastating cumulative effect of stroke on the cohort. Figure 2 also suggests that a large number of people had ADLs, even though Figure 1 shows that relatively few ever had such difficulties. Figure 2 refers only to people alive at each time, and so does not represent the trajectory of the entire cohort who experienced a stroke.

[Figure 1 and Figure 2 about here]

Other variables: This approach is not limited to categorical variables, since any continuous variable can be categorized. For example, Figure 3 shows the distribution of systolic blood pressure (SBP) in 5 categories, plus death. Unlike the ADL graphs, the percent of the cohort with healthier blood pressure levels did not change at the time of the stroke, while the percent with unhealthy blood pressure (>135) decreased. This might be interpreted as evidence that low blood pressure before the stroke was protective against mortality. Surprisingly, analysis of the person-level transitions found that blood pressure in the year before stroke was completely unrelated to one-year survival. Rather, many survivors improved their blood pressure, no doubt due to more aggressive post-stroke treatment for hypertension. Thus, although these graphs explain the trajectory for the cohort, individual transitions may also need to be examined to explain the results fully.

[Figure 3 about here]

Approach 2. Dichotomize Y as Healthy yes/no, and define death as "not healthy"

ADL: The stacked bar graphs show the distribution of ADL health states over time, but are inconvenient if the goal is to compare the trajectories of different groups. We next transformed ADL, with 7 categories plus death, into a binary variable that takes on the value of 100 if the person is healthy (no ADL difficulties) and 0 if not healthy (1 or more ADL difficulties). Death can be considered a form of not being free of ADL difficulties, and so can also be coded as zero. The mean of this new variable is interpreted as the percent of the cohort who were healthy (alive and ADL-free). The lower-most line in Figure 4 shows the percent of the stroke cohort who were healthy, before and after the stroke, similar to what was seen in Figure 1. The percent alive after the stroke is also shown, as a dotted line. The decline in health shown in Figure 5 was probably not due entirely to the stroke, since there is also a general increase in disability with age. [8] The percent healthy and percent alive for the random-date comparison group are also shown, with the comparison group a little healthier than the stroke group before the event, and declining only slightly over time. The area between the two survival curves is .75 years in the 3 years after stroke. The area between the two percent healthy curves is .24 years before the stroke and 1.07 years after the stroke. This suggests that stroke was responsible for the loss of about .75 years of life in the 3 years after stroke. After adjusting for prior differences in the percent healthy, stroke accounted for a loss of about 1.07-.24 = .84 years of healthy life in the three years after the stroke. Thus most of the observed decline can be ascribed to the stroke rather than to aging. Only .84-.75 = .09 years of healthy life were lost due to increased ADLs; the rest were due to mortality. A more detailed analysis would adjust for baseline differences, and also account for multiple observations per person.

[Figure 4 about here]

Other variables: Table 1 gives our definitions of "healthy". Note that for BMI, both very high and very low values are considered "unhealthy". Table 2 part A shows, for each variable, the percent who were healthy in the 3 years before and after stroke (numbered years 1 to 6); the change from year 1 to year 3; the change from year 3 to year 4 (the stroke); the change from 4 to year 6; and the change from year 1 to year 6. For example, for ADL, the percent healthy (no ADLs) was 79 75 75 in the 3 years before the stroke and 47 39 34 in the 3 years after. The change from year 1 to year 3 was only 4 points, from yr 3 to year 4 (the stroke) was 27 points, from year 3 to year 6 was 40 points, from year 4 to year 6 was 13 points, and total decline was 45 percentage points.

Among the different variables, the biggest change coincident with the stroke (year 3 minus year 4) was in being hospital-free, which reasonably dropped 42 percentage points. To make comparisons easier, Table 2B shows the “healthiness ratio” (HR), roughly interpreted as the % who were as healthy as they were 3 years before the stroke. (Specifically, each row in Table 2A was divided by the value at year 1 and multiplied by 100 to yield the Table 2B values). Most variables were stable in the 3 years before the stroke, but EVGGFP and IADL decreased, and SBP was highly variable. All of the variables but SBP and number of blocks walked showed an abrupt drop after the stroke, with the biggest drops for EVGGFP and for not being hospitalized. Most variables declined further after the stroke, with the exception of bed days and hospitalizations, which after an initial decline showed improvement approximately 1 year after stroke. (If we had instead divided each value in Table 2A by the corresponding percent alive at each time, the resulting curves would be analogous to those of Figure 2, showing the percent of those still alive who were healthy at each time).

[Table 2 about here]

Figure 5 shows the HR before and after stroke for some of the variables. Most curves are consistently below the survival curve, indicating that the cohort loss in health was more than just the loss due to death. That is, the loss in the percent healthy was greater than the loss in the percent alive. The curve for BMI coincides almost perfectly with the survival curve, which indicates that most of the drop in the percent with healthy BMI can be accounted for by the deaths, which seems sensible because no immediate effect of stroke on weight was expected.

[Figure 5 about here]

The fact that the health ratio for SBP was sometimes above the survival curve indicates that the percent with good blood pressure did not drop as much as the percent alive, as was already noted in Figure 3. Twenty-two percent of the stroke cases died in the year following the stroke. Under a "null case", if death after stroke were unrelated to previous SBP, and there was no effect of stroke on SBP, and little change over time due to aging, the number of people in each non-death category would be expected to decrease by about 22%. The relative increase in the percent with good blood pressure that was actually observed was no doubt due to more aggressive treatment.

