The Number of Sick Persons in a Cohort
Paula Diehr, Ann Derleth, Anne B. Newman and Liming Cai
Research on Aging 2007; 29; 555
DOI: 10.1177/0164027507305727

The online version of this article can be found at:
http://roa.sagepub.com/cgi/content/abstract/29/6/555
The Number of Sick Persons in a Cohort

Paula Diehr
*University of Washington, Seattle*

Ann Derleth
*University of Washington, Seattle*

Puget Sound Health Care System, Seattle

Anne B. Newman
*University of Pittsburgh School of Medicine*

Liming Cai
*Centers for Disease Control and Prevention, Hyattsville, Maryland*

Planning for future health needs of older adults requires a better understanding of the trajectories of health and sickness with age. The authors calculated the number of sick persons over time in a “research” cohort of older adults followed for up to 14 years, and also in a synthetic birth cohort. In the research cohort, the authors calculated the actual number of persons in each health state over time, using eight different definitions of “sick.” For the birth cohort, the authors estimated the number of sick persons each year after birth. The number of sick persons in the research cohort was approximately constant for 14 years, for all definitions of “sick.” The number sick in the birth cohort was approximately constant from ages 55 to 80 and then declined. The relative excess of sick persons in later life is caused by a decline in the number of healthy persons rather than an increase in the number who are sick. The number of sick current Medicare enrollees may be approximately constant for 14 years. These insights may help in planning for the aging population.

**Keywords:** self-rated health; equilibrium; health state; projection; older adults; Medicare

Authors’ Note: The research reported in this article was supported by contracts N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01-HC-75150, N01-HC-45133, and N01 HC-15103 from the National Heart, Lung, and Blood Institute. Additional funding came from the Department of Veterans Affairs, Health Services Research and Development Service, Grant TPP61-017, and from the University of Washington School of Public Health and Community Medicine, Department of Biostatistics, under subcontract from the National Center for Injury Prevention and Control of the Centers for Disease Control and Prevention, Grant R49/CCR316840 (Frederick P. Rivara, Ellen MacKenzie, and Gregory J. Jurkovich, Principal Investigators). The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.
Introduction

The American population is aging and life expectancy is increasing. These trends are likely to increase the number of persons at risk for costly age-associated chronic diseases, injuries, and disability, although there are conflicting scenarios about the effect on society of these trends (Lee and Skinner 1999). It is important to understand the natural trajectory of health and morbidity over time, but the patterns of health and death and their associated health care needs and expenditures over time are complex. One approach is to follow a well-defined “research” cohort from some point in time until death, and tabulate at each time how many persons are healthy (by some definition), how many are sick, and how many have died. The research cohort might be a random sample of the population, and usually include persons with a mix of ages at baseline. The measure of time is the number of years since enrollment in the cohort. Another approach is to estimate the age-specific probability of transition from one health state to another and use multistate life table methods to estimate the number who are healthy, sick, or dead each year after birth, where the measure of time in this birth cohort is age. Here, we present the number of sick persons over time in a research cohort of older adults followed for up to 14 years, using eight different definitions of “sick.” We also estimated the number sick each year in a birth cohort, using transition probabilities estimated from three large data sets.

Method

Study Design: 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, designed to identify factors related to the occurrence of coronary heart disease and stroke (Fried et al. 1991). Participants were recruited from a random sample of the Medicare eligibility lists in four U.S. counties. Persons not expected to be able to participate for the next three years (planning to move, too sick, used wheelchairs at home) were ineligible, and about 59% of those eligible agreed to enroll. The resulting sample was younger, more highly educated, more likely to be married, and had fewer activity limitations than those who declined to participate (Tell et al. 1993). Two CHS research cohorts were followed, one with nine years of follow-up ($n = 5,201$) and the second (all African American, $n = 687$) with six years of full follow-up. Some analyses omit the second cohort because of its shorter follow-up. At baseline the mean age was 73 (range 65 to
105), 58% were women, and 84% were White. Data collection began in about 1990, and follow-up is virtually complete for all surviving subjects in the year 1999 (and for a few measures through 2004).

