

Predicting 12-Month Mortality for Persons With Dementia

Robert Newcomer,¹ Kenneth E. Covinsky,² Ted Clay,³ and Kristine Yaffe⁴

Departments of ¹Social & Behavioral Science, ²Medicine, and
⁴Psychiatry, University of California, San Francisco.
³Clay Software and Statistics, Ashland, Oregon.

Objectives. We develop and test a model of 12-month mortality among persons ($N = 3,858$) with organic dementia.

Methods. Data are from caregiver interviews and claims records for persons enrolled in the Medicare Alzheimer's Disease Demonstration Evaluation. Information covers the year prior to enrollment through 36 months. We used Proportional hazards models to predict time to death. We estimated two starting points: first, the date of randomization, and second, the date of maximum difficulty in performing two or more activities of daily living (ADLs).

Results. The final model includes age, gender, ADL difficulty, medical conditions, prior year hospitalizations, and whether a daughter was the primary caregiver. We combined hazard ratios to produce a cumulative mortality risk score. Model discrimination is reasonable for both models (c statistics of .72 and .69, respectively), and calibration tests were nonsignificant.

Discussion. The model's efficiency, as measured by the ratio of false positives (those predicted to die, but who lived) to true positives (those predicted to die and who did die) ranged from fewer than 1:1 to more than 4:1 as the model's sensitivity increased. This ratio was lower in the two or more ADL difficulty model. A validation test of the prediction model found comparable sensitivity and specificity (c statistic of .69) to the reference model.

ALTHOUGH patients with dementia have considerably higher mortality rates than age-matched controls, survival in patients with dementia is highly variable. A prognostic measure that could stratify patients with dementia into groups at differential risk for death would be useful to both clinical providers and policy makers. For example, such a measure would be helpful to providers who counsel patients and their families about advance planning and palliative care options such as hospice. In addition, a prognostic measure would be useful for risk adjustment, either to compare outcomes across different groups of providers or to more fairly compensate health systems that care for a greater proportion of higher risk patients.

Our model for the prediction of mortality in patients with dementia is guided by a conceptual framework, adapted from Iezzoni (1997). She views poor outcomes such as mortality as mediated by multiple domains of risk factors that include demographic, biomedical, and social factors. Risk factors signified by disease (or comorbidity), social factors, functional status, and health behaviors interact with each other and the effects of age and gender to increase the risk for mortality. The theoretical basis for Iezzoni's work is formed in part by a biopsychosocial model of health and illness. This perspective, as postulated by George Engel (1977), complements the prevailing biomedical formulations of illness severity (which focus primarily on disease diagnoses) by including the notion that the patient, the social context in which he lives, and the system devised by society to deal with illness should all be considered. When one views illness severity and prognosis through the eyes of the patient as in the biopsychosocial perspective, it becomes apparent that illness severity and

prognosis are to some extent socially determined and must include dimensions important to patients, such as functional status (Covinsky & Landefeld, 1996). It is also apparent that social and environmental issues interact with biology, health practices, and individual definitions of health.

The importance of considering the multiple domains postulated by the biopsychosocial model (including those used by Iezzoni) when defining health status and predicting health outcomes in older people is well established by empirical work. For example, across all age ranges, men have higher rates of mortality than women (Fried et al., 1998). The number and severity of comorbid conditions are associated with mortality. Particular comorbid conditions such as congestive heart failure are often important components of comorbidity indices (Charlson, Pompei, Ales, & MacKenzie, 1987). Physical function, such as the ability to perform activities of daily living (ADLS) without assistance, has also been shown to predict mortality independently, and it may be at least as important a predictor as diagnoses. One example of this work is that ADL function has been shown to strongly predict mortality in hospitalized elders (Covinsky, Justice, Rosenthal, Palmer, & Landefeld, 1997; Walter et al., 2001). Within the current sample of persons with dementia, when comorbidity and other risk factors were held constant, the rate of 12-month mortality was 10% among those with no ADL limitations and more than 20% among those having two or more limitations.

The quality of the social networks available to an elder, another element in Iezzoni's framework, may be as important a predictor of health outcomes as traditional biomedical predictors. For example, Berkman reported that among older people hospitalized with myocardial infarction, lack of

emotional support was associated with a threefold higher mortality over a 6-month period (Berkman, Leo-Summers, & Horwitz, 1992). The importance of social support is particularly important in patients with dementia, who are heavily reliant on the care of others. Socioeconomic status (SES), whether measured by education, income, occupation, or insurance status, has also been consistently associated with mortality and functional loss (Adler, Boyce, Chesney, Folkman, & Syme, 1993). Minority ethnicity has also been demonstrated to be a risk factor for higher mortality, although this is an effect only partially explained by SES (Guralnik, Land, Blazer, Fillenbaum, & Branch, 1993).

Although the discussion just given and most of our analysis focus on baseline risk factors, adverse outcomes in older people are probably the manifestation of both baseline risk factors and precipitating events. In older people, hospitalization is often used as a model of precipitating events (Gill, Williams, & Tinetti, 1999). For example, Morrison has demonstrated that in patients with end-stage dementia, mortality rates are very high following hospitalization for hip fracture or pneumonia (Morrison & Siu, 2000). Within the current sample, persons with dementia having a hospital stay for any cause had almost twice the mortality rate of those without a hospital stay. This rate increased fourfold to almost 50% among those discharged to nursing homes.

The development of any prognostic model is limited by the elements available in the data set. For example, we were not able to consider all of the risk domains (e.g., smoking) that have been formulated and tested by Iezzoni and others. However, the use of her framework informed our use of available elements and informs our thinking about how future studies may improve this model.

Mortality studies typically use an index event, such as a hospital stay or date of diagnosis, to provide the starting point or time zero for the time frame of interest. However, an index event such as one of these ignores information about the natural history and management of dementia that may have contributed to the hospital stay. An alternative event, the loss of independence in ADLs, has been suggested as a potentially more appropriate index event for persons with dementia (Gauthier, 1998). Our analysis examines whether this crucial outcome for dementia patients and their caregivers may be an effective start date in the prediction of mortality.