Approach 3 (Probability of being healthy).

The mean of the dichotomized variable (x100) is the percent of the people who are "healthy" (percent healthy, or PCTH). PCTH is completely interpretable, but lost some information in going from the 7 original ADL categories (+ death) to 2. Further, some interventions such as one that kept people from dying but left them with two ADLs, would not show any benefit using PCTH. Finally, some would find it unsatisfying to group persons with fair or poor health together with the dead. All of these objections can be satisfied by recoding the variable as the **probability of being healthy** 1 year in the future, conditional on the current health status.[3] The mean of this transformed variable is the expected percent who will be healthy 1 year later, or EPCTH. For categorical variables, the probability can be estimated by calculating the percent in each category who were "healthy" 1 year later with dead set to 0. For example, in the complete CHS dataset, 88% of those with no ADL disabilities were disability-free one year later, and so "no disabilities" is coded as 88; 1 difficulty is coded as 42; 2 as 21; 3 as 14; and 4-6 disabilities is coded as 7 because only 7% with 4-6 ADLs had no ADLs 1 year later. Dead persons have no chance of being disability free one year later, and so are coded as 0. The area under the PCTH curve over time is the years without disability, sometimes referred to as disability-adjusted life-years, or DALYs. Similarly, the area under the EPCTH curve is the expected DALYs starting one year in the future. The transformed variable can be graphed or used as the dependent variable in regression analyses.

The probability of being healthy may also be estimated using logistic regression. For example, for ADLs, we estimated logit (No ADLs 1 year later) =1.97+.55*(# ADLs now)-4.064*ln(# ADLs now+1) =a+b*ADL+c*ln(ADL+1). (The fit was not as good without the logarithm term). The estimated probability of being healthy (ADL-free) one year later for a given ADL value is then: exp[a+b*ADL+c*ln(ADL+1)]/[1+exp(a+b*ADL+c*ln(ADL+1))]*100. The resulting probability estimates for 0 through 6 ADLs, plus death, are: 88, 43, 20, 12, 9, 7, 7, and 0, which are quite close to the values above obtained from the raw data in the previous paragraph. Other worked-out examples of the logistic regression transformation are available. [3] Transformation equations for the other variables are in the first 3 columns of Appendix Table A1.

Another possibility is to improve the estimate of the probability of being healthy by adding other variables, such as age and sex, to the logistic regression equations. For example, we estimated logit(no ADLs one year later) = -3.81 + .562*(# ADLs now) -3.94*ln(#ADLs now +1) +.126*(male)-.117*(baseline age) + 7.81*ln(age). Appendix Table A2 provides the necessary transformation equations for the other variables. Preliminary work showed that this transformation did improve the estimates when a long time horizon was considered because it allowed for aging.

Table 2 defined "healthy" as a threshold on the scale of interest (e.g., no ADLs). Research on the physical component score (PCS) of the SF-36 found that defining "healthy" as being in excellent, very good, or good health, rather than as a threshold of the PCS itself, gave transformed variables with a less bimodal distribution and better discriminatory power, suggesting that external definitions of being healthy should be considered. [3] For example, we estimated logit(in excellent/very good/ good health one year later) = 1.19 + .22*(# ADLs now) - 1.96*ln(#ADLs now +1). The estimated probabilities of being healthy (E/VG/G) 1 year later for the ADL categories are then: 77, 51, 37, 30, 25, 23, 21, 0, which decrease less abruptly than the previous transformations. In particular, the distance between "0" and "1" ADLs is only 26 percentage points using this definition of healthy, as opposed to 45 points under the previous transformation. The mean of the new transformed variable would be interpreted as the expected percent healthy (defined as E/VG/G) 1 year later as a function of ADLs today. Coefficients for all variables are in columns 4-6 of Appendix Table 1.

Similarly, we could estimate probabilities of being healthy (E/VG/G) at the same time as the original variable was measured, which avoids the time shift in the other probability of being healthy measures. The equation is logit (E/VG/G now) = 1.39 + .239*(# ADLs now) - 2.00*ln(#ADLs now +1). Table 2C shows the probability of being healthy (in excellent, very good, or good health) at the same time as the measurement, estimated from the variable in the first column. (The actual % alive is also included). Although we will not discuss this method in detail, it is interesting to note that the means for all transformed variables are about 75 in Year 1, meaning that persons were in a sense equally healthy on all variables 3 years before their stroke.

This transformation is, in a sense, self-standardizing across different variables. Coefficients for this transformation are in columns 7-9 of Appendix Table 1.

Figure 6 shows the trajectory before and after the stroke for the five different transformations of ADL. The curves differ a little, particularly before the stroke. Before the stroke, the percent healthy is higher than the others because it refers to health now rather than a year later. The transformation accounting for age and sex gives lower values than the transformation that does not, but they are not very different. The curves for the probability of being E/VG/G are lower than the others because the percent E/VG/G is lower than the % having no ADLs, as shown in Table 2A. Interestingly, the greatest change associated with the stroke was for the simple dichotomous variable, % with no ADLs. The other transformations, which used all of the information in the data rather than simply dichotomizing, did not show as much change. More work is needed to compare the properties of these transformations.

