

Development and Validation of a Prognostic Index for 4-Year Mortality in Older Adults

Sei J. Lee, MD

Karla Lindquist, MS

Mark R. Segal, PhD

Kenneth E. Covinsky, MD, MPH

PROGNOSTIC INFORMATION IS valuable to many of the stakeholders in the US health care system. For patients and caregivers, prognostic information is needed to inform clinical decision making.¹ For example, patients with limited life expectancy may benefit more from advance care planning and discussions on the goals of care.^{2,3} Older patients with better than average life expectancy may be more likely to benefit from cancer screening since recent guidelines suggest clinicians target screening in the elderly to those with life expectancies greater than 5 years.⁴⁻⁶ For policymakers, prognostic information is essential when comparing the quality of care between different health care organizations, such as hospitals and insurance plans. Accurate risk adjustment levels the playing field by accounting for the differences in health status of the underlying populations making fair comparisons between health care organizations possible. Finally, prognostic information is essential for epidemiological studies.⁷ Observational studies require accurate risk adjustment so that baseline differences do not confound the relationship between the risk factor of interest and the outcome.

Despite the clinical, policy, and epidemiological importance of prognostic information, there are few prognostic indexes that can easily be used in large seg-

Context Both comorbid conditions and functional measures predict mortality in older adults, but few prognostic indexes combine both classes of predictors. Combining easily obtained measures into an accurate predictive model could be useful to clinicians advising patients, as well as policy makers and epidemiologists interested in risk adjustment.

Objective To develop and validate a prognostic index for 4-year mortality using information that can be obtained from patient report.

Design, Setting, and Participants Using the 1998 wave of the Health and Retirement Study (HRS), a population-based study of community-dwelling US adults older than 50 years, we developed the prognostic index from 11 701 individuals and validated the index with 8009. Individuals were asked about their demographic characteristics, whether they had specific diseases, and whether they had difficulty with a series of functional measures. We identified variables independently associated with mortality and weighted the variables to create a risk index.

Main Outcome Measure Death by December 31, 2002.

Results The overall response rate was 81%. During the 4-year follow-up, there were 1361 deaths (12%) in the development cohort and 1072 deaths (13%) in the validation cohort. Twelve independent predictors of mortality were identified: 2 demographic variables (age: 60-64 years, 1 point; 65-69 years, 2 points; 70-74 years, 3 points; 75-79 years, 4 points; 80-84 years, 5 points, >85 years, 7 points and male sex, 2 points), 6 comorbid conditions (diabetes, 1 point; cancer, 2 points; lung disease, 2 points; heart failure, 2 points; current tobacco use, 2 points; and body mass index <25, 1 point), and difficulty with 4 functional variables (bathing, 2 points; walking several blocks, 2 points; managing money, 2 points, and pushing large objects, 1 point). Scores on the risk index were strongly associated with 4-year mortality in the validation cohort, with 0 to 5 points predicting a less than 4% risk, 6 to 9 points predicting a 15% risk, 10 to 13 points predicting a 42% risk, and 14 or more points predicting a 64% risk. The risk index showed excellent discrimination with a *c* statistic of 0.84 in the development cohort and 0.82 in the validation cohort.

Conclusion This prognostic index, incorporating age, sex, self-reported comorbid conditions, and functional measures, accurately stratifies community-dwelling older adults into groups at varying risk of mortality.

JAMA. 2006;295:801-808

www.jama.com

ments of the population. Many have focused on specific segments of the population such as hospitalized elders⁸⁻¹⁰ or veterans.¹¹ Other indexes have focused on single domains of risk, such as comorbidity¹² or function.^{13,14} Finally, some indexes require laboratory or performance

Author Affiliations: Division of Geriatrics, San Francisco Veterans Affairs Medical Center (Drs Lee and Covinsky and Ms Lindquist) and Division of Geriatrics (Drs Lee and Covinsky and Ms Lindquist) and Department of Epidemiology and Biostatistics (Dr Segal), University of California, San Francisco.

Corresponding Author: Sei J. Lee, MD, Division of Geriatrics, San Francisco Veterans Affairs Medical Center, San Francisco, CA 94121 (sei.lee@va.gov).

testing, which is costly, time-intensive, and not always available.^{8,15}

To address these issues, we developed and validated a 4-year mortality prediction index from a representative sample of the US population older than 50 years. We considered and combined demographic, comorbidity, behavioral, and functional predictors into our final index. To maximize the usability of the index in a variety of settings, we relied solely on patient report. Our goal was to develop a single prognostic index that could be used for community-dwelling individuals older than 50 years for clinical, health policy, and epidemiological purposes.

METHODS

Participants

We studied community-dwelling participants interviewed in 1998 as part of the Health and Retirement Survey (HRS), which was initiated in 1992 and was expanded in 1998 to become a representative sample of all persons in the contiguous United States older than 50 years.^{16,17} Data were collected primarily through telephone interviews, with an overall response rate of 81%.¹⁸

A total of 20 447 persons were enrolled in HRS in 1998. We excluded 429 of those residing in nursing homes and 308 whose vital status could not be determined in 2002. This yielded a final sample size of 19 710. Of this final sample, 2433 individuals died by 2002. We developed the model from the eastern, western, and central regions of the country ($n=11\ 701$) with a total of 1361 deaths. We then validated the model in participants from the southern region ($n=8009$), which included 1072 deaths.^{8,13,19} We chose our validation cohort to be geographically distinct from our development cohort to test our model's geographic transportability as well as our model's accuracy.¹⁹

Measures

The outcome of interest was death by December 31, 2002. We assessed mortality using the HRS follow-up procedures, which entailed cross-referencing HRS information with the National

Center for Health Statistics National Death Index to determine vital status.²⁰

We considered 3 classes of variables as potential predictors of mortality: demographic variables, behavior and comorbidity variables, and functional status variables. In the demographic category, we examined sex and age. The relationship between age and mortality was exponential, with an additional year of age providing more mortality risk among older individuals. However, we found that categorizing age into 5-year intervals up to age 85 years resulted in a model that was easier to interpret than the exponential model with minimal losses in discrimination (c statistic 0.819 vs 0.823). When categorized, the 2 youngest age groups had similar mortality. Thus, these groups were combined to form 1 reference age group (50-59 years).