METHODS

Data came from caregiver interviews and claims records compiled for persons enrolled in the Medicare Alzheimer's Disease Demonstration Evaluation (MADDE). This was an eight-site program funded by the Centers for Medicare & Medicaid Services (CMS). It operated from December of 1989 through November of 1994. MADDE eligibility included a physician-certified diagnosis of dementia (e.g., Alzheimer's disease, vascular dementia, dementia caused by degenerative diseases, infections, and trauma), eligibility for Parts A and B of the Medicare program, residence in a program catchment area, and not living in a nursing home at time of enrollment. Dementias resulting from tumors, toxins, drugs, or nutritional or psychiatric disorders were excluded. The intervention and control groups have been combined here, as the case management and home care service intervention has been

shown to have *no* association with mortality risk or nursing home placement (Miller, Clay, Fox, & Newcomer, 1999), or with caregiver outcomes such as depression and burden that could be thought to influence the quality of informal care (Newcomer, Yordi, DuNah, Fox, & Wilkinson, 1999).

Excluded from the combined sample of 8,108 MADDE participants were (a) those who enrolled as managed care members ($n = 2,307$) at program enrollment or during the year prior to enrollment; (b) those with no match to the Medicare eligibility claims file ($n = 203$); (c) those with less than 6 months of Medicare eligibility in the prior year ($n = 63$); and (d) those whose primary caregiver was either a group home operator or a paid provider and for whom no family member was defined as the primary caregiver ($n = 118$). The exclusion of managed care, those without a match to claims records, or less than 6 months of Medicare eligibility was necessary for the present analysis because we needed Medicare claims data to track hospital stays and diagnoses. The paid providers (which includes group home operators as well as live-in aides) were omitted because this group was too small for separate analysis.

We have divided the individuals included in the analysis into two cohorts: those who enrolled into MADDE between December of 1989 and March of 1991 ($n = 3,858$) and those who enrolled between April and November of 1991 ($n = 1,559$). We interviewed the initial cohort at baseline (in their homes) and semiannually (by telephone) for 24 months (if in the community), and then at 36 months. We obtained assessment data for the second cohort only at baseline, and at 24 and 36 months. We used the initial sample to develop the mortality risk prediction model. We used the second cohort here in validation tests of the prediction model.

We obtained information on subject and primary caregiver attributes from assessment interviews conducted with the primary caregivers, and a cognitive screening examination conducted with the program applicants. Linked to these attributes were the subjects' Medicare claims, and, if applicable, date of permanent nursing entry and date of death. A primary caregiver was the one person who provided the "most assistance" to the person with dementia. No cases in the analysis were lost from claims or the observation of death, as we compiled Medicare eligibility and claims files for at least 36 months after study enrollment for all participants.

Measures and Prediction Model Specification

The dimensions of risk are shown in Tables 1–3. We have used two alternative measures to define the time zero or starting period for the prospective analysis of mortality risk: first, the date of randomization into the demonstration, and second, the date when the caregiver reported that the subject had maximum difficulty performing at least two of the following ADLs: bathing, dressing, eating, grooming, toileting, transferring, or walking (Katz, Ford, Moskowitz, Jackson, & Jaffee, 1963). We do not expressly consider date of institutionalization (defined by a permanent nursing home entry date), but we do retain cases in which subjects entered nursing homes after time zero in the analysis. This represents approximately 25% of the study sample, and just over 30% of those who died in the observation year.

Subject Characteristics

Assessments include demographic information as well as clinically relevant measures of physical health, function, and cognitive status. We discuss chronic health conditions in a separate section. The number and percent of cases with each attribute, and the proportion with the attribute who died during the ensuing 12 months, are shown in Table 1. Subject limitations in instrumental ADLs (IADLs; i.e., shopping, doing housework, doing laundry, managing medications, managing money, using the telephone, and getting to places outside the house) are not included, as a preliminary analysis found little variation among subjects.

We divided counts of subject behavior problems (Zarit, Reeves, & Bach-Peterson, 1980), such as asking repetitive questions, being suspicious or accusative, having trouble recognizing familiar people, and engaging in behavior potentially dangerous to self or others, into groups. This was to accommodate any curvilinear association between behavior problems and the risk of death. Another measure of cognitive status was the Mini-Mental State Examination (MMSE). The MMSE assesses orientation, recall, and ability to name objects. It was scored on a 30-point scale (Folstein, Folstein, & McHugh, 1975) and then grouped to address curvilinear relationships.

All measures for the randomization group were obtained at baseline. Among those having two or more maximum difficulty ADL limitations, 67% had this status at baseline.

Two other sets of measures are included in Table 1. These are the communities from which the sample was compiled and whether the case was in the MADDE treatment or control group. These measures were included in the initial Cox proportional hazards models as contextual controls, but they failed to retain statistical significance in the final model.

Caregiver Attributes

Caregiver measures are shown in Table 2. These include relationship to the subject, the caregiver's individual attributes, and items reflecting possible consequences of caregiving. These items were obtained at baseline for the randomization date group, and they were updated to reflect the information current at the time of the two maximum difficulty ADL classification. Caregivers' ADL limitations and IADLs (Lawton & Brody, 1969) used a slightly different set of activity items from the subject ADL and IADL scores, and asked whether "some" or "no" (but not "maximum") difficulty was experienced. The presence of some difficulty was used to count the number of limitations. Caregiver burden was measured by the 7-item Zarit Burden Scale (Zarit et al., 1980). Responses range from 0 = "never" to 4 = "almost always." The scale score was grouped into approximately equal aggregations as a way to adjust for any curvilinear relationships with mortality or nursing home placement (McCarty et al., 2000). Caregiver depression, measured by the Geriatric Depression scale, includes 15 items (Yesavage, Rink, Rose, & Aday, 1983).

Chronic Health Conditions

Comorbidities associated with mortality risk among adults include myocardial infarction, coronary artery disease, congestive heart disease, peripheral vascular disease, cerebrovascular

disease, dementia, chronic pulmonary disease, connective tissue disease, ulcer disease, mild to severe liver disease, diabetes, hemiplegia, moderate or severe renal disease, and any tumor or cancer other than mild skin cancer (e.g., Charlson et al., 1987; Fillenbaum, Pieper, Cohen, Comoni-Huntley, & Guralnik, 2000). We incorporate these conditions into the analysis.