**Health-Related Variables in CHS**

To provide information on the broad spectrum of health, we selected eight commonly measured health-related variables and defined being “sick” for each variable. These definitions included having a Modified Mini Mental State Examination score below 80 (MMSE) (Teng and Shui 1987), having one or more difficulties with activities or instrumental activities of daily living (ADL and IADL), a score above 10 on the Center for Epidemiologic Studies Depression score (CESD) (Radoff 1977), any days spent in bed in the previous two weeks (Bed Days), requiring more than 10 seconds to walk 15 feet (Timed Walk), and prevalent heart disease (CVD). CVD is defined as having angina, coronary heart disease, congestive heart failure, claudication, myocardial infarction, stroke, transient ischemic attack, angioplasty, or coronary artery bypass surgery at the survey time or earlier. A person who is sick by this definition (has CVD) cannot become healthy (no CVD) in the future. These eight measures include both self-report and clinical definitions, as well as one definition for which recovery is not possible. Findings that are similar for all of these disparate definitions of “sick” may be considered to hold in general.

The primary health variable used in this article is the self-rating of health (EVGGFP) as either excellent, very good or good (E/VG/G = healthy), or fair or poor (F/P = sick). EVGGFP was singled out because it was collected semi-annually and continued to be collected by telephone even after 1999, providing up to 29 measures per person that could be used to calculate transition probabilities. The other variables were collected only annually and were not collected after 1999. The amount of missing data was low, and persons often returned after a missed visit, allowing us to impute the missing data using simple imputation methods that have been shown to perform well in the CHS data set (Engels and Diehr 2003). Data that were missing between the first and last known measures were imputed from a person-specific regression of the variable on the log of time from the last known measure.

When measurements were missing between the last known observation and death, we imputed values using the last known observation carried forward, so that the direction of the bias in transition probabilities close to death would be clear. One potential bias is that there would be no imputed transitions from sick to healthy (which is probably reasonable) or from healthy to sick for those persons in that period. Given the large number of
additional transition pairs available for estimating those transitions, this should not be a problem. More important is a potential bias in the death rate for healthy persons. Of the two-thirds of deaths that were preceded by a known value of EVGGFP, about 62% were sick (fair or poor) in the year before death. EVGGFP was missing in the year before death for about a third of the deaths and was imputed to be sick about 56% of the time. Since persons with missing data are more likely to be sick than persons who provide data (Engels and Diehr 2003), we would have expected more than 62% to be sick in the year before death. About 14% of all the values just before death were thus imputed to be “healthy,” and some of them may have been misclassified. The estimated death rate for healthy persons may be biased upward for that reason. For persons still alive in 1999, approximately 5% of the relevant data had to be imputed. For persons who died, 20% to 29% of the data for Timed Walk were imputed; 10% to 19% were imputed for CESD, Mini Mental, EVGGFP, ADL, and IADL; and less than 10% were imputed for Bed Days.

We defined a transition pair as two EVGGFP values for the same person measured one year apart. The 5,888 CHS participants contributed about 150,000 transition pairs, which were used to estimate the probability of moving from one state to another at different ages.

**Additional Data**

To increase the quantity and the age range of the transition data, we also used EVGGFP data from two large national surveys, the Medicare Current Beneficiary Study (MCBS, 1998-2002) (Adler 1994) and the Medical Expenditures Panel Survey (MEPS, 1996-2001) (Cohen 2000). In MCBS, persons were followed two to six years, and about 41,000 persons contributed about 98,000 transition pairs. Unlike CHS, the MCBS sample was slightly sicker than the general population (Kautter et al. 2006). In MEPS, approximately 93,000 persons contributed about 224,000 transition pairs for ages 0 to 64, and 29,000 transition pairs for age greater than or equal to 65. MEPS did not survey institutionalized persons. The small number of persons not surveyed for that reason were assigned “poor” health and we imputed the data missing for other reasons, as in CHS. MEPS ages above 90 (and later 85) were set to 90 (85). One person answered for everyone in the family, meaning that MEPS information was not usually “self” reported. We ignored the survey weights, to make the data consistent with the CHS data, and so cannot claim unbiased estimates for the U.S. population. The number of persons lost to follow-up were low in both surveys.
Data on life expectancy came from the Statistical Abstracts of the United States (Statistical Abstract of the United States 2005). Estimates of the health of Americans by age came from the National Health Interview Survey, which surveys noninstitutionalized persons only (Vital and Health Statistics 1990). Data on medical expenditures by age and EVGGFP were estimated from MEPS 2002 data using the MEPSnet program (MEPSnet 2005). We fit a regression of expenditures on age and log age and used the regression estimate for ages greater than 85.