We considered a total of 18 behavioral and comorbidity variables: current tobacco use, alcohol use, hypertension, diabetes mellitus, non-skin cancers, chronic lung disease, heart failure, other heart problems, stroke, psychiatric disease, memory-related disease, arthritis, history of falls, history of pain, incontinence, visual or hearing impairment, and body mass index (BMI) of less than 25. (Body mass index is calculated as weight in kilograms divided by the square of height in meters.) We dichotomized BMI because our multivariate analysis showed that when other variables were considered, more extreme values of BMI did not improve the performance of our model. Disease status was determined by asking participants, "Have you ever had or has a doctor told you that you have/had X?" Initially some comorbidity variables were coded to represent the severity of disease. However, all variables were ultimately dichotomized into disease present and disease absent categories when multivariate analysis showed that severity levels did not yield significantly different risks of mortality. Participants were asked to self-identify their race with the question, "Do you consider yourself to be primarily white or Caucasian, black or African American, American Indian, Asian, Hispanic or Latino, or some-

thing else?" We combined American Indian, Asian, and other categories into "Other" due to the small numbers in those categories. We considered all participants who self-identified as Hispanic to be Hispanic regardless of whether they also identified as white, black, or other.

Participants were asked whether they had difficulty with 21 functional measures. These included 6 activities of daily living (bathing, dressing, toileting, eating, transferring, and walking across the room); 5 instrumental activities of daily living (shopping, preparing meals, using the telephone, managing medications, and managing finances); and 10 other functional variables derived from the Rosow-Breslau Functional Health Scale and the Nagi Index²¹⁻²³ (getting up from a chair; walking several blocks; pushing or pulling heavy objects; climbing a flight of stairs; stooping, kneeling, or crouching; picking up a dime; reaching above one's shoulders; lifting 10 lb; using a map; and vigorous physical activity). These functional domains have been shown in previous studies to be important predictors of institutional care and death.^{9,13,14}

In developing our model, we did not consider some variables that, although associated with mortality, could have reduced the generalizability or usability of our index. Specifically, we did not include race or socioeconomic status because their association with mortality may be due, at least in part, to lower quality of care. Since risk adjustment indexes are often used to measure quality of care, including race and socioeconomic status would have precluded the use of our model for risk adjustment.⁷ We also did not include the number of hospitalizations and surgeries because of marked regional variability in these rates among patients with similar disease severity.²⁴

Statistical Analysis

Using mortality as the dichotomous outcome variable, we measured the bivariable relationship between each of the 41 risk factors and mortality within the development cohort. Because all pro-

posed variables were significant predictors of mortality in the bivariable analysis, all variables were analyzed using multivariable logistic regression. We used backward elimination ($P < .05$ to retain) to determine which variables remained independent predictors of mortality. This process yielded a model with 19 independent risk factors for mortality.

Because of our large sample size, many predictors were found to be significantly associated with mortality but only marginally improved the predictive accuracy of the model. To simplify the model further, while minimizing losses in predictive ability, we subjected all remaining predictors to further selection with the Schwarz Bayesian Information Criterion (BIC).^{25,26} The BIC penalizes the log likelihood of a model (a measure of its fit) by a factor related to the number of predictor variables in the model (a measure of its complexity). This provides an objective way of assessing whether the improved fit provided by an additional variable is justified by the added complexity of the model.²⁶ Applying the BIC criteria to our model yielded the final model with 12 predictor variables.

To test the stability of our final model, we tried alternate methods to determine whether the resultant model would differ from our original final model. First, we used forward and bidirectional selection techniques instead of backward selection. These alternative strategies resulted in differences of a few variables at the early modeling stages, but these different variables were eliminated in the BIC selection step leading to the same final model. Second, we bypassed the backward elimination steps, subjecting all 41 of the initial variables to BIC selection. This also resulted in the same final 12-variable model. Third, we derived the model in stages, in which the functional variables were initially modeled separately from the comorbidity variables. The resultant submodels were then merged into single model, which was then subjected to BIC selection. This staged approach is an accepted method of variable reduction and has been used successfully in the development of many previous prognostic indexes.^{9,13,27-29} Af-

ter the BIC selection step, this staged approach resulted in the same final 12-variable model. Fourth, prespecified

interactions between age and other predictors were checked, and these terms were also eliminated in the stepwise se-