The MADDE data have two sources of information on chronic health conditions. One of these is a listing of conditions included on the physician referral form required as part of the application for MADDE participation. This was completed by each subject's physician, but its emphasis was on delineating the nature of the dementia rather than other conditions. A more inclusive source of diagnoses is Medicare claims. Preliminary analyses found that claims-identified condition prevalence, except for diabetes, was more highly associated with mortality. The prevalence and mortality rates associated with a selected list of conditions are shown in Table 3. These conditions reflect the array of chronic conditions included in the Charlson Index (Charlson et al., 1987) and other recent analyses of mortality risk in the aged (e.g., Fillenbaum et al., 2000). A number of studies have tested the predictive validity of the Charlson Index and similar weighting processes and extended the application from medical records to claims records (e.g., Cleves, Sanchez, & Draheim, 1997; Deyo, Cherkin, & Ciol, 1992; D'Hoore, Bouckaert, & Tilquin, 1996; Kieszak, Flanders, Kosinski, Shipp, & Karp, 1999).

A total of 13,561 Medicare Part A claims were used to compile the diagnoses in the 12 months prior to the time-zero randomization date period. Any condition mentioned defined the condition as present. Outpatient departments (59.6%), home health (24.5%), and inpatient services (14.0%) accounted for the majority of claims. Skilled nursing (1.8%) and hospice care (0.1%) accounted for the balance. There were 9,542 claims available for the any two ADL limitation group: outpatient departments (49.8%), home health (33.3%), inpatient services (14.2%), skilled nursing (2.5%), and hospice (0.2%). The use of a single year rather than multiple years and only Part A claims undercounts the actual number of comorbidities (Newcomer, Clay, Luxenberg, & Miller, 1999), but this approach enabled us to identify unstable or severe conditions.

Hospital Use

An additional adjustment for condition classification was a measure counting the number of inpatient stays in the 12 months immediately preceding time zero. The hospital use distribution is shown in Table 3. The enrollment process into the MADDE program created a systematic truncation of hospital effects on mortality risk for those cases measured prospectively from date of randomization. Individuals could not enroll into MADDE if they were in a hospital. They had to be living in the community. Thus mortality occurring during or immediately following an inpatient stay is underrepresented in the MADDE enrollment. Hospital stays after date of enrollment were not constrained in this manner.

Statistical Methods

We modeled survival time in months by Cox proportional hazard regression, using PROC PHREG in SAS Version 8.2. We censored time at 12 months following the starting date,

Table 1. Subject Characteristics

Characteristic	Randomization Date		MD Any 2 ADLs	
	No. of Cases	% Died	No. of Cases	% Died
Gender				
Female	2336	11.6	1417	16.7
Male	1522	18.5	838	27.9
Race				
Black	375	13.9	262	19.5
Hispanic-White	128	14.8	95	18.9
White or other	3355	14.3	1898	21.2
Age (years)				
20-64	52	15.4	36	19.4
65-69	383	9.9	213	16.4
70-79	1639	11.3	943	18.9
80-84	977	15.4	569	20.4
85-89	580	19.5	346	25.1
90+	227	25.6	148	32.4
Income (\$)				
<5K	317	15.5	209	16.7
5-10K	1054	14.8	632	21.4
10-20K	1311	14.5	754	21.9
20-30K	637	13.3	359	20.3
>30K	505	13.1	288	20.1
Income missing	34	17.6	13	38.5
Education				
<High school	1065	16.2	691	23.3
Some high school	682	16.4	378	23.0
High school graduate	1058	13.3	616	18.8
Some college	522	11.7	289	16.6
College graduate or more	518	12.7	274	21.5
Education missing	13	0.0	7	0.0
ADLs needing maxi. help				
0	1886	8.4	NA	NA
1	440	14.5	NA	NA
2	356	17.4	641	15.3
3	325	18.2	569	18.5
4	212	17.0	311	21.2
5	161	23.6	200	25.5
6	170	24.7	201	26.4
7	308	29.9	333	29.4
Behavior problems				
0-4 (the best)	540	16.5	378	22.5
5-7	834	13.3	468	19.9
8-10	967	13.1	541	23.5
11-13	878	15.6	470	18.9
14-19	632	13.3	397	19.1
Problems missing	7	42.9	1	100.0
MMSE score				
0 severe (the worst)	495	24.8	463	26.1
1-5 severe	357	16.0	283	19.8
6-10 moderately severe	387	16.5	284	21.1
11-15 moderate	633	15.5	380	22.1
16-20 mild-moderate	792	11.4	372	16.7
21-25 mild-moderate	671	9.2	241	20.3
26-30 early-mild	365	7.9	125	16.8
Score missing	158	18.4	107	16.8
Location				
Florida	548	18.4	368	26.6
Illinois	624	13.1	311	20.3
Minnesota	380	10.5	173	16.8
New York	454	14.3	235	19.6

(Table 1 continues)

Table 1. Subject Characteristics (Continued)

Characteristic	Randomization Date		MD Any 2 ADLs	
	No. of Cases	% Died	No. of Cases	% Died
Ohio	632	14.9	394	20.6
Oregon	304	14.5	182	22.5
Tennessee	574	14.3	382	19.4
West Virginia	342	12.9	210	18.6
Assign. in MADDE demonstration				
Treatment group	1947	14.4	1157	20.3
Control group	1911	14.2	1098	21.5
Total	3858	14.3	2255	20.9

Notes: MD = maximum difficulty; ADL = activity of daily living; MMSE = Mini-Mental State Examination; MADDE = Medicare Alzheimer’s Disease Demonstration Evaluation; NA = not applicable.

which was defined in two ways, as already discussed. We performed the initial model-reduction process by using backward stepwise procedures to address any collinearity among measures. We performed the model validation using the April–November MADDE enrollment cohort by applying the parameter estimates fitted on the main cohort. We assessed model calibration by comparing predicted deaths versus actual deaths within deciles of probability of death. This was tested with the Hosmer–Lemeshow chi-square test.

RESULTS

The prevailing unadjusted likelihood of death among the mortality prediction sample was 14.3% when date of randomization was used as the starting point, and 20.9% when time zero shifted to the presence of at least two ADLs where maximum assistance was needed. Three sets of findings are presented. These attempt to broaden and refine these unadjusted estimates. The initial analyses used Cox proportional hazards methods to derive multivariate models of mortality. The linear parameter estimates obtained in these models were then combined into a cumulative risk score for each participant. The sensitivity and specificity of the derived risk scores and their relationship with mortality is tested in a second analysis. A third analysis assesses the predictive value of the risk scores by applying them to a second cohort of MADDE enrollees (i.e., those enrolling between April and December of 1991).