[Figure 6 about here]

DISCUSSION

We have suggested and operationalized several approaches for incorporating death into longitudinal health variables. We then used the transformed variables to illustrate and compare the effect of a stroke on eleven different health-related variables. We next discuss which variables appear to "make sense" when they are transformed, which transformations to choose, and limitations of this research.

It is possible to add a category for death to any variable (e.g., number of books read lately), and also in many cases to know what value to assign the dead on this variable (no books).

Here, however, we wish to consider health-related variables whose underlying or latent construct may sensibly include death. Death is probably a part of the construct for measures of function and health status, and perhaps for health-related mobility, but probably not for risk factors such as BMI, SBP, or number of blocks walked. Table 1 notes the category in which we consider each variable.

With respect to dichotomization, it is always possible to separate people into 2 groups, with and without a particular characteristic. The question here is whether we feel comfortable classifying death as "not healthy" on the variable of interest. Again, this seems reasonable for measures of health status and function, but perhaps less so for mobility and risk factors. We are particularly uncomfortable with considering the dead to have "bad weight" or "high blood pressure" or "insufficient walking". All of the variables in Table 1 except SBP have been shown to become worse close to death, independent of age, a finding which may suggest that death is part of their construct.[9] Investigators must decide whether the construct underlying their variable of interest includes death. It should make sense to say that the dead are "not healthy" on the construct underlying the variable.

Which transformation to choose?

Adding a category for death can yield useful graphs and some ordinal analyses. Incorporating a value for death (PCTH or EPCTH) allows for more flexible analyses. If the goal is to compare groups, or conduct regression analyses with the health measure as the dependent variable, a definition of "healthy" must be chosen. PCTH can then be calculated directly, or EPCTH can be calculated either from the regression equations in the appendix tables or from newly developed equations. PCTH has the advantage of simplicity, but EPCTH has the advantages of not losing information and of having a distinct value for the dead, and so may permit more powerful analyses than PCTH. [10] In the stroke example, however, the biggest change was seen for PCTH, suggesting that retaining the additional information will not necessarily improve power. The better statistical properties of the transformation using an external definition of healthy (e.g., E/VG/G for all variables) may make it worth consideration, although its interpretation is a less straightforward.

In practice, we usually present stacked bar graphs and graphs of mean PCTH, after checking that the results do not change when some form of EPCTH is used. It is usually advisable to impute values for data missing for other reasons than death, since values for the dead will always be known, and the dead might have too much influence.[2]

Limitations:

We estimated a small amount of intermittent missing data from known data before and after the missing assessment.[3 9 11] This may have dampened the estimated drop at the time of the stroke. In data sets with fewer deaths and more loss to follow-up, death may not be the most important threat to validity, and methods presented here may be less salient. It is possible to misinterpret Figure 1 as representing a “typical” pattern for individuals, as already noted in the SBP results. The relatively smooth trajectories represent the over-all picture of the cohort, but individual transitions were highly variable. Person-level questions require person-level analysis.

Conclusion

Graphs of available longitudinal data may give misleading pictures of the trend over time for a cohort. Incorporating death into the analysis can provide a better picture of the cohort's trajectory over time, or before and after an event such as a stroke. Transformation of health-related variables to include death is reasonable and should be considered at least for supporting or sensitivity analyses in longitudinal studies where deaths occur.

Acknowledgments

Wake Forest University School of Medicine: Gregory L. Burke MD. **Wake Forest University—ECG Reading Center**: Pentti M. Rautaharju MD PhD. **University of California, Davis**: John Robbins MD MHS. **The Johns Hopkins University**: Linda P. Fried MD MPCTH. **The Johns Hopkins University—MRI Reading Center**: Nick Bryan MD PhD, Norman J. Beauchamp MD. **University of Pittsburgh**: Lewis H. Kuller, MD DrPCTH. **University of California, Irvine—Echocardiography Reading Center** (baseline): Julius M. Gardin MD. **Georgetown Medical Center—Echocardiography Reading Center** (follow-up): John S. Gottdiener MD. **New England Medical Center, Boston—Ultrasound Reading Center**: Daniel H. O'Leary MD. **University of Vermont—Central Blood Analysis Laboratory**: Russell P. Tracy PhD. **University of Arizona, Tucson—Pulmonary Reading Center**: Paul Enright MD. **Retinal Reading Center-University of Wisconson**: Ronald Klein MD. University of Washington—Coordinating Center: Richard A. Kronmal PhD. **NHLBI Project Office**: Jean Olson, MD MPCTH

Grants: The research reported in this article was supported by contracts N01-HC-85079 through N01-HC-85086, N01-HC-35129, and N01 HC-15103 from the National Heart, Lung, and Blood Institute.