**Transition Probability Calculations**

We combined the transition pairs from all three data sets, calculated the transition probabilities for each age, and smoothed the curves over age using a moving average. Estimates for ages below 65 were based on about 3,400 transition pairs per year of age, and those over 65 were based on about 8,000 transition pairs per year.

**Projection of Future Health Status**

The research cohort showed the actual behavior of the CHS cohort over the years they were followed. To obtain more generalizable findings, we used the transition probabilities for EVGGFP to project the number of healthy, sick, and dead persons over time for a synthetic birth cohort with 98,000 healthy persons and 2,000 sick persons at baseline, chosen to agree with national statistics. This was done using standard multistate life table methods (Guralnik et al. 1993) implemented in a spreadsheet and partly in a Stata program (Weden 2005). We also created a synthetic CHS cohort of 5,201 persons whose initial distribution of age and health matched the CHS baseline population, for comparison with the CHS research cohort. An example of the calculations is given in the Synthetic Cohorts section.

**Analysis**

We first calculated the number of sick persons over the 14 years of follow-up in the CHS research cohort, for all eight definitions of being sick. We next used the estimated transition probabilities to project the number of healthy, sick, and dead persons over time in the synthetic birth cohort and in the synthetic CHS cohort, using standard multistate life table methods. We multiplied the health- and age-specific expenditure estimates by the number of persons projected to be in each age and health state to estimate lifetime medical expenditures.
Findings

Number of Healthy, Sick, and Dead Persons in the CHS Research Cohort

Figure 1 describes the first CHS cohort \( n = 5,201 \). It shows the number who were healthy (E/VG/G), sick (F/P), or dead in the 14 years after baseline. The number healthy decreased over time, and the number dead increased, but the number who were sick (the solid line) was approximately constant. This was the graph that suggested the current investigation. The pattern was similar for men and women separately (not shown).

Figure 2 shows the number of healthy, sick, or dead persons in each year of follow-up, using different definitions of “sick,” as explained above. (For convenience in plotting, the Y axis is the percentage of 5,201 persons in each health state rather than the count). The number in parentheses is the mean number of sick persons per year based on this variable. The number sick (the solid line) was always approximately constant over time, even
though “sick” is defined differently in each graph. The graph labeled EVGGFP-1 is equivalent to the first nine years in Figure 1, and EVGGFP-2 shows the same variable but for the second cohort (n = 687 and all African American). The percentage sick in EVGGFP-2 was higher than that in EVGGFP-1 but was also reasonably stable over time. The area under the upper line for EVGGFP is sometimes called “years of healthy life,” and the corresponding area for ADL is known as “active life expectancy” (Crimmins, Hayward, and Saito 1994).

**Number of Healthy, Sick, and Dead Persons in a Synthetic Cohort**

We used the estimated transition probabilities among the three health states (healthy, sick, and dead) to project the number of healthy, sick, and dead persons over time in the synthetic cohorts.
Health States and Transition Probabilities

Here, we define healthy as “E/VG/G” health and sick as “F/P” health. The probabilities of transition among the three states were estimated from the 500,000 transition pairs, and are listed in an online technical report (Diehr et al. 2005). Persons who are healthy at the first observation have the probability $P(H|H)$ of being healthy 1 year later; the probability $P(S|H)$ of being sick 1 year later; and $P(D|H)$ of being dead 1 year later. Similarly, persons who are initially sick have the associated probabilities $P(H|S)$, $P(S|S)$, and $P(D|S)$. The estimated transition probabilities change with age, as shown in Figure 3. The probabilities below age 65 are based only on MEPS data, while those for 65 and above are calculated from all three data sets. Despite the smoothing, there is an apparent discontinuity near age 65 and above age 95 for some of the probability estimates.