Proportional Hazard Models of Mortality

Table 4 shows the results for two models representing alternative approaches to measuring the effect of functional disability on the likelihood of death within 12 months. One uses the randomization date; the other uses the presence of at least two ADLs with maximum performance difficulty. We adjusted both models for the simultaneous risk associated with the subject, caregiver, and other attributes shown in Tables 1–3. Only those sets of measures having a significant association with the mortality are shown.

These models produce relatively similar results; namely, the final set of measures is the same and these attributes have similar hazard ratios. This similarity may be due in part to the high proportion (67%) of those with maximum difficulty in at least two ADLs reaching this status in their baseline assessments. The final model risk factors encompass four broad categories. One of these was subject demographics. Male

subjects were associated with a higher risk of death than female subjects, as were those beyond the age of 80. Clinical conditions of heart, pulmonary, cancer, and diabetes were also found to have a substantial association with the risk of death.

One caregiver attribute emerged as a risk factor. Daughters as caregivers were associated with a lower subject mortality risk, independent of subject age and gender, than were spouses. The final factor in the model is the history of inpatient utilization. This measure serves as a proxy measure for severity of illness. Three or more prior year inpatient stays were associated with an increased hazard of death. The absence of similar effects with recent hospital stays may be an artifact of the sample exclusion criteria noted earlier.

Predicting Risk of Mortality

The risk factors in the final models were combined to produce a cumulative risk score for each individual in the sample. We did this by summing the linear predictor (i.e., the logarithmic form of the hazard ratios) for each of the measures applicable to the case. We then tested these summed risk scores (here converted back to hazard ratios) relative to the likelihood of death. We used a series of cut points in the risk score to divide the population into high-risk cases predicted to die or low-risk cases predicted to live. This prediction has the following components:

- true positive cases (those predicted to die who do die) = *A*,
- false positive cases (those predicted to die who live) = *B*,
- false negative cases (those predicted to live who die) = *C*, and
- true negative cases (those predicted to live who live) = *D*.

Sensitivity and specificity are commonly used measures of how well a model discriminates or correctly distinguishes persons who live from those who die. In this case, sensitivity [$A/(A + C)$] is the percent of deaths correctly classified by our prediction rule (true +); specificity [$D/(B + D)$] is a measure of the percent of cases correctly classified as living (true –). A particular advantage of sensitivity and specificity measures is that their values are unaffected by the prevalence of death in the population (Ash & Shwartz, 1997).

The columns in Table 5 show the distribution of the sample cases among those predicted to die and their outcomes relative to death or remaining alive after 12 months. The rows show the distribution of risk scores divided into cut points based on approximately 10% increments in sensitivity. Evident from the

Table 2. Caregiver Characteristics

Characteristic	Randomization Date		MD Any 2 ADLs	
	No. of Cases	% Died	No. of Cases	% Died
Relationship to client				
Spouse	1846	16.1	1119	23.6
Daughter	1134	11.8	673	16.3
Other relative	878	13.7	463	21.0
Caregiver income (\$)				
<10K	503	17.1	347	21.3
10–40K	2579	14.8	1521	21.6
40K+	776	11.0	387	17.8
Caregiver education				
<High school	837	18.8	539	27.3
High school graduate	1154	13.0	678	17.6
Some college	916	14.6	523	20.5
College graduate	951	11.7	515	19.0
Caregiver age (years)				
<70	2407	12.7	1402	18.5
70–74	483	14.1	299	21.4
75–79	490	18.6	286	27.6
80–84	340	18.5	183	26.8
85+	138	18.1	85	22.4
Caregiver health status				
Excellent	1131	12.1	586	18.1
Good	1815	14.4	1062	21.7
Fair	765	16.6	513	22.6
Poor	147	17.7	94	20.2
Caregiver ADL limitations				
0	3260	13.9	1905	20.1
1	382	16.5	215	25.6
2+	598	16.7	350	25.4
Caregiver IADL limitations				
0	2678	13.2	1527	19.3
1	374	17.4	237	26.6
2	247	13.0	150	20.0
3+	559	18.1	341	24.3
Caregiver stress–burden score				
0–6 problems (best)	779	12.6	393	22.6
7–10 problems	731	12.0	393	17.6
11–14 problems	833	14.5	464	20.7
15–18 problems	756	15.1	428	18.7
19–28 problems (worst)	742	17.5	565	23.7
Problem score missing	17	5.9	12	25.0
Caregiver depression score				
0 (Best)	325	10.5	143	17.5
1–2	1053	12.9	528	19.3
3–4	898	13.7	525	18.7
5–6	620	14.7	396	22.2
7–8	418	14.1	277	22.0
9–15 (Worst)	523	20.5	371	25.3
Score missing	21	4.8	15	20.0
Total	3858	14.3	2255	20.9

Notes: Various other caregiver variables were used: in preliminary data runs, for example, total caregiving hours, insurance coverage, marital status, and working status. These were either highly intercorrelated or had no relationship to mortality risk. The burden score excludes family and work related burden items. These were tested separately. MD = maximum difficulty; ADL = activities of daily living; IADL = instrumental ADL.

table is a relatively constant trade-off between sensitivity and specificity, with the gain in sensitivity being achieved with a growth in the number of false positive cases (i.e., more persons are predicted to die than do die). In general, the randomization time-zero model performs marginally better than

the any two maximum difficulty ADL model, with higher specificity and somewhat fewer false positive cases as sensitivity increases. The performance of these models is reflected in a *c* statistic, which is a comparison of the proportion of pairs in which the predicted probability of death is higher for the