Table 1. Characteristics of the Health and Retirement Study Participants Included in This Study*

| Characteristics | No. (%) | |
|--|-----------------------------|--------------------------|
| | Development (n = 11 701) | Validation (n = 8009) |
| Demographics | | |
| Age, y | | |
| 50-59 | 3154 (27) | 2328 (29) |
| 60-64 | 2145 (18) | 1547 (19) |
| 65-69 | 1798 (15) | 1188 (15) |
| 70-74 | 1669 (14) | 1017 (13) |
| 75-79 | 1362 (12) | 909 (11) |
| 80-84 | 856 (7) | 533 (7) |
| ≥85 | 717 (6) | 487 (6) |
| Women | 6639 (57) | 4516 (56) |
| Race or ethnicity | | |
| White, non-Hispanic | 9467 (81) | 5656 (71) |
| Black, non-Hispanic | 1225 (10) | 1484 (19) |
| Other, non-Hispanic | 225 (2) | 149 (2) |
| Hispanic | 772 (7) | 706 (9) |
| Less than a high school education | 2845 (24) | 2745 (34) |
| Comorbidities and behaviors | | |
| Hypertension | 5366 (46) | 3960 (49) |
| Diabetes mellitus | 1608 (14) | 1246 (16) |
| Cancer | 1349 (12) | 862 (11) |
| Lung disease | 443 (4) | 376 (5) |
| Heart failure | 298 (3) | 240 (3) |
| Coronary artery disease | 2032 (17) | 1481 (19) |
| Stroke or transient ischemic attack | 754 (6) | 602 (8) |
| Dementia | 189 (2) | 179 (2) |
| Incontinence | 1881 (16) | 1292 (16) |
| BMI <25 | 4402 (38) | 2981 (37) |
| Current smoker | 1813 (16) | 1416 (18) |
| Functional measures | | |
| Activities of daily living difficulty | | |
| Bathing | 754 (6) | 644 (8) |
| Dressing | 1014 (9) | 848 (11) |
| Eating | 312 (3) | 308 (4) |
| Toileting | 575 (5) | 496 (6) |
| Transferring | 863 (7) | 715 (9) |
| Walking across room | 1280 (11) | 1068 (13) |
| Instrumental activities of daily living difficulty | | |
| Shopping | 1307 (11) | 1118 (14) |
| Preparing meals | 1154 (10) | 990 (12) |
| Using the telephone | 463 (4) | 485 (6) |
| Managing medications | 307 (3) | 276 (3) |
| Managing finances | 887 (8) | 737 (9) |
| Other measures of functional status, difficulty | | |
| Walking several blocks | 3167 (27) | 2542 (32) |
| Pushing or pulling heavy objects | 3329 (29) | 2604 (33) |
| Climbing stairs | 5324 (46) | 4234 (53) |

Abbreviations: BMI, body mass index, which is calculated as weight in kilograms divided by the square of height in meters.

*Percentages may not sum to 100 due to rounding.

Table 2. Bivariable Analysis of Risk Factors and 4-Year Mortality in Development Cohort*

| Risk Factor | No. (%) of Deaths | Odds Ratio (95% Confidence Interval) |
|---|-------------------|---|
| Demographics | | |
| Age, y | | |
| 50-59 | 95 (3) | 1.0 |
| 60-64 | 124 (6) | 2.0 (1.5-2.6) |
| 65-69 | 155 (9) | 3.0 (2.3-3.9) |
| 70-74 | 199 (12) | 4.4 (3.4-5.6) |
| 75-79 | 239 (18) | 6.9 (5.4-8.8) |
| 80-84 | 224 (26) | 11.4 (8.8-14.7) |
| ≥85 | 325 (45) | 26.7 (20.8-34.3) |
| Sex | | |
| Women | 686 (10) | 1.0 |
| Men | 675 (13) | 1.3 (1.2-1.5) |
| Comorbidities and behaviors | | |
| Diabetes mellitus | | |
| Absent | 1053 (10) | 1.0 |
| Present | 308 (19) | 2.0 (1.8-2.3) |
| Cancer | | |
| Absent | 1060 (10) | 1.0 |
| Present | 299 (22) | 2.5 (2.2-2.9) |
| Lung disease | | |
| Absent | 1197 (11) | 1.0 |
| Present | 164 (37) | 4.9 (4.0-6.0) |
| Heart failure | | |
| Absent | 1235 (11) | 1.0 |
| Present | 120 (40) | 5.5 (4.4-7.0) |
| Coronary artery disease | | |
| Absent | 990 (10) | 1.0 |
| Present | 371 (18) | 2.0 (1.7-2.2) |
| Dementia | | |
| Absent | 1271 (11) | 1.0 |
| Present | 90 (48) | 7.3 (5.5-9.8) |
| Body mass index† | | |
| ≥25 | 648 (9) | 1.0 |
| ≤24.9 | 713 (16) | 2.0 (1.8-2.2) |
| Tobacco | | |
| Never and former | 1123 (11) | 1.0 |
| Current | 238 (13) | 1.2 (1.0-1.4) |
| Functional measures | | |
| Activities of daily living—bathing | | |
| No difficulty | 1032 (9) | 1.0 |
| Difficulty or needs help | 325 (43) | 7.3 (6.2-8.5) |
| Instrumental activities of daily living | | |
| Preparing meals | | |
| No difficulty | 973 (9) | 1.0 |
| Difficulty, can't, or don't | 386 (33) | 4.9 (4.3-5.7) |
| Using the telephone | | |
| No difficulty | 1149 (10) | 1.0 |
| Difficulty, can't, or don't | 210 (45) | 7.3 (6.0-8.8) |
| Managing finances | | |
| No difficulty | 1035 (10) | 1.0 |
| Difficulty, can't, or don't | 323 (36) | 5.4 (4.6-6.3) |
| Other measures of functional status | | |
| Walking several blocks | | |
| No difficulty | 546 (6) | 1.0 |
| Difficulty, can't, or don't | 813 (26) | 5.0 (4.5-5.7) |
| Pushing heavy objects | | |
| No difficulty | 614 (7) | 1.0 |
| Difficulty, can't or don't | 746 (22) | 3.6 (3.2-4.1) |
| Climbing stairs | | |
| No difficulty | 378 (6) | 1.0 |
| Difficulty, can't, or don't | 980 (18) | 3.6 (3.1-4.0) |
| Vigorous physical activity | | |
| Yes | 303 (6) | 1.0 |
| No | 1058 (16) | 3.0 (2.6-3.4) |