The finding that $P(H|H)$ and $P(H|S)$ decrease with age reflects clinical experience that older adults are less likely to remain healthy or to recover from illness. $P(D|H)$, $P(D|S)$, and $P(S|H)$ increase monotonically with age,
which is also not surprising. But we did not know in advance how the probability of remaining sick would change with age. By definition, \( P(S|S) = 1 - P(H|S) - P(D|S) \), but since \( P(H|S) \) decreases with age while \( P(D|S) \) increases, the effect of aging on \( P(S|S) \) was not obvious. Here, \( P(S|S) \) increases until about age 50, and is fairly flat until about age 80, after which it declines, presumably because sick persons become more likely to die than to remain sick. The trends for age 65 and older were reasonably similar in the three data sets, with the most differences in the probability of remaining sick, \( P(S|S) \) (not shown).

**Synthetic Cohorts**

For those unfamiliar with life-table calculations, we present a simple example. Consider a cohort with a specified number of healthy and sick persons at baseline, and also specified transition probabilities among the three states: healthy, sick, and dead. For specificity, use the probabilities at age 65 (see Figure 3), which are as follows: \( P(H|H) = .90; P(S|H) = .09; P(D|H) = .01; P(H|S) = .34; P(S|S) = .61; \) and \( P(D|S) = .05 \). For an artificial cohort of 65-year-olds, of whom (say) 100 are healthy and 100 are sick at age 65, we can estimate the number who will be in each health state one year later, at age 66. Of the 100 healthy persons, the probabilities indicate \( .09 \times 100 = 9 \) will be sick one year later, 1 will be dead, and 90 will still be healthy. Of those who start out sick, 61 will remain sick, 34 will become healthy, and 5 will die. Thus, at age 66, there would be \( 90 + 34 = 124 \) healthy persons, 70 sick persons, and 6 dead persons. These calculations can be repeated at age 66, using the transition probabilities specific to that age, and repeated again until all participants have died. At that point, the total number of person-years spent in the healthy and sick states can be calculated, to yield estimates of the years of healthy life, years of sick life (morbidity), and years of life (life expectancy).

We estimated the number of healthy, sick, and dead persons over time in a birth cohort of 100,000 persons, of whom 98% were healthy at age 0 (Vital and Health Statistics 1990), as shown in Figure 4. The number healthy declines and the number dead increases, as expected. The number sick increases slowly until about age 55, is fairly flat until about age 80, and declines after that. The area under the Dead curve (divided by 100,000) is the average years lost to death = 23.8, and 100 – 23.8 is the average life expectancy, or 76.2 years. Given the limitations of the data, this is reasonably close to the life expectancy of the U.S. population (77.3 years). The area below the Healthy curve is 66.9 years of healthy life from birth to age 100. The area below the Sick curve is 9.3 years of sick life or morbidity.
The actual CHS research cohort is a mix of birth cohorts and baseline health states. It is clear from Figure 4 that a cohort consisting primarily of persons aged 55 to 80 at baseline will have a fairly constant number of sick persons over time, whereas an excess of persons below 55 will result in an increase over time, and an excess above 80 will result in a decrease. Using the starting age and health status distribution for the CHS cohort, the estimated number of sick persons in the synthetic CHS cohort was approximately constant for 14 years of follow-up, as was seen in Figure 1 for the real data.

Medical Expenditures for the Birth Cohort

In Figure 4 the number of sick persons, reasonably expected to have the highest medical expenditures, is approximately constant from age 55 to 80, while the less expensive (healthy) group declines in size over time. If medical expenditures were a function only of health state, cost for the cohort would decrease after age 55, as the number of sick persons would be constant but the number healthy would decline. However, expenditures increase with age as well as with health status (Machlin et al. 1996;
Spillman and Lubitz 2000; Miller 2001). We multiplied the estimated mean age- and health-specific cost by the projected number of persons in each health state and age. Total estimated annual costs for the birth cohort of 100,000 would increase monotonically with age after about age 10, would be fairly constant at about $420 to $430 million per year from ages 61 to 73, and would decline after that. Medical expenditures for a cohort with new entries are discussed in the next section. (We provided only crude expenditure estimates because of uncertainty in the imputed transition probabilities and the extrapolated expenditure data.)

**Summary and Discussion**

In this article, we used data from two CHS research cohorts to suggest that the number of sick persons was surprisingly stable over time, no matter how “sick” was defined. We then combined three large data sets to estimate transition probabilities, and used multistate life table methods to create a synthetic birth cohort. We also created a synthetic CHS cohort from these transition probabilities, to demonstrate that the stability noted in the CHS data could be predicted from the transition model.