Table 3. Selected Chronic Conditions Among Clients

Chronic Condition	Randomization Date		MD Any 2 ADLs	
	No. of Cases	% Died	No. of Cases	% Died
Selected from physician referral listing				
Dementia (all causes)				
Alzheimer's disease	2718	13.3	1551	19.8
Degenerative disease, CNS infection, trauma or other	700	15.9	449	22.3
Vascular dementia	906	18.3	556	25.4
Other conditions				
Angina or other cardiac conditions	1010	20.6	557	28.2
Arthritis	1197	16.2	737	21.4
Cancer	205	22.0	112	30.4
Diabetes	436	20.6	271	25.1
Paralysis	138	21.0	118	23.7
Renal Insuff.	219	25.6	160	27.5
Respiratory	384	20.8	222	28.4
Selected from claims in prior 12 months				
Cerebrovascular				
All cerebrovas. disease events	334	21.6	250	26.0
Cerebrovascular aneurysm	152	21.7	136	22.8
Cancer (except minor skin cancer)	117	30.8	83	38.6
Coronary Conditions				
CAD	322	20.8	210	29.0
CHF	214	29.4	162	34.6
MI	49	28.6	40	27.5
Severe heart condition (not CHF, MI, or CAD)	28	35.7	17	64.7
Circulatory & renal conditions				
Chronic renal problem	51	15.7	27	33.3
Gangrene	7	42.9	10	40.0
Kidney infection	369	24.1	311	28.3
Renal failure	41	39.0	35	48.6
Peripheral vascular disease	83	19.3	66	24.2
COPD-asthma-emphysema	199	31.2	140	42.1
Diabetes (w/ and w/o complications)	250	18.8	184	25.0
Infectious arthritis-osteomyelitis	14	21.4	9	33.3
Liver-gall bladder-pancreas disease	60	15.0	40	20.0
Paralysis (excludes coma or injuries)	104	17.3	94	17.0
Inpatient stays				
None in the 12 prior months	2647	12.8	1419	19.2
Any in ≤ 90 days	567	18.9	391	27.1
1 in 91+ prior days	464	12.9	311	16.1
2 in 91+ prior days	131	22.1	96	26.0
3+ in 91+ prior days	49	36.7	38	47.4
Total	3858	14.3	2255	20.9

Notes: Medicare claims records for the 12 months prior to the starting period were the basis of claims condition classifications and inpatient counts. Physician referral listings were provided at the time of application to the Medicare Alzheimer's Disease Demonstration. Dementia cases exceed the total, as some cases had more than one cause of dementia identified by their physician. MD = maximum difficulty; ADL = activities of daily living; CNS = central nervous system; CAD = coronary artery disease; CHF = congestive heart failure; MI = myocardial infarction; COPD = chronic obstructive pulmonary disease.

patient who died than for the patient who lived. Each tied pair is counted as one half. A *c* statistic also equals the area under a receiver operating characteristic curve. Both time zero models perform relatively similarly, although the randomization date model has a slightly higher *c* statistic (0.72 vs. 0.69). These values are consistent with *c* statistics reported from a number of risk-adjusted mortality models (e.g., Ash & Shwartz, 1997), and they reflect a 2 standard deviation discrimination improvement by both models over chance, which would be reflected by a *c* statistic of 0.5. There is approximately a 0.10 standard deviation difference between the two time-zero models.

Validating the Predictive Model

A second test of the predicted mortality risk classification model is reflected in a comparison of the model's calibration. Calibration in this context is the comparison of predicted versus actual deaths across the range of risk scores. These results are shown in Table 6. This table again compares randomization date with the any two maximum difficulty ADLs models, and it adds mortality risk information from a replication or validation sample. This sample consists of persons who enrolled in the MADDE program between April and December of 1991. The validation sample has attributes that are similar, but not

Table 4. Likelihood of Mortality in 2 Months

Measures	Randomization Date				≥ 2 ADL Max. Help			
	No. of Cases	Haz. Ratio	LCI	UCI	No. of Cases	Haz. Ratio	LCI	UCI
Client gender								
Female	2336				1417			
Male	1522	1.64	1.34	2.00	838	1.62	1.31	2.00
Client age (years)								
65–69	383				213			
20–64	52	1.30	0.60	2.79	36	1.00	0.44	2.25
70–79	1639	1.11	0.78	1.58	943	1.08	0.75	1.56
80–84	977	1.67	1.17	2.40	569	1.36	0.93	2.00
85–89	580	2.33	1.60	3.39	346	1.90	1.27	2.85
90+	227	2.90	1.90	4.43	148	2.42	1.54	3.81
ADLs needing max. help								
0	1886				NA			
1	440	1.70	1.27	2.28	NA			
2	356	2.06	1.53	2.77	641			
3	325	2.33	1.72	3.15	569	1.27	0.96	1.68
4	212	1.91	1.32	2.75	311	1.41	1.03	1.93
5	161	2.91	2.04	4.17	200	1.78	1.26	2.52
6	170	2.64	1.87	3.74	201	1.56	1.11	2.19
7	308	3.87	2.98	5.04	333	2.09	1.57	2.78
Caregiver relationship								
Spouse	1846				1119			
Daughter	1134	0.70	0.56	0.89	673	0.68	0.53	0.89
Other	878	0.94	0.74	1.20	463	0.98	0.75	1.28
Inpatient stays								
None in prior 12 months	2647				1419			
Most recent 1–90 days prior	567	1.01	0.79	1.28	391	1.06	0.82	1.35
1 before 91+ days	464	0.67	0.50	0.90	311	0.62	0.46	0.85
2 before 91+ days	131	1.05	0.70	1.56	96	0.90	0.59	1.37
3 before 91+ days	49	1.66	1.01	2.72	38	1.98	1.21	3.26
Chronic conditions								
CHF or severe cardiac condition	232	1.84	1.40	2.43	174	1.61	1.21	2.15
COPD–asthma–emphysema	199	2.04	1.53	2.71	140	2.03	1.50	2.73
Cancer (no minor skin cancer)	117	1.92	1.34	2.74	83	1.70	1.16	2.48
Diabetes	436	1.51	1.20	1.90	271	1.22	0.94	1.58
Total	3858				2255			

Notes: ADL = activities of daily living; LCI and UCI = lower and upper confidence interval, respectively; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; NA = not applicable.

identical, to those in the randomization prediction model cohort. For example, the unadjusted mortality rate is 15.7 percent (vs. 14.3%) in the validation cohort. Consistent with this, the validation sample generally has higher proportions of cases with each of the following risk factors: being female (59.5% vs. 60.5%), being 85 years old or older (24.4% vs. 20.9%), having no inpatient stays in the prior year (65.3% vs. 68.6%), and having one or more of the target chronic conditions (18.5% vs. 17.4%). The validation sample has proportionately fewer cases with only two of the risk dimensions, that is, having five or more ADLs (14.4% vs. 16.6%) and having daughter caregivers (29.4% vs. 31.8%).