Table 1

Variable Definitions and Symbols

VARIABLE	DEFINITION	HEALTHY	Construct
ACTIVIES OF DAILY LIVING (ADL)	Do you have any difficulty walking around your home? Getting out of bed or a chair? Because of health or physical problems do you have any difficulty or are you unable to ... eat, including feeding yourself? Dress yourself? Bathe or shower? Use the toilet, including getting to the toilet?. # (of 6).	0 ADLs	function
BED DAYS	Days spent in bed in past 14 days.	0 Days	mobility
BLOCKS	# of blocks walked in last week.	> 9 blocks	risk factor
BODY MASS INDEX (BMI)	Weight (kg) / ht^2 (meters).	20-30	risk factor
DEPRESSION	CESD has 10 depressive symptoms rated on a scale from 0 to 3. The sum of items has a maximum of 30.	< 10	mental health status
EVGGFP	Is your health: excellent, very good, good, fair, or poor).	Excellent, very good, or good	health status
HOSPITAL	1 if hospitalized in previous 6 months [or is it 100?], 0 if not	Not hospitalized	mobility
INSTRUMENTAL ACTIVITIES OF DAILY LIVING (IADL)	Because of health or physical problems do you have any difficulty or are you unable to ... Do heavy housework like scrubbing floors or washing windows; or yard work, like raking leaves or mowing? Do light housework? Shop for personal items? Prepare your own meals? Manage your money, such as paying bills? Use the telephone? # (of 6).	0 IADLs	function
MODIFIED MINI MENTAL STATE EXAM (MM)	Coded 0 to 100, 100 is best	80-100	function
SYSTOLIC BLOOD PRESSURE (SBP)	Systolic Blood Pressure	# 125	risk factor
TIMED WALK	# of Seconds to walk 15 feet.	< 10 seconds	function

Table 2
Percent Healthy over 6 years b/a stroke

A. Percent Healthy (above cut-off on variable in Column 1) Now

	-----% Healthy---------						-----Changes-----				
Variable	**YR1**	**YR2**	**YR3**	**YR4**	**YR5**	**YR6**	**D13**	**D34**	**D36**	**D46**	**D16**
ALIVE	100	100	100	78	69	61	0	22	39	17	39
Bed Days	95	95	93	68	64	56	1	25	38	12	39
Blocks	60	52	52	35	30	26	8	16	26	9	34
Depression	79	77	77	53	45	39	2	24	38	14	39
Hospital	90	88	84	42	56	53	6	42	31	-10	37
MMSE	84	80	77	54	47	40	7	22	37	14	44
Systolic	47	51	47	42	40	36	0	5	11	6	11
Timed walk	89	88	85	62	53	46	3	23	40	17	43
EVGGFP	73	67	64	32	31	30	8	33	35	2	43
BMI	75	73	74	58	52	46	0	16	28	11	28
ADL	**79**	**75**	**75**	**47**	**39**	**34**	**4**	**27**	**40**	**13**	**45**
IADL	62	57	53	32	27	24	9	21	28	7	37

B. Percent Healthy Now divided by Value at Year 1

	-----% Healthy---------						-----Changes-----				
Variable	**YR1**	**YR2**	**YR3**	**YR4**	**YR5**	**YR6**	**D13**	**D34**	**D36**	**D46**	**D16**
ALIVE	100	100	100	78	69	61	0	22	39	17	39
Bed Days	100	101	99	72	67	59	1	27	40	13	41
Blocks	100	88	86	59	51	43	14	27	43	16	57
Depression	100	98	98	68	58	50	2	30	48	18	50
Hospital	100	99	94	47	62	59	6	46	35	-11	41
MMSE	100	95	92	65	56	48	8	27	44	17	52
Systolic	100	110	100	90	85	76	0	11	24	13	24
Timed walk	100	99	96	70	59	52	4	26	44	19	48
EVGGFP	100	92	88	44	42	41	12	45	48	3	59
BMI	100	97	99	77	69	62	1	22	37	15	38
ADL	**100**	**96**	**94**	**60**	**50**	**43**	**6**	**34**	**51**	**17**	**57**
IADL	100	93	86	52	44	39	14	34	46	12	61

C. Probability of being in E/VG/G Health Now given variable in Column 1

	-----% Healthy---------						-----Changes-----				
Variable	**YR1**	**YR2**	**YR3**	**YR4**	**YR5**	**YR6**	**D13**	**D34**	**D36**	**D46**	**D16**
ALIVE	100	100	100	78	69	61	0	22	39	17	39
Bed Days	74	74	74	56	51	45	0	18	29	11	29
Blocks	75	73	72	56	49	43	3	16	29	13	32
Depression	74	73	71	55	48	42	3	16	29	13	31
Hospital	78	78	78	61	53	48	0	17	30	13	30
MMSE	73	72	71	54	48	42	2	17	30	12	31
Systolic	75	75	75	59	52	46	0	15	29	13	29
Timed Walk	71	70	69	52	45	39	2	17	30	13	32
EVGGFP	72	66	61	32	32	29	11	29	32	3	43
BMI	75	75	75	59	52	46	0	16	29	13	30
ADL	**73**	**72**	**71**	**51**	**45**	**39**	**2**	**20**	**33**	**13**	**35**
IADL	73	71	69	49	43	38	4	20	31	12	35

Appendix Table A1

Logistic Regression Coefficients for Estimating the Probability of being Healthy 1 year later.*