**How Does the Number of Sick Persons Change Over Time?**

The number of sick persons was approximately constant in the CHS research cohort for every definition of “sick” that was considered. This finding is strengthened by the variety of variables that were considered. Timed walk was assessed at the clinic (not by self-report). Unlike the other variables, CVD (prevalent heart disease) was cumulative, in that P(H|S) is zero. Persons in Cohort 2 (all African American) were substantially sicker than those in Cohort 1, but the number sick was still fairly constant over time.

We were able to reproduce Figure 1 from the transition probabilities and the initial CHS age and health distribution. In addition, we projected that in a birth cohort there would be a constant number sick from about age 55 to 80. The increase in the number dead and the decrease in the number healthy in the birth cohort, shown in Figure 4, were expected. It now seems obvious that the number sick must increase at earlier ages (since most persons are healthy at birth) and must decrease at the end (as most persons will be dead). The large range over which the estimated number of sick persons was constant was not expected.
This long period of stability may not have been noticed earlier, because there are few lengthy longitudinal series and most analyses have dealt with the percentage of the living who were sick, rather than the percentage of all the persons in the cohort. That is, many had noted that the percentage of survivors who were sick increased over time (Spillman and Lubitz, 2000), but our finding is equivalent to saying that the percentage of the initial cohort who were sick was constant for a long period.

The strong similarity of the patterns in Figure 2 suggests that we can estimate the stable level of the number sick for the variables other than EVGGFP. For example, in a figure similar to Figure 4, but with sick defined as having ADL difficulties, we would expect the number sick in a birth cohort of 100,000 to increase over time, then to be stable for a long period, and finally to decrease, with about $756/5,201*100,000 = 14,535$ persons with ADL difficulties each year during the stable period. This supposition needs to be verified, as does the age interval over which stability occurs.

If the number of sick persons does not change (in a certain time period), this implies (under certain assumptions) that the rate of decline in the number of healthy persons is the same as the rate of decline in the number of living persons. It also implies that the proportion of future lifetime spent in the healthy (as opposed to the sick) state is the square of the proportion who are healthy at baseline (see Appendix A). These relationships could lead to a useful approximation of the number of sick persons over time and the proportion of lifetime spent in the healthy state when longitudinal health data are not available, but baseline health data and an appropriate life table are available. This possibility should be explored further.

The decrease in the medical expenditures of a birth cohort after about age 65 may provide an alternative way to think about future Medicare expenditures, although more detailed models are needed to provide better expenditure estimates. The lifetime distribution of health care costs has recently been examined by others (Alemayehu and Warner 2004).

**Transition Probabilities**

The transition probabilities were estimated from half a million transition pairs. Even so, there were relatively few sick and dying young persons and few healthy persons over age 90. The data sets were different in that positive selection bias at age 65 was a factor for CHS, the three study samples were drawn from somewhat different populations, and “healthy” was reported by the head of the household for the MEPS data. However, our only assumption was that the transition probabilities among the health
states for each age were the same in all data sets, which could be true even if the survey populations were different. (That is, a “healthy” person of a certain age may have the same transition probabilities no matter how the sample was obtained). Life expectancies calculated from the separate data sets (not shown) indicated that CHS and MEPS overestimated survival, and MCBS underestimated survival, but the combined estimates were reasonably close to the national figures. In the National Health Interview Survey, proxy respondents were found to give a more positive assessment of the health-related quality of life of others, which may explain why MEPS data produced more optimistic estimates (Stineman et al. 2004). Bias caused by imputation of the missing data is unknown but is probably small. With these caveats, we next discuss some features of the transition probabilities, which have not, to our knowledge, been published in such detail for such a large age range.

EVGGFP is strongly related to mortality (Idler and Benyamini 1997), as was found here; that is, in Figure 3, \( P(D|S) \) is considerably higher than \( P(D|H) \). Considering only the age range where neither of these death rates was based on small numbers, the estimated relative risk of death for healthy versus sick is greater than seven from age 30 to 55, and then declines to about two at age 90 (data not shown).