*Due to missing comorbidity and functional data, the sum of deaths for each variable may vary from 1355 to 1361.

†Body mass index is calculated as weight in kilograms divided by the square of height in meters.

lection and BIC selection processes, leading to the same final model.³⁰ Fifth, 6 collinear predictor variable pairs were identified. However, replacement of the selected variable with the collinear variable resulted in a poorer fit in every case. Thus, we reverted back to our original 12-predictor model. Sixth, we used the number of comorbid conditions in place of the individual comorbidities and found the resultant model to be slightly less accurate. The final model was checked using the Hosmer-Lemeshow goodness-of-fit test with 10 groups, which showed the absence of a gross lack of fit ($P=.27$).

To determine the performance of our index across age groups, we applied our index to 3 age subgroups. We attempted to maintain a similar number of deaths in each subgroup, which resulted in our youngest subgroup having many more participants than our older subgroups. To maximize power, we combined the development and validation cohorts for this analysis. We also dropped age as a predictor in our point score for this analysis because age was used to divide our sample.

We assessed the predictive accuracy of the final model by looking at the 2 components of accuracy: calibration and discrimination. Calibration of the model was assessed by comparing the predicted mortality with the actual mortality in the development and validation cohorts. The discrimination of the model was assessed by calculating the receiver operating characteristic (ROC) curves for the development and validation cohorts.

We describe the results of our final model in 2 ways. First, we determined the predicted risk for each participant using the coefficients from the final logistic regression model. We then divided the sample into quartiles of risk and compared the mortality rates in the development and validation cohorts. Second, a point-based risk scoring system was developed in which points were assigned to each risk factor by dividing each β coefficient by the lowest β coefficient (ability to push or pull heavy objects) and rounding to the

nearest integer. A risk score was assigned to each participant by summing the points for each risk factor present.^{8,30} We compared the mortality rates by point score in the development and validation cohorts.

The Human Subjects Committee of the University of California, San Francisco, and the San Francisco Veterans Affairs Research and Development Committee approved this study. All statistical analyses were performed using Intercooled Stata software version 8.2 (Stata Corporation, College Station, Tex).

RESULTS

Characteristics of the Participants

The mean (SD) age of participants in the development cohort was 67 (10) years. Fifty-seven percent were women, 81% were white, and 10% were black. Twenty-five percent reported completing less than a high school education. Sixteen percent reported difficulty in 1 or more activities of daily living and 12% reported difficulty in at least 1 instrumental activities of daily living (TABLE 1). During 4-year follow-up, 1361 participants (12%) died, leading to a total of 40 471 person-years of observation. The mean (SD) age of participants in the validation cohort was 67 (10) years. Fifty-six percent were women, 71% were white, and 19% were black. Thirty-four percent reported completing less than a high school education. Eighteen percent reported difficulty in 1 or more activities of daily living and 16% reported difficulty in at least 1 instrumental activities of daily living (Table 1). During 4-year follow-up, 1072 participants (13%) died, leading to a total of 27 647 person-years of observation.

Bivariable Results

All risk factors were associated with 4-year mortality in the bivariable analyses ($P < .05$; TABLE 2). Age was the most powerful predictor of mortality, yielding an ROC area of 0.75 in the development cohort.

Multivariable Results

Because all 41 proposed predictor variables were significant predictors

of mortality in the bivariable analysis, all variables were entered into a multivariable logistic regression using stepwise backward selection. This resulted in 19 variables remaining as significant predictors of mortality. Application of the BIC criteria to this model resulted in a final model of 12 independent predictors with only a small loss in discrimination, as seen by the minimal drop in ROC curve area between the 19-term model (ROC area = 0.85) vs the 12-term model (ROC area = 0.84).

The final model with 12 predictors included 2 measures of demographics (age and sex), 6 measures of comorbidities and behaviors (diabetes mellitus, cancer, chronic lung disease, heart failure, current tobacco use, and $BMI < 25$), and 4 measures of functional difficulty (bathing, walking several blocks, managing finances, and pushing or pulling heavy objects; TABLE 3). Former tobacco use was not independently associated with mortality after adjustment for comorbid conditions and functional measures.

Risk Stratification by Quartiles

When divided into quartiles of risk, the 4-year mortality ranged from 1% in the lowest-risk quartile to 33% in the highest-risk quartile in the development cohort and 2% to 34% in the validation cohort (TABLE 4). The calibration of the model was good, with close agreement between the observed mortality in the development and validation cohorts. The discrimination of the model was also good, with an ROC area of 0.82 in the development cohort and 0.84 in the validation cohort.