The rows in Table 6 reflect a distribution of the sample into deciles, ordered by the risk of mortality going from lowest to highest rank. The validation model used the weights from the randomization date development sample, and it applied them to the subjects in the validation sample. The number of cases in each row may not equal exactly 10% of the cases as a result of ties in risk scores. A Hosmer–Lemeshow chi-square test

compares the predicted versus actual mortality for each decile, and for the model overall; *c* statistics of each model's discrimination are also shown. The results for both the randomization and the any two ADL maximum difficulty models are based on risk scores derived from prediction models discussed previously. These models, in spite of minor differences in their *c* statistics, show similar performances in predicting the total number of deaths, and the number of deaths at each decile in the risk score rankings. Both models also have nonsignificant Hosmer–Lemeshow chi-square results, indicating that they are well calibrated.

The validation sample results (the third panel in Table 6) show somewhat lower discrimination (*c* statistic = .69) and calibration than those of the prediction model. Calibration is particularly problematic at deciles 4 and 6. The net effect is that over most of the risk deciles, the validation sample underestimates actual deaths by approximately 6%. These differences suggest some overfitting of the models and problems with transportability. However, the reduction in discrimination and

Table 5. Sensitivity and Specificity of the Prediction Models

Linear Pred.	Haz. Ratio	A	B	C	D	Sensitivity [A/(A+C)]	Specificity [D/(B+D)]	1-Specificity (% false +)
A: All Cases From Date of Randomization into the MADDE Program								
3.82	45.48	1	0	551	3336	0.2	100.0	0.0
2.56	12.97	56	45	496	3261	10.1	98.6	1.4
2.14	8.50	111	137	441	3169	20.1	95.6	4.1
1.90	6.66	166	258	386	3048	30.1	92.3	7.8
1.69	5.40	221	427	331	2879	40.0	87.1	12.9
1.52	4.55	276	585	276	2721	50.0	81.7	17.7
1.24	3.47	332	969	220	2337	60.1	70.7	29.3
1.02	2.79	388	1293	164	2013	70.3	60.9	39.1
0.85	2.33	442	1622	110	1684	80.1	50.9	49.1
0.52	1.67	498	2278	54	1028	90.2	31.1	68.9
-0.34	0.71	552	3253	0	53	100.0	1.6	98.4
B: Those Having Maximum Difficulty ≥ 2 ADLs								
3.22	25.09	1	0	470	1784	0.0	100.0	0.0
1.87	6.50	48	35	423	1749	10.2	98.0	2.0
1.53	4.61	98	97	373	1687	20.8	94.6	5.4
1.30	3.65	142	207	329	1577	30.1	88.4	11.6
1.13	3.08	196	311	275	1473	41.6	82.6	17.4
1.00	2.72	236	421	235	1363	50.1	76.4	23.6
0.82	2.27	286	591	185	1193	60.7	66.9	33.1
0.70	2.02	330	751	141	1033	70.1	57.9	42.1
0.50	1.64	377	1051	94	733	80.0	41.1	58.9
0.28	1.32	424	1325	47	459	90.0	25.7	74.3
-0.85	0.43	471	1784	0	0	100.0	0.0	100.0

Notes: MADDE = Medicare Alzheimer’s Disease Demonstration Evaluation; A and B = number of high risk cases that died and lived, respectively; C and D = number of low risk cases that died and lived, respectively. Sensitivity is the percent of deaths correctly classified by our prediction rule (true +); specificity is the percent of cases correctly classified as living (true -). The c statistics for panels A and B are .725 and .686, respectively. ADL = activities of daily living.

the calibration problems are consistent with what is usually seen with predictive models. One possible explanation for the differences in model calibration examined was the geographic distribution of the sample cases, with location perhaps serving as a proxy for practice patterns and service supply. Expressed as a percentage of total subjects, the proportion of cases varied by less than 2% between the two samples among four of the sites, and by up to 6% among three sites.

DISCUSSION AND CONCLUSIONS

This article has identified factors associated with mortality over 12 months. The analysis used 3 years of cohort data from a sample of persons known to have some form of organic dementia. Mortality over 12 months was 14.3% from the date of enrollment into the MADDE demonstration, and 20.9% among those having maximum difficulty performing two or more ADLs. The individual-level attributes of age, gender, the need for maximum assistance in ADLs, and selected medical conditions (i.e., congestive heart failure, chronic obstructive pulmonary disease, cancer, and diabetes) were each associated with an increase of 50% or more in the likelihood of mortality (i.e., a hazard ratio greater than 1.5). The number of prior year hospitalizations (an indicator of illness severity) was also associated with an increased likelihood of death. A daughter as the primary caregiver, rather than the spouse, was associated with reduced mortality.

These measures were combined to produce a cumulative risk score. This score was used to predict 12-month mortality. Its effectiveness, as measured by the ratio of false positives (those predicted to die, but who lived) to true positives (those

predicted to die and who did die) ranged from fewer than 1:1 to more than 4:1 as the model’s sensitivity increased. The ratio of false positives to true positives was generally lower when the prospective starting point for predicting mortality was defined as having maximum difficulty performing two or more ADLs, rather than when the start time of enrollment into the study was used, adjusting for the number of ADL limitations.

The validation test of the model was rigorous because it tested the model in subjects with some difference in their geographic distribution, and who enrolled at a different point of time than the subjects in the development sample. As a result, the validation tests not only the degree of model overfitting, but the degree of geographic and temporal transportability of the model (Justice, Covinsky, & Berlin, 1999). In spite of this rigorous validation test, there was only a moderate drop in the discrimination of the model and a moderate fall in the calibration of the model. The validation application underpredicted actual mortality by an average of 6%, or 15 cases out of 245 deaths. This suggests that our model may be useful in predicting mortality among patients with dementia in other geographic settings. Application among group home residents, those with paid care providers, and those after nursing home entry has not been tested.