It is not always recognized that older adults have a meaningful probability of recovering from being sick. Figure 3 shows that for EVGGFP, \( P(H|S) \) is about 70% under age 20, about 30% at age 65, and then declines monotonically to about 10% at age 100. This meaningful rate of recovery is of interest because some statistical procedures such as the quality-adjusted time without symptoms or toxicity (Q-Twist) need to assume that recovery is not possible (Schwartz et al. 1995), which is clearly not true in a general population. Our recovery rate is consistent with rates reported by others (Gill, Hardy, and Williams 2002; Hardy et al. 2005). Estimated recovery probabilities based on other definitions of being sick are available elsewhere (Diehr et al. 2005).

It is also interesting that so many healthy persons die. Probability \( P(D|H) \) is about 1% at age 65 but increases to about 20% at age 95. (As mentioned above, this probability may have been overestimated because of the imputation of some missing data just before death.) As these are one-year transitions, some healthy persons may have fallen sick and then died within the year. However, 18% of the CHS enrollees who died reported being healthy within three months of death, suggesting that a substantial number may have died while assessing themselves as healthy (Diehr et al. 2001).
Implications for Medicare

Because two of the data sets used here were samples from the Medicare population, some of our findings may apply directly to Medicare. The group of current Medicare enrollees may be thought of as a type of research cohort (with a mix of ages), suggesting that for this group, the number of sick persons will be approximately constant for the next 14 years, no matter how sick is defined. Medicare, of course, enrolls a new birth cohort of 65-year-olds each year. If the number of new enrollees per year (M) is approximately constant over time, and the age-specific transition probabilities do not change, it is easy to show that the number of sick persons enrolled in Medicare at any time will be M times the area under the “# Sick” curve from age 65 to 100 (we estimate about 4.5 * M) (see Appendix B). Similarly, the number of healthy persons in any year will approach M times the area under the “# Healthy” curve (about 12.6 * M). For example, if there are two million new enrollees every year, and 20% of them are sick at enrollment, then at any time there will be about 9.0 million sick enrollees and 25.2 million healthy enrollees. This is an oversimplification because M is expected to increase substantially in the near future, but demonstrates in a different way the stability of the number of sick persons. It would be easy to perform this calculation for M that increased over time, but we have not done so.

Limitations

The longitudinal data in Figure 1 and Figure 2 came from a single study. Additional lengthy longitudinal datasets should be examined. There was not perfect agreement among the three datasets, particularly at the oldest ages, and some missing data were imputed. CHS and MEPS data overestimated the national life expectancy, and MCBS data underestimated it. We did not use survey weights, in part because persons followed over time had different weights at each survey wave. This limits our claims of generalizability, but probably had little effect since we were estimating only transition probabilities from these data.

Because our goal was to gain insight into the process, rather than to provide detailed probability or projection estimates, we made many simplifications. We restricted analysis to only three health states, and ignored gender, which is related to transition probabilities (Diehr and Patrick 2001). We ignored other important information, such as the person’s socio-economic status, prior health, or the presence of chronic or acute illness. Although such information would greatly improve estimates of transition for an individual, it was not needed for this article, which required only
estimates of the average transition probabilities at each age. For example, if the healthy state at baseline a contained two equal subgroups, one with 90% probability of remaining healthy one year later and one with a 70% probability of remaining healthy, our estimated transition probability would be 0.8, and 80% would indeed remain healthy one year later, even though 0.8 was not the correct probability for any person in the group. The transition probabilities at each age should be reasonably accurate, since they were computed using data from general populations. Some of the change over time in the probabilities is due to aging, and some due to the changing makeup of each health state.

The cohort projections make the standard assumption that the transition probabilities are stable over time (no birth cohort effect), but this may not be the case (Freedman, Martin, and Schoeni 2002; Freedman et al. 2004). A cohort initiated in 2007 and followed until death might not have the same number in the health states if the probabilities change substantially during their lifetime. However, the estimated probabilities were able to reproduce Figure 1, and the estimated life expectancy was close to national values, suggesting that the projections are reasonably valid. Short-term validity of similar estimates has been demonstrated elsewhere (Diehr and Patrick 2001; Diehr et al. 1998).

We calculated transition probabilities for only one measure of sickness, but Figure 2 suggests that similar results will be found when using other definitions of “sick.” This conjecture needs verification. Although different populations may have different transition probabilities, we have already in a sense considered some different transition probabilities by considering different definitions of being sick. Our understanding is limited in that the transition probabilities were calculated by brute force, without an underlying theoretical model. Further research should include development of a theory-based model that also takes sex, prior health, and incident health events into account.