Risk Stratification by Points

The points assigned to each of the final 12 predictor variables are listed in Table 3. A risk score was calculated for each participant by adding the points for each risk factor present. For example, a 78-year-old (3 points) non-smoking (0 points) woman (0 points) with a BMI of 28 (0 points) with diabetes (1 point) and difficulty managing finances (2 points) would have a

Table 3. Independent Risk Factors for 4-Year Mortality in the Development Cohort in the Multivariable Analysis

| Risk Factor | Adjusted OR (95% CI)* | Points |
|-------------------------------|--------------------------|--------|
| Demographics | | |
| Age, y | | |
| 60-64 | 1.9 (1.4-2.5) | 1 |
| 65-69 | 2.8 (2.1-3.7) | 2 |
| 70-74 | 3.7 (2.8-4.9) | 3 |
| 75-79 | 5.4 (4.1-7.1) | 4 |
| 80-84 | 8.3 (6.3-11.0) | 5 |
| ≥85 | 16.2 (12.2-21.6) | 7 |
| Male sex | 2.0 (1.8-2.3) | 2 |
| Comorbidities and behaviors | | |
| Diabetes mellitus | 1.8 (1.5-2.1) | 1 |
| Cancer | 2.1 (1.7-2.4) | 2 |
| Lung disease | 2.3 (1.8-2.9) | 2 |
| Heart failure | 2.3 (1.8-3.1) | 2 |
| $BMI < 25$ | 1.7 (1.4-1.9) | 1 |
| Current smoker | 2.1 (1.7-2.5) | 2 |
| Functional measures | | |
| Bathing | 2.0 (1.6-2.4) | 2 |
| Managing finances | 1.9 (1.6-2.3) | 2 |
| Walking several blocks | 2.1 (1.8-2.4) | 2 |
| Pushing/pulling heavy objects | 1.5 (1.3-1.8) | 1 |

Abbreviation: BMI , body mass index, which is calculated as weight in kilograms divided by the square of height in meters; CI, confidence interval; OR, odds ratio.

*Each OR was adjusted for the other risk factors in the table.

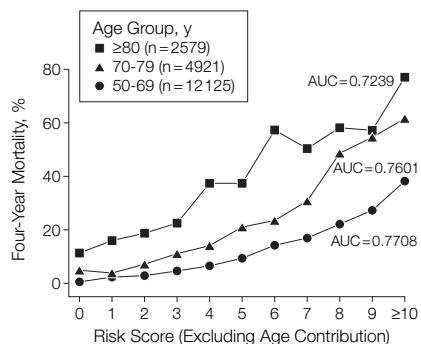
total risk score of 6 points. Development cohort risk scores ranged from 0 to 22 (mean [SD], 5.3 [3.5]). Validation cohort risk scores ranged from 0 to 23 (mean [SD], 5.5 [3.6]).

The point score effectively divided the cohort into groups at varying risk of 4-year mortality. In the development group, the mortality risk ranged from 3% in those with 0 to 5 points, 15% in those with 6 to 9 points, 40% in those with 10 to 13 points, and 67% in those with 14 or more points. In the validation cohort, the mortality risk ranged from 4% in those with 0 to 5 points, 15% in those with 6 to 9 points, 42% in those with 10 to 13 points, and 64% in those with 14 or more points (Table 4). The point-based model showed excellent discrimination, with a c statistic of 0.84 in the development cohort and 0.82 in the validation cohort. Translating our risk model into a point-based index resulted in very small losses in discrimination (c statistic of 0.819 vs 0.817). The model was also well calibrated with close

Table 4. Validation of the Prognostic Index: Comparing the Model Performance by Quartiles of Risk and Point Score

| Risk Strata | Development Cohort | | Validation Cohort | |
|------------------|----------------------------------|----------------------------------|----------------------------------|-----------------|
| | No. Who Died/ No. at Risk (%) | No. Who Died/ No. at Risk (%) | No. Who Died/ No. at Risk (%) | No. at Risk (%) |
| Quartile of risk | | | | |
| 1 | 24/2853 (1) | 31/1934 (2) | | |
| 2 | 123/3086 (4) | 112/2042 (5) | | |
| 3 | 232/2780 (8) | 187/1822 (10) | | |
| 4 | 968/2933 (33) | 733/2175 (34) | | |
| ROC area | 0.842 | 0.819 | | |
| Point score | | | | |
| 0 | 2/488 (0) | 2/355 (1) | | |
| 1 | 5/742 (1) | 10/494 (2) | | |
| 2 | 13/1368 (1) | 15/889 (2) | | |
| 3 | 48/1480 (3) | 35/973 (4) | | |
| 4 | 58/1448 (4) | 61/994 (6) | | |
| 5 | 68/1334 (5) | 65/845 (8) | | |
| 6 | 104/1168 (9) | 72/762 (9) | | |
| 7 | 110/890 (12) | 96/638 (15) | | |
| 8 | 146/762 (19) | 99/501 (20) | | |
| 9 | 133/552 (24) | 79/404 (20) | | |
| 10 | 112/409 (27) | 87/310 (28) | | |
| 11 | 139/324 (43) | 106/233 (45) | | |
| 12 | 118/244 (48) | 85/192 (44) | | |
| 13 | 95/177 (54) | 94/159 (59) | | |
| ≥14 | 210/315 (67) | 166/260 (64) | | |
| ROC area | 0.840 | 0.817 | | |

Abbreviation: ROC, receiver operating characteristic.

Figure. Four-year Mortality by Risk Score in Differing Age Groups

AUC indicates area under the curve.

agreement between the mortality rates in the development and validation cohorts for various levels of risk (Table 4).