"Healthy" based on:	**(1) Healthy 1 year later**			**(2) E/VG/G 1 year later**			**(3) E/VG/G NOW**		
Coefficients:	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**
Variable:									
ACTIVIES OF DAILY LIVING	1.97	.55	-4.06	1.19	.22	-1.96	1.39	.239	-2.01
BED DAYS	2.55	.113	-1.73	.968	-.018	-.84	1.15	.029	-1.06
BLOCKS**	-2.47	3.29	-1.48	-.188	.782	.364	.006	.896	.185
BODY MASS INDEX	-352.7	-6.06	155.	-19.61	-.317	8.77	-19.87	-.334	9.04
DEPRESSION	3.25	-.254	-.024	1.99	-.077	-.368	2.36	-.076	-.466
EVGGFP (5=excellent, 4 is very good, …, 1=poor)	-7.448	.139	5.96	-7.45	.139	5.967	n/a	n/a	n/a
HOSPITAL	2.009	-.011	--	1.07	-.0087	--	1.24	-.0091	n/a
INSTRUMENTAL ACTIVITIES OF DAILY LIVING	1.51	.56	-3.90	1.43	.17	-1.90	1.66	.266	-2.16
MODIFIED MINI *** MENTAL STATE EXAM	-19.709	.236	.372	.282	.019	-.461	2.02	.006	-.636
SYSTOLIC BLOOD PRESSURE	31.198	-.042	-5.22	-13.47	-.033	3.852	-11.59	-.029	3.40
TIMED WALK	13.36	.126	-5.90	7.46	.178	-3.992	7.78	.188	-4.08

* logit $(\text{healthy}_1) = a + b(\text{variable}_0) + c \ln(\text{variable}_0 + 1)$

** uses Y = lg10(bk) instead of bk and ln(lg10(bk)+1) instead of ln(bk+1)

*** uses Y = ln (101 - MM SCORE) instead of ln(mmscore+1).

Appendix Table A2

Probability healthy in 1 year conditional on current health, age, and sex

Regression Coefficients* for Transformation (4)

VARIABLE= Y	constant (a)	Y (b)	ln(Y) (c)	Male (d)	Age (e)	ln(Age) (f)
ACTIVITIES OF DAILY LIVING	-3.81	.562	-3.94	.126	-.117	3.391805
BED DAYS	-38.60	.136	-1.79	-.013	-.242	13.78002
BLOCKS **	.379	3.20	-1.56	.321	-.064	0.469902
BODY MASS INDEX	-492.2	-6.17	158.6	-.179	-.574	40.97526
DEPRESSION	-51.68	-.257	.014	.024	-.270	17.42371
EVGGFP (5=excellent, 4 is very good, …,1=poor)	-2.13	.154	5.84	-.066	-.047	-0.35916
HOSPITAL	-18.31	-.010	n/a	-.256	-.141	7.204871
INSTRUMENTAL ACTIVITIES OF DAILY LIVING	-10.41	.63	-3.91	.223	-.125	4.929192
MODIFIED MINI *** MENTAL STATE EXAM	-20.95	.229	.411	-.178	-.090	2.015106
SYSTOLIC BLOOD PRESSURE	25.95	-.038	-5.56	.060	-.063	2.601397
TIMED WALK	45.21	.096	-5.29	-.333	.029	-8.04305

* logit (healthy_1) = a + b(variable_0) + c ln(variable_0 + 1)+d*male + e*age + f*ln(age)