Conclusion

We found that the number of persons in fair or poor health is low at birth, reaches a maximum, and is approximately constant from ages 55 to 80, after which it declines. The number of sick persons does not increase in old age, but rather the number of healthy persons declines. This results in a constant number of sick persons over time in “research” cohorts of older adults for 14 years and perhaps longer. This phenomenon is likely to hold when
“sick” is defined in different ways. Medical expenditures for the cohort are highest at about age 65 after which they decline, because the smaller number of people remaining alive in the cohort more than offsets the rising expenditures per surviving person. These interesting findings may be useful in conceptualizing how the health and medical expenditures of a population or a cohort will change over time. It might also suggest that health promotion and disease prevention would be the best public health strategies at younger ages because there are relatively few sick persons to be cured or prevented from dying (Diehr et al. 2007). The approximations in Appendix A may also be useful when longitudinal data are not available. Further research in these areas would be welcome.

Appendix A

We found that the number of sick persons was approximately constant over a large range of ages. This finding can be expressed in terms of $A_t$, $H_t$, and $S_t$, the number alive, healthy, or sick at time $t$, where $t$ is 0 at baseline (at enrollment in the research cohort, or perhaps at age 55 in a life table calculation for older adults). Our main finding is that for older adults, $S_t \sim S_0$. This has some interesting consequences. Suppose that at least approximately, $A_t$ is linear in time, or $A_t = a + bt$. At $t = 0$, $A_0 = a$, so the equation can be rewritten as

$$A_t = A_0 + bt$$

[1]

and $b$ will be negative. This line reaches zero (no-one alive) when $t = -A_0/b$. Life expectancy is then the area under the curve from $t = 0$ to $-A_0/b$, and because this is just a right triangle, the area is

life expectancy $= -A_0^2 / (2b)$. [2]

Because the number alive is the number sick plus the number healthy, equation [1] can be written as $H_t + S_t = A_0 + bt$. Because $S_t \sim S_0$, $H_t + S_0 = A_0 + bt$, and $H_t = A_0 - S_0 + bt$. Because $A_0 - S_0 = H_0$,$H_t = H_0 + bt$. [3]

Thus, our findings predict approximately the number healthy at any time if we know the number healthy at time zero and have life table data to estimate $b$. The area under the curve (healthy life expectancy, or years of healthy life) will be
The proportion of future life that is spent in the healthy state is $\frac{YHL}{\text{years alive}}$, or

$$\text{Proportion healthy} = \frac{H_0^2}{A_0^2}. \quad [5]$$

Our findings thus suggest that in the relevant age range, the number of healthy people (by whatever definition) will decrease at the same rate, $b$, as the number alive. They also suggest that the square of the proportion who are healthy at baseline, by whatever definition, is an estimate of the proportion of future life that will be spent in that healthy state.

In the Cardiovascular Health Study (CHS) data shown in Figure 2, the slopes of the number healthy, $b$ (row-wise, respectively) are $-.046$, $-.028$, $-.035$, $-.036$, $-.037$, $-.040$, $-.037$, $-.035$, $-.033$. The slope for survival was $-.034$ in cohort 1 and $-.039$ in cohort 2. Thus, the slopes are fairly similar but not identical. A more detailed model might find either more or fewer similarities. These simple estimates may provide ballpark estimates when longitudinal health data are not available.

**Appendix B**

This article has discussed primarily the number of sick persons in a closed cohort. The number of sick persons in Medicare, which has a new entering group of 65-year-olds every year, can also be estimated if the number of new enrollees each year is approximately constant. This calculation requires an appropriate multistate life table that gives the proportion of persons alive, sick, and healthy at age $t$ (respectively, $A_t$, $H_t$, and $S_t$). For Medicare all persons are alive at age 65, and $A_{65} = 1$. At all ages, $H_t + S_t = A_t$. Let $M$ be the number of new enrollees each year, assumed to be approximately constant over time.