Risk Stratification in Special Populations

When the point score (excluding the points assigned to age) was evaluated in different age groups, we found that

our index discriminated well in all 3 of the age subgroups, with *c* statistics ranging from 0.77 to 0.72 (FIGURE). The *c* statistic in the validation cohort with the point-based model was 0.79 among black participants vs 0.84 among whites. Similarly, those with less than a high school education had a *c* statistic of 0.80 vs 0.83 for those with a high school diploma.

COMMENT

We developed and validated a prognostic index that can be used in the clinical setting to stratify patients 50 years or older into high-, intermediate-, and low-risk groups for 4-year mortality. Our index is easy to use and can be obtained in a few minutes with an interview or intake form (Box), with no need for medical record review or laboratory testing. Our index shows good calibration, as seen by the similar mortality rates in the development and validation cohorts, and good discrimination, as seen by a *c* statistic of 0.84 in the development cohort and 0.82 in the validation cohort. As seen from our subgroup analysis, our index appears valid in the entire population of individuals older than 50 years, as well as among black individuals and those with less than a high school education. Our index is parsimonious without significant losses in discriminative ability: the *c* statistic using all 41 candidate variables is 0.847, whereas the *c* statistic of the 12 variable model is 0.842.

Our prognostic index demonstrates the importance of considering both the presence of disease and functional status in prognostic systems. We found that models considering both disease status and functional status perform better than models considering just 1 of these domains. Measures of function are particularly useful in prognostic systems because they reflect the severity and end consequence of disease.³¹ Thus, comorbid conditions and function are markers at different points along the trajectory of frailty. We believe that the inclusion of functional measures is the reason that our index discrimination

compares favorably with other widely used prognostic indexes, such as the Charlson-Deyo (0.60-0.78)^{32,33} and Framingham (0.77-0.74).³⁴

Our index has several potential uses in clinical, policy, and research settings. In clinical settings, our index may be useful in identifying both high- and low-risk patients so that specific interventions can be targeted to each group. Because the benefits of cancer screening are not realized for 5 years, recent cancer screening guidelines call for targeting screening to individuals with a life expectancy of more than 5 years.^{4-6,35} Our index may be useful in identifying older low-risk patients who may benefit from screening as well as identifying younger high-risk patients for whom the benefits of screening are outweighed by the harms. For example, a 75-year-old male smoker with heart failure, difficulty bathing, walking, and managing finances may not be an appropriate candidate for colorectal cancer screening because his probability of 4-year mortality is greater than 64% (16 points). On the other hand, an 85-year-old woman with no major comorbid conditions and excellent functional status has a high probability of surviving 4 years and would be a good candidate for screening despite her advanced age. In addition, our index may be useful in identifying patients with whom advance directives would be especially important to discuss.

Our index may also be useful when risk adjustment is needed to compare the patient outcomes among different health care organizations. Currently there is much interest at the federal,³⁶⁻³⁸ state,^{39,40} and private⁴¹ levels in evaluating care organizations through patient outcomes. As these efforts proceed, accurate methods of risk adjustment are needed to avoid penalizing clinicians and organizations that care for sicker patients.

Finally, our index may be useful in epidemiological studies that examine the impact of exposures and treatments on mortality. A primary concern in observational studies is baseline differences between groups, making any difference at the end of the study difficult to interpret. For example, the results of ran-

domized trials of postmenopausal estrogen therapy suggest that the initial observational studies were biased because the treated participants were healthier before treatment than the untreated participants.⁴² More thorough risk adjustment of these observational studies may have led researchers to the correct conclusion and prevented harmful estrogen exposure in many women.⁴³

Although risk adjustment is used extensively in research, prognostic indexes have made little headway into routine clinical practice.⁴⁴ The reasons for this are numerous but include competing demands on the physician's memory and difficulty in collecting all necessary information for an index. We attempted to maximize the usability of our index by addressing these issues. First, to minimize demands on physician memory, we made our index as simple as possible. Unlike previous indexes, we used a more stringent criterion than statistical significance to select our predictor variables.^{8,11,13,15} By using the BIC, we were able to remove variables that were statistically significant but contributed little to our predictive accuracy. Second, we limited our predictor variables to information that would be readily available through patient report. Although it is possible that clinical evaluation and observation may be a more objective measure of comorbidity and function, there is strong evidence that patient report of both comorbidity and functional status is reliable and has predictive validity.^{45,46} Given the high time costs for chart review and laboratory testing, our reliance on patient report will make our index easier to use across a variety of care settings while compromising little in terms of reliability and accuracy.

Our mortality index has several limitations. Since our index focused on community-dwelling older adults, it is likely that our index will not be applicable to nursing home and other institutional populations. Since the participants were predominantly white with a high school education, our model's ac-