** uses Y = lg10(bk) instead of bk and ln(lg10(bk)+1) instead of ln(bk+1)

*** uses Y = ln (101 - MM SCORE) instead of ln(mmscore+1).

FIGURE 1

# of ADLs in 3 Yrs Before/After Stroke

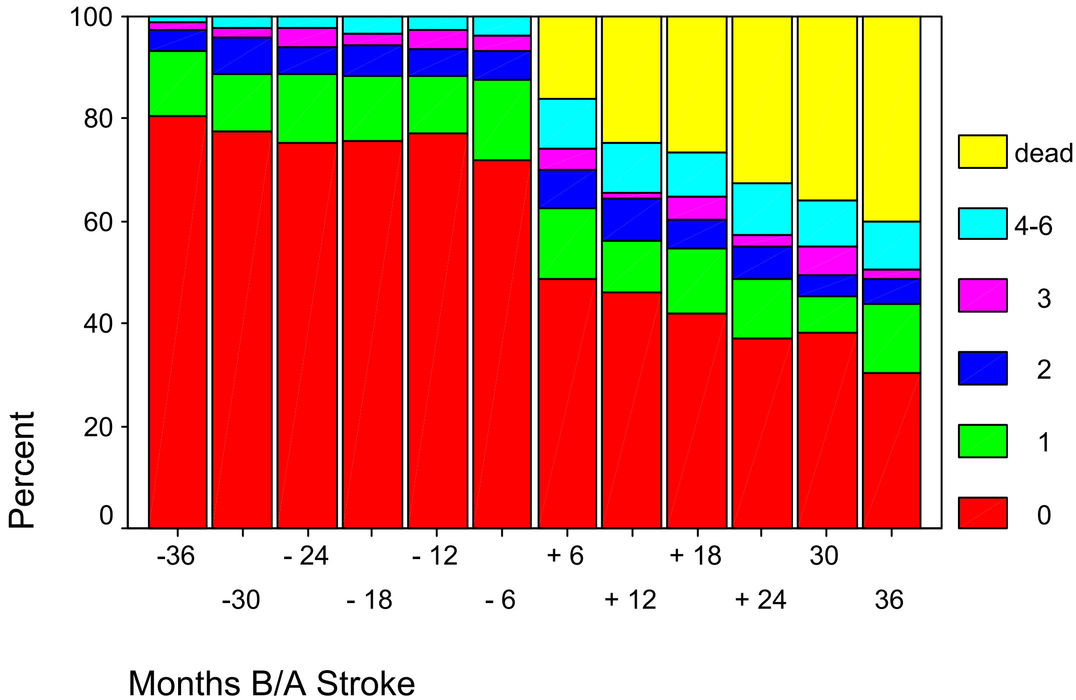
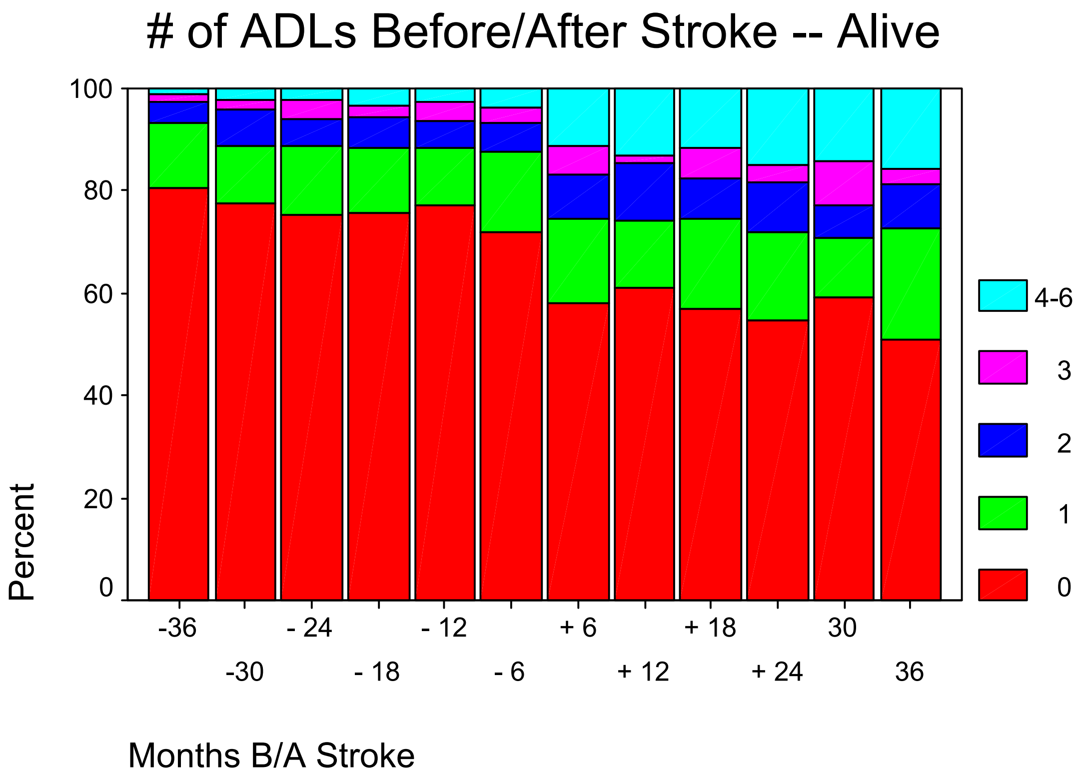


Figure 2



# of ADLs Before/After Stroke -- Alive
Percent
100
80
60
40
20
0
-36
-30
- 24
- 18
- 12
- 6
+ 6
+ 12
+ 18
+ 24
30
36
Months B/A Stroke
4-6
3
2
1
0