These findings are encouraging. Moreover, the predictive efficiency of the models can likely be increased in future applications where limitations of the MADDE data set can be rectified. For example, approximately two thirds of the demonstration enrollees had at least two ADL limitations at the time of enrollment. It is possible that establishing a more precise measurement of the onset of maximum ADL dependency would improve predictive reliability. However,

Table 6. Testing the Calibration of the Prediction and Validation Models

Decile Pred. Mortality	Sample Size in Decile	Pred. % Death	Actual % Died	Pred. Deaths	Actual Deaths	Contrib. to H-L χ^2 Stat.
A: Randomization Date						
1	385	3.9	4.4	15.2	17	0.22
2	397	5.4	3.9	21.0	15	1.82
3	388	7.3	7.5	28.5	29	0.01
4	385	8.2	5.7	31.6	22	3.18
5	376	9.9	11.2	37.3	42	0.65
6	394	12.1	14.7	47.7	58	2.51
7	380	14.6	14.2	55.6	54	0.54
8	392	17.9	16.3	70.4	64	0.70
9	382	23.6	24.3	90.2	93	0.11
10	389	39.9	40.6	155.3	158	0.08
B: ≥ 2 ADLs With Maximum Difficulty						
1	225	8.2	8.0	18.6	18	0.02
2	228	10.8	10.1	24.7	23	0.13
3	223	12.9	13.0	28.8	29	0.00
4	234	15.0	14.1	35.0	33	0.13
5	224	16.9	14.3	37.8	32	1.09
6	232	19.4	20.3	45.1	47	0.10
7	214	21.8	22.9	46.7	49	0.14
8	224	25.7	27.7	57.6	62	0.46
9	224	31.2	30.8	70.0	69	0.02
10	227	47.6	48.0	108.2	109	0.01
C: Validation Sample Using Date of Randomization						
1	156	3.9	3.2	6.1	5	0.19
2	155	5.3	5.8	8.3	9	0.07
3	156	7.2	9.0	11.3	14	0.69
4	156	8.3	14.7	12.9	23	8.57
5	156	10.1	10.9	15.8	17	0.10
6	155	12.3	16.8	19.0	26	2.94
7	161	14.9	16.1	24.0	26	0.20
8	152	18.5	19.1	28.1	29	0.04
9	156	24.9	26.9	38.9	42	0.33
10	156	42.5	34.6	66.3	54	3.97

Notes: H-L = Hosmer-Lemeshow; ADL = activities of daily living. The total predicted deaths for panels A, B, and C are 552.8, 472.4, and 230.6, respectively; total actual deaths are 552, 471, and 245; *c* statistics are .725, .686, and .689; H-L statistic χ^2 tests are 9.34, $p < .391$, 8 *df*; 2.11, $p < .98$, 8 *df*; and 17.10, $p < .03$, 8 *df*.

whether simply having more accurate information on the movement from one to two or two or higher limitations will be informative versus knowing the onset of particular limitations such as being unable to eat or transfer without maximum assistance has not been tested. What was observed is that mortality rates are higher among those with more ADL limitations.

A second area of limitation is the measurement of chronic conditions. A diagnosis establishes the types of care and resources required to treat the illness, and for this reason diagnoses have been a focal point for the development of risk-adjustment methods. The current analysis attempted to enumerate all known conditions, but claims sources likely underenumerated the conditions. Further, among the identified conditions, we were not able to distinguish primary disease process (other than dementia) from coexisting conditions; we also could not adjust for "severity." Studies having access to medical records would be able to more reliably determine the number of conditions, and possibly delineate the level of severity or stage, or the conditions directly associated with death. Such information could change the mortality risks for selected conditions over those reported here.

A third issue is more subtle, but it is related to the timing associated with diagnosis or functional limitation; it involves the selection of cases or patients. The MADDE sample had an inherent selection bias among program enrollees. All had to be living in the community (and presumably medically stable enough for the physician and families) to apply for program admission. This selection rule diminished the salience of the hospital stays in the weeks prior to admission, and it likely undercounted the proportion of the population with behavior and other challenges for which a nursing home placement was eminent. Although this is a possible limitation to the study findings, it is also important to consider that the selection of cases into any program or benefit will have to confront similar enrollment decision rules. Cases with eminent nursing home placements or with very high mortality risk (such as those who die during their hospital stay or shortly afterward) would likely be excluded from a palliative care or end-of-life risk-adjusted payment oriented to people in the community. Similarly, those without family caregivers might also be excluded, as they were in these analyses. In other words, replications and extensions of the present analyses have to recognize that the decision-making process permitting application or admission to a program could

in various ways affect the salience of time-varying risk factors. In the current analyses, we likely underrepresented the prevalence of disease or condition severity in the dementia population (by the exclusion criteria). If this is true, then mortality risk may be higher in the presence of the identified risk factors than was represented by these analyses.

A fourth issue is that of caregiver attributes and their collinearity with each other. The models shown used caregiver relationship and found that the presence of a daughter as the primary caregiver lowered the risk of mortality compared with the presence of spousal caregivers. The underlying cause of this effect is not known. It may reflect a better ability to provide both instrumental assistance and disease management, or the attainment of better management through earlier nursing home placement than when there is a spousal caregiver. Among the expectations of a nursing home placement are close monitoring of medical conditions and drug management. Those in the MADDE sample entering nursing homes from the community (almost half of those entering nursing homes) had approximately a 40% lower 12-month mortality risk than the other half of the sample who entered nursing homes following a hospital stay. An effort was made in the analysis to unbundle the effects embedded within caregiver relationships. Those directly measured in the data set included education, health status, and age. There is overlap between relationship and caregiver age (highly correlated with spouse on the high end and with daughter on the low end). Age and education are also collinear, with older individuals generally having less education than their children. Caregiver education, age, and health status were separately considered in the original models, but they were not retained by the backward stepwise process that also included relationship. Attributes that may distinguish daughters from spouse caregivers, such as the comparative tolerance for the burdens of caregiving, were not directly measured. These are inferred from outcomes such as nursing home placements. Caregiver-perceived stress and burden was directly measured, but this measure was not associated with mortality risk, perhaps because of differential burden tolerance among caregivers. In any clinical application of the model, it would be prudent to consider caregiver physical and cognitive capabilities, educational level, and the ability to read and follow care management instructions. Further interacting with all of this are living arrangements and relationship. Spouses living with the subject likely have greater tolerance for the caregiver tasks than does a nonspouse, or caregivers not living with the subject. In short, the strengths and limitations of the caregiver must be appreciated and understood for their potential effects on patient outcomes, as well as for the nature and intensity of support that may be required.