Box. Four-Year Mortality Index for Older Adults

| | |
|---|---|
| 1. Age _____ | 60-64: 1 point 65-69: 2 points 70-74: 3 points 75-79: 4 points 80-84: 5 points ≥85: 7 points |
| 2. Sex (Male/Female) | Male: 2 points |
| 3. a. Weight: _____ b. Height: _____ $703 \times (\text{weight in pounds} / \text{height in inches}^2)$ BMI = _____ | BMI <25: 1 point |
| 4. Has a doctor ever told you that you have diabetes or high blood sugar? (Y/N) | Diabetes: 1 point |
| 5. Has a doctor told you that you have cancer or a malignant tumor, excluding minor skin cancers? (Y/N) | Cancer: 2 points |
| 6. Do you have a chronic lung disease that limits your usual activities or makes you need oxygen at home? (Y/N) | Lung Disease: 2 points |
| 7. Has a doctor told you that you have congestive heart failure? (Y/N) | Heart Failure: 2 points |
| 8. Have you smoked cigarettes in the past week? (Y/N) | Smoke: 2 points |
| 9. Because of a health or memory problem do you have any difficulty with bathing or showering? (Y/N) | Bathing: 2 points |
| 10. Because of a health or memory problem, do you have any difficulty with managing your money—such as paying your bills and keeping track of expenses? (Y/N) | Finances: 2 points |
| 11. Because of a health problem do you have any difficulty with walking several blocks? (Y/N) | Walking: 2 points |
| 12. Because of a health problem do you have any difficulty with pulling or pushing large objects like a living room chair? (Y/N) | Push or Pull: 1 point |

Total Points: _____

curacy and discrimination are slightly less for black individuals and those with less than high school education. Finally, because this was a community-dwelling sample, which included participants as young as 50 years, our cohort was generally healthy. In a sicker cohort, the ideal model would likely include other predictor variables that capture the severity of comorbidity and functional impairment.

In summary, our index provides a potentially useful tool to estimate the 4-year mortality risk in community-dwelling older adults in the United States. Our in-

dex relies on 12 variables, which are available by patient report in a variety of care and research settings. The index had good discrimination and calibration and was successfully validated in a geographically distinct cohort. These characteristics suggest that our index may be useful for clinical, policy, and epidemiological applications.

Author Contributions: Dr Lee had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Lee, Covinsky.
Analysis and interpretation of data: Lee, Lindquist, Segal, Covinsky.

Drafting of the manuscript: Lee, Covinsky.

Critical revision of the manuscript for important intellectual content: Lee, Lindquist, Segal, Covinsky.
Statistical analysis: Lee, Lindquist, Segal, Covinsky.
Obtained funding: Covinsky.

Administrative, technical, or material support: Covinsky.
Study supervision: Segal, Covinsky.

Financial Disclosures: None reported.

Funding/Support: The National Institute on Aging (NIA) supports the Health and Retirement study (U01AG09740). This project was sponsored by a grant R01AG023626 from the NIA. Dr Lee was supported by the Veterans Affairs National Quality Scholars Fellowship Program. Dr Covinsky was supported by grant K02-HS00006 from the Agency for Health Care Research and Quality.

Role of the Sponsor: The funding sources had no role in the design or conduct of the study, data management or analysis, or manuscript preparation or review.

REFERENCES

1. Reilly BM, Magnussen CR, Ross J, Ash J, Papa L, Wagner M. Can we talk? inpatient discussions about advance directives in a community hospital. Attending physicians' attitudes, their inpatients' wishes, and reported experience. *Arch Intern Med.* 1994;154:2299-2308.

2. Christakis NA, Iwashyna TJ. Attitude and self-reported practice regarding prognostication in a national sample of internists. *Arch Intern Med.* 1998;158:2389-2395.

3. Christakis NA. *Death Foretold: Prophecy and Prognosis in Medical Care.* Chicago, Ill: University of Chicago Press; 1999.

4. US Preventive Services Task Force. Screening for colorectal cancer: recommendation and rationale. *Ann Intern Med.* 2002;137:129-131.

5. American Geriatrics Society Clinical Practice Committee. Breast cancer screening in older women. *J Am Geriatr Soc.* 2000;48:842-844.

6. US Preventive Services Task Force. Screening for breast cancer: recommendations and rationale. *Ann Intern Med.* 2002;137:344-346.

7. Iezzoni L. *Risk Adjustment for Measuring Health Care Outcomes.* 3rd ed. Chicago, Ill: Health Administration Press; 2003.

8. Walter LC, Brand RJ, Counsell SR, et al. Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. *JAMA.* 2001;285:2987-2994.

9. Inouye SK, Peduzzi PN, Robison JT, Hughes JS, Horwitz RI, Concato J. Importance of functional measures in predicting mortality among older hospitalized patients. *JAMA.* 1998;279:1187-1193.

10. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40:373-383.

11. Fan VS, Au D, Heagerty P, Deyo RA, McDonell MB, Fihn SD. Validation of case-mix measures derived from self-reports of diagnoses and health. *J Clin Epidemiol.* 2002;55:371-380.

12. Fillenbaum GG, Pieper CF, Cohen HJ, Cornoni-Huntley JC, Guralnik JM. Comorbidity of five chronic health conditions in elderly community residents: determinants and impact on mortality. *J Gerontol A Biol Sci Med Sci.* 2000;55:M84-M89.

13. Carey EC, Walter LC, Lindquist K, Covinsky KE. Development and Validation of a Functional Morbidity Index to Predict Mortality in Community-dwelling Elders. *J Gen Intern Med.* 2004;19:1027-1033.

14. Saliba D, Elliott M, Rubenstein LZ, et al. The Vulnerable Elders Survey: a tool for identifying vulnerable older people in the community. *J Am Geriatr Soc.* 2001;49:1691-1699.

15. Fried LP, Elliott M, Rubenstein LZ, et al. Risk factors for 5-year mortality in older adults: the Cardiovascular Health Study. *JAMA.* 1998;279:585-592.

16. The Health and Retirement Survey. Ann Arbor: University of Michigan; 1999. Available at: http://hrsonline.isr.umich.edu/intro/sho_uinfo.php?hfyle=overview&xtyp=2. Accessed January 19, 2006.