After equilibrium is reached, the number of current Medicare beneficiaries can be estimated as follows. There will be $M^*A_{65}$ 65-year-olds, the newest enrollees. There will be only $M^*A_{66}$ 66-year-olds, those who enrolled one year earlier at age 65 and are still alive. There will be $M^*A_{67}$ 67-year-olds, who enrolled two years earlier at age 65, and so on. The total number of enrollees is thus $\sum M^*A_t = M^* \sum A_t$ (where 150 was chosen as an age at which all persons will surely be dead). Life expectancy can be written as $\sum A_t$. Under certain conventions 0.5 would be subtracted from this quantity,
but we have omitted that term here for simplicity.) The total number of Medicare beneficiaries in a particular year will thus be approximately $M^\text{life expectancy}$. The number of healthy persons or sick persons in Medicare is obtained by replacing $A_t$ by $H_t$ or $S_t$ respectively in the above calculations. For example, assume that $M = 2$ million new enrollees per year, life expectancy from age 65 = 17.1 years, expected years of healthy life from age 65 = 12.6 years, and expected years of sick life = 4.5 years (Diehr et al. 2007). Under these assumptions, the number of current beneficiaries in Medicare who are sick is expected to be 4.5 times 2 million, or 9 million, while the number of healthy beneficiaries is 25.2 million and the total number of current beneficiaries is $25.2 + 9 = 34.2$ million. These numbers will hold as long as the life table remains accurate and $M$ is approximately constant.

If $M$ is not constant (as is expected in the future) this calculation would require a different value of $M$ for each age cohort, and would not be so simple. If $M$ increases over time, the total number of beneficiaries and the proportion who are healthy will increase, and the proportion who are sick will decrease, because there will be relatively more beneficiaries at younger ages. For example, suppose Medicare enrolled $M$ new persons this year, but only $(1-\alpha)M$ last year, $(1-2\alpha)M$ the year before that, $(1-3\alpha)M$ the year before that, and so on. In that case, 26.6% of all current beneficiaries would be sick if $\alpha = 0$ (if $M$ is constant), 25.5% would be sick if $\alpha = 0.02$, and 22.2% would be sick if $\alpha = 0.05$. Similarly, if the number of new enrollees is decreasing with time, the proportion of all current beneficiaries who are sick will become larger.

Note

1. Participants in the Cardiovascular Health Study are as follows: Wake Forest University School of Medicine: Gregory L. Burke, MD. Wake Forest University, ECG Reading Center: Pentti M. Rautaharju, MD, Ph.D. 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, Ph.D., 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. Gotttdiener, MD. New England Medical Center, Boston, Ultrasound Reading Center: Daniel H. O’Leary, MD. University of Vermont, Central Blood Analysis Laboratory: Russell P. Tracy, Ph.D. University of Arizona, Tucson, Pulmonary Reading Center: Paul Enright, MD. Retinal Reading Center, University of Wisconsin: Ronald Klein, MD. University of Washington, Coordinating Center: Richard A. Kronmal PhD. NHLBI Project Office: Jean Olson, MD, MPCTH.
References


**Paula Diehr**, PhD, MS, is a professor of biostatistics and health services in the School of Public Health at the University of Washington, in Seattle. Her earlier work included diagnosis of community-acquired pneumonia and ankle fracture, small-area variation analysis, health insurance and use and cost of services, and community randomized trials for health promotion. She is currently involved in issues surrounding aging, such as obesity, exercise, diet, years of healthy life, and quality of life at the end of life. Her current statistical methodology interest is in methods for accounting for death in longitudinal data where many people die.

**Ann Derleth**, PhD, MSPH, is currently a postdoctoral fellow at the VA Puget Sound Health Services Research & Development Center of Excellence. Her work has focused on transforming health status measures to obtain single index measures of health-related quality of life and predicting years of healthy life.
Anne B. Newman, MD, MPH, is a professor of epidemiology and medicine with a joint appointment in the Division of Geriatric Medicine at the University of Pittsburgh. Her research focuses on the epidemiology and implications of subclinical diseases in older adults and its contribution to frailty, a concept that she has helped to define. She has helped to pioneer the application of clinical tools to the evaluation of entire populations. Dr. Newman was recently appointed as director of the Center on Aging and Population Health.

Liming Cai, PhD, is an economist and senior service fellow in the National Center for Health Statistics, Office of Analysis and Epidemiology. His research interest is in modeling health changes at older ages and understanding the demographic and economic implication of aging.