A final issue is understanding the importance of age, gender, and race in the predictive models. We have presented these measures as individual-level attributes, but they potentially reflect dimensions that are subject to manipulations. For example, two dimensions are potentially adjusted for when age is controlled for. One of these is the physiological changes associated with aging. Increasing age increases the risk of death and the risk of complications from many procedures and treatments (see, e.g., Mangione et al., 1993). A second dimension is that of potential "ageism" in therapeutic choices for elderly patients. Separating ageism from changes in preferences among older patients further complicates the

interpretation of treatment differences across age groups or by cognitive status.

Women and men differ anatomically, physiologically and hormonally. Whether the observed differences are a function of physiology or of patient and family preferences has not been resolved. Stratifying analyses by gender and documenting treatment selection decisions would be a possible means for addressing these concerns in future medical effectiveness studies. Several parallels arise when race and ethnicity are viewed as risk factors. In addition to well-documented differences in disease prevalence, there is the concern that ethnic minorities may have preferences that lead to differences in treatment selection. Such differences were not documented in the MADDE data set. The data used did separately measure income and education, but any cultural differences in health practices and behaviors (e.g., smoking, blood pressure, cholesterol level, body-mass index, and alcohol intake) were not measured. Analyses of the effect of race or ethnicity on outcomes of care can be better accomplished in results stratified by race and ethnicity, and where behaviors, practices, preferences, and treatments are documented.

Although a number of limitations and further measurement refinements have been posited for the mortality prediction model, the current model's potential salience should not be ignored. First, a prognostic measure that can stratify patients with dementia into groups at differential risk for death is useful to providers counseling patients and their families about advance planning and palliative care options, such as current treatment and future interventions; decisions about hospitalization; and referrals for nursing home care, hospice, or home health care. Our model identifies a combination of factors that contribute to an end-stage prognosis. A clinician using this model would have a starting prognosis and treatment plan that could be modified for individual patients on the basis of their comorbid condition severity. The main point is that the current work opens up an examination of the issue of end-of-life care for persons with dementia.

Second, the measures in this model are available in many administrative systems, as well as national and catchment area surveys. This lends itself to replication tests (such as those suggested previously) and application in clinical trials or outcomes studies designed to test the efficacy of interventions directed to those with dementia at the end of life, or studies of community-support programs directed at the caregivers of those at the end of life.

ACKNOWLEDGMENTS

This article was supported in part by Grant IIRG-01-2390 from the Alzheimer's Disease and Related Disorders Association, and Contract 500-89-0069 from the Centers for Medicare & Medicaid Services. The authors remain solely responsible for the conclusions presented.

Address correspondence to Robert Newcomer, Department of Social & Behavioral Sciences, University of California, 3333 California Street, Suite 455, San Francisco, CA 94118. E-mail: rjn@itsa.ucsf.edu

REFERENCES

- Adler, N. E., Boyce, W., Chesney, M., Folkman, S., & Syme, S. (1993). Socioeconomic inequalities in health. No easy solution. *Journal of the American Medical Association*, 269, 3140-3145.
- Ash, A., & Shwartz, M., (1997). Evaluating the performance of risk-adjustment methods: Dichotomous outcomes. In L. I. Iezzoni (Ed.),

- Risk adjustment for measuring healthcare outcomes* (pp. 427–469). Chicago: Health Administration Press.
- Bennett, C., Greenfield, S., Aronow, H., Ganz, P., Vogelzang, N., & Elashoff, R. (1991). Patterns of care related to age of men with prostate cancer. *Cancer*, *67*, 2633–2641.
- Berkman, L. F., Leo-Summers, L., & Horowitz, R. (1992). Emotional support and survival after myocardial infarction. A prospective, population-based study of the elderly. *Annals of Internal Medicine*, *117*, 1003–1009.
- Charlson, M., Pompei, P., Ales, K., & MacKenzie, R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Disease*, *40*, 373–383.
- Cleves, M., Sanchez, N., & Draheim, M. (1997). Evaluation of two competing methods for calculating Charlson's comorbidity index when analyzing short-term mortality from administrative data. *Journal of Clinical Epidemiology*, *50*, 903–908.
- Covinsky, K. E., Justice, A. C., Rosenthal, G. E., Palmer, R. M., & Landefeld, C. S. (1997). Measuring prognosis and case mix in hospitalized elders. The importance of functional status. *Journal of Internal Medicine*, *12*, 203–208.
- Covinsky, K. E., & Landefeld, C. S. (1996). Using the biopsychosocial model in practice: Improving the care of patients with late life depression. *Journal of General Internal Medicine*, *11*, 249–250.
- Deyo, R., Cherkin, D., & Ciol, M. (1992). Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of Clinical Epidemiology*, *45*, 613–619.
- D'Hoore, W., Bouckaert, A., & Tilquin, C. (1996). Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *Journal of Clinical Epidemiology*, *49*, 1429–1433.
- Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, *196*, 129–136.
- Fillenbaum, G., Pieper, C., Cohen, H., Cornoni-Huntley, J., & Guralnik, J. (2000). Comorbidity of five chronic health conditions in elderly community residents: Determinants and impact on mortality. *Journal of Gerontology: Medical Sciences*, *55A*, M84–M89.
- Folstein, M., Folstein, S., & McHugh, P. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.
- Fried, L. P., Kronmal, R., Newman, A. B., Bild, D. E., Mittelmark, M. B., Polak, J. F., et al. (1998). Risk factors for 5-year mortality in older adults: The Cardiovascular Health Study. *Journal of the American Medical Association*, *279*, 585–592.
- Gauthier, S. (1998). Update on diagnostic methods, natural history and outcome variables in Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *9*(Suppl. 3), 2–7.
- Gill, T. M., Williams, C. S., & Tinetti, M. E. (1999). The combined effects of baseline vulnerability and acute hospital events on the development of functional dependence among community-living older persons. *Journal of Gerontology: Medical Sciences*, *54A*, M377–M383.
- Guralnik, J. M., Land, K. C., Blazer, D., Fillenbaum