17. The Health and Retirement Survey. Ann Arbor: University of Michigan; 1996. Available at: http://hrsonline.isr.umich.edu/docs/sample/sho_samp.php?hfyle=ref023b&xtyp=1. Accessed January 19, 2006.

18. The Health and Retirement Study. Ann Arbor: University of Michigan; 2002. Available at: http://hrsonline.isr.umich.edu/intro/sho_uinfo.php?hfyle=sample&xtyp=2. Accessed January 19, 2006.

19. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999;130:515-524.

20. HRS Restricted Data: What's Available. Ann Arbor: University of Michigan; 2005. Available at: <http://hrsonline.isr.umich.edu/meta/years/iy6.php?iyear=1027>. Accessed January 19, 2006.

21. Fonda SH, Herzog AR. Documentation of physical functional measured in the health and retirement study and the asset and health dynamics among the oldest old study. Ann Arbor: Survey Research Center, University of Michigan; 2004:14-16. Available at: <http://hrsonline.isr.umich.edu/docs/userg/dr-008.pdf>. Last updated December 21, 2004. Accessed January 19, 2006.

22. Nagi SZ. An epidemiology of disability among adults in the United States. *Milbank Mem Fund Q Health Soc.* 1976;54:439-467.

23. Rosov I, Breslau N. A Guttman health scale for the aged. *J Gerontol.* 1966;21:556-559.

24. Welch WP, Miller ME, Welch HG, Fisher ES, Wennberg JE. Geographic variation in expenditures for physicians' services in the United States. *N Engl J Med.* 1993;328:621-627.

25. Schwarz G. Estimating the Dimension of a Model. *Ann Statist.* 1978;6:461-464.

26. Hook EB, Regal RR. Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation. *Am J Epidemiol.* 1997;145:1138-1144.

27. Inouye SK, Charpentier PA. Precipitating factors for delirium in hospitalized elderly persons. Predictive model and interrelationship with baseline vulnerability. *JAMA.* 1996;275:852-857.

28. Inouye SK, Viscoli CM, Horwitz RI, Hurst LD, Tinetti ME. A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Ann Intern Med.* 1993;119:474-481.

29. Vittinghoff EG, Glidden DC, Shiboski SC, McCulloch CE. Regression methods in biostatistics. In: Gail M, Krickerberg K, Samet J, Tsatis A, Wong W, eds. *Statistics for Biology and Health.* New York, NY: Springer; 2005:147-149.

30. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med.* 1993;118:201-210.

31. Verbrugge LM, Jette AM. The disablement process. *Soc Sci Med.* 1994;38:1-14.

32. Cleves MA, Sanchez N, Draheim M. Evaluation of two competing methods for calculating Charlson's comorbidity index when analyzing short-term mortality using administrative data. *J Clin Epidemiol.* 1997;50:903-908.

33. Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. *Med Care.* 2004;42:355-360.

34. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation.* 1998;97:1837-1847.

35. Walter LC, Covinsky KE. Cancer screening in elderly patients: a framework for individualized decision making. *JAMA.* 2001;285:2750-2756.

36. *Guide to Inpatient Quality Indicators: Quality of Care in Hospitals—Volume, Mortality, and Utilization.* Rockville, Md: Agency for Healthcare Research and Quality; 2002. Available at: http://www.qualityindicators.ahrq.gov/downloads/iqi/iqi_guide_rev4.pdf. Accessed January 19, 2006.

37. *Guide to Prevention Quality Indicators: AHRQ Quality Indicators, Agency for Healthcare Research and Quality.* Rockville, Md: Agency for Healthcare Research and Quality; 2001. Available at: http://www.qualityindicators.ahrq.gov/downloads/pqi/pqi_guide_v30.pdf. Accessed January 19, 2006.

38. *AHRQ Quality Indicators—Guide to Patient Safety Indicators.* Rockville, Md: Agency for Healthcare Research and Quality; 2003. Available at: http://www.qualityindicators.ahrq.gov/downloads/psi/psi_guide_rev3.pdf. Accessed January 19, 2006.

39. The California Report on CABG Surgery. 2000-2002 Hospital Data. Sacramento, Calif: Office of Statewide Health Planning and Development; 2005.

40. *Adult Cardiac Surgery in New York State, 1998-2000.* Albany: New York State Dept of Health; February 2004. Available at: http://www.health.state.ny.us/nysdoh/heart/pdf/1998-2000_cabg.pdf. Accessed January 19, 2006.

41. Krumholz HM, Rathore SS, Chen J, Wang Y, Radford MJ. Evaluation of a consumer-oriented internet health care report card: the risk of quality ratings based on mortality data. *JAMA.* 2002;287:1277-1287.

42. Hulley S, Grady D, Bush T, et al; Heart and Estrogen/progestin Replacement Study (HERS) Research Group. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA.* 1998;280:605-613.

43. Petitti DB. Hormone replacement therapy and heart disease prevention: experimentation trumps observation *JAMA.* 1998;280:650-652.

44. Redelmeier DA, Lustig AJ. Prognostic indices in clinical practice. *JAMA.* 2001;285:3024-3025.

45. Covinsky KE, Palmer RM, Counsell SR, Pine ZM, Walter LC, Chren MM. Functional status before hospitalization in acutely ill older adults: validity and clinical importance of retrospective reports. *J Am Geriatr Soc.* 2000;48:164-169.

46. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care.* 1996;34:73-84.