Figure 3

SBP in 3 Yrs Before/After Stroke

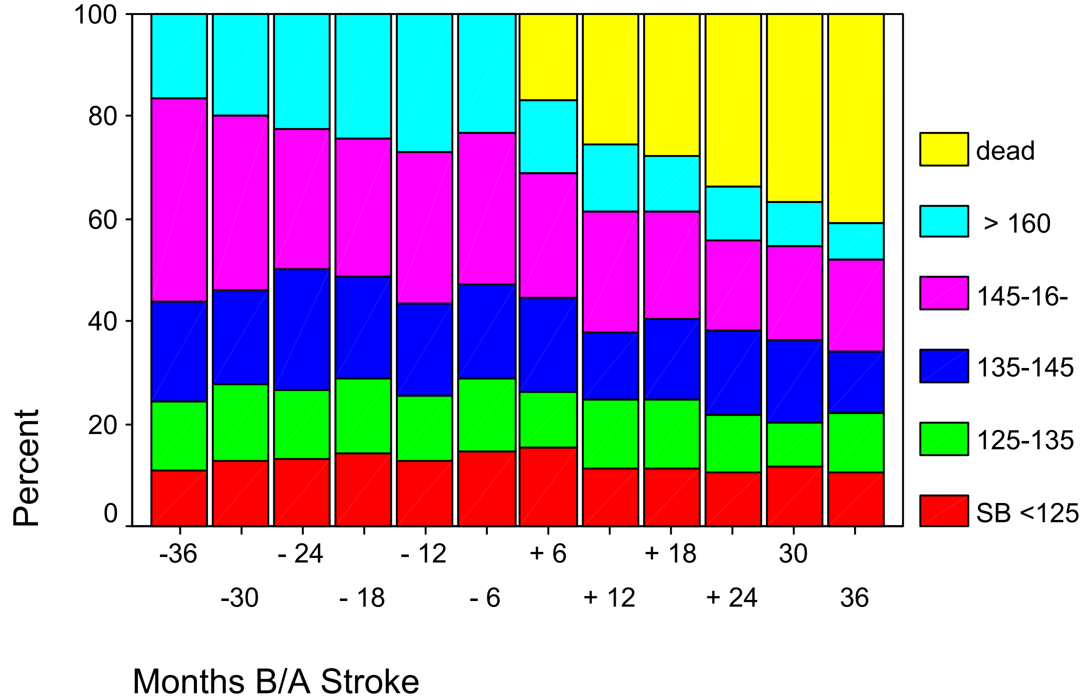
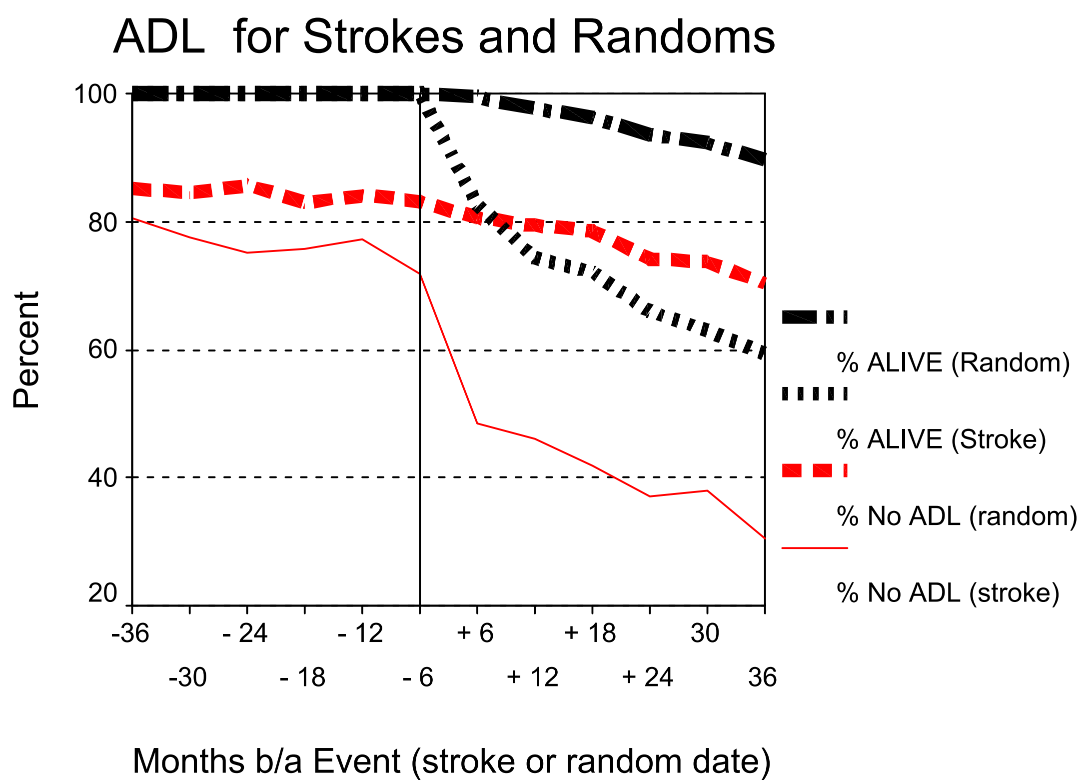
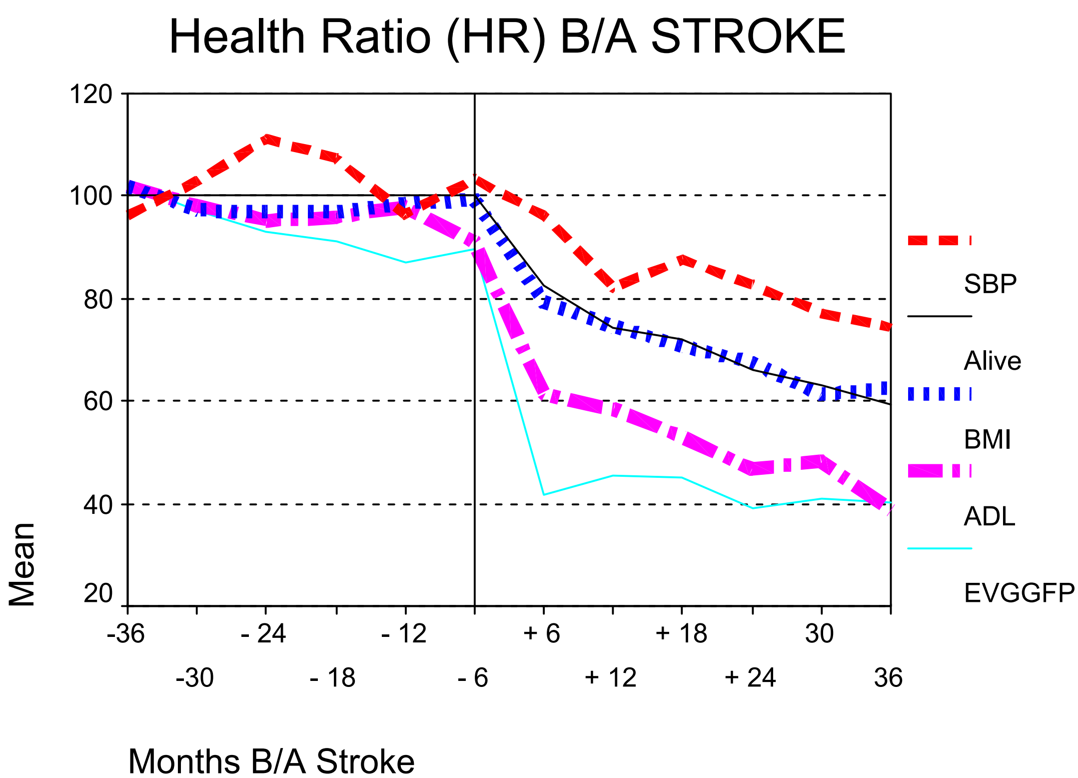


Figure 4



ADL for Strokes and Randoms
Percent
100
80
60
40
20
-36
- 24
- 12
+ 6
+ 18
30
-30
- 18
- 6
+ 12
+ 24
36
% ALIVE (Random)
% ALIVE (Stroke)
% No ADL (random)
% No ADL (stroke)
Months b/a Event (stroke or random date)

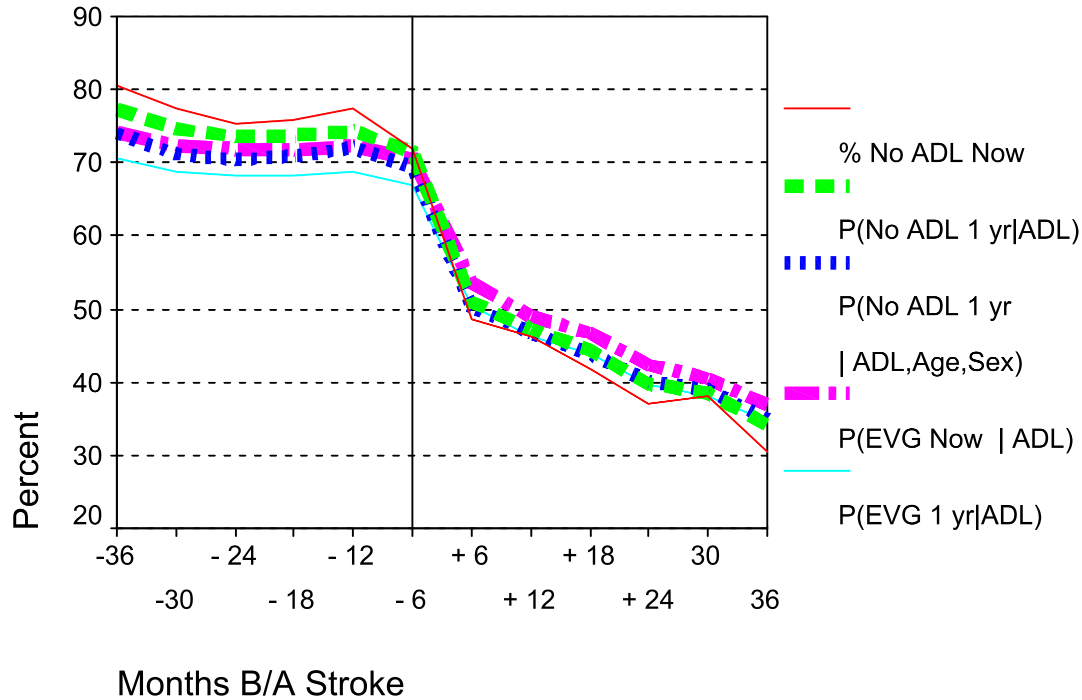
Figure 5



Health Ratio (HR) B/A STROKE
Mean
120
100
80
60
40
20
-36
-30
- 24
- 18
- 12
- 6
+ 6
+ 12
+ 18
+ 24
30
36
Months B/A Stroke
SBP
Alive
BMI
ADL
EVGGFP

Figure 6

5 Transformations of ADL



References

1. Diehr P, Patrick DL. Trajectories of health for older adults over time: accounting fully for death. Annals of Internal Medicine, 2003; 139: 416-421.

2. Diehr P, Patrick DL, McDonell MB, Fihn SD. Accounting for deaths in longitudinal studies using the SF-36: the performance of the PCS and the PCTD. Medical Care. 2003; 41:1065-1073.

3. 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. Medical care 2001; 39:670-680.

4. Diehr P, Patrick D, Hedrick S, Rothman M, Grembowski D, Raghunathan TE, Beresford S. Including deaths when measuring health status over time. Medical Care. 33:AS164-AS172. 1995.

5. Diehr P, Williamson J, Patrick DL, Bild DE, Burke GL. Patterns of self-rated health in older adults before and after sentinel health events. J Am Geriatr Soc 49:36-44, 2001.

6. Bozzette SA, Kanouse DE, Berry S, Duan N. Health status and function with zidovudine or zalcitabine as initial therapy for AIDS. A randomized controlled trial. JAMA. 1995;273:295-301.

7. Fried LP, Borhani NO, Enright PL, et al. The Cardiovascular Health Study: design and rationale. Annals of Epidemiology 1991. 1:263-276.

8. Ives G, Fitzpatrick A, Bild D, et al. Surveillance and ascertainment of cardiovascular events: the cardiovascular health study. Annals of Epidemiology 1995. 5:278-285.

9. Diehr P, Williamson J, Burke G, Psaty B. The aging and dying processes and the health of older adults. Journal of Clinical Epidemiology. 2002. 55:269-278.

10. Diehr P, Patrick DL, Burke G, Williamson J. Survival versus years of healthy life: which is more powerful as a study outcome? Controlled Clinical Trials. 20:267-279. 1999.

11. Engels JM, Diehr P. Imputation of Missing Longitudinal Data. Journal of Clinical Epidemiology. 2003; 56:968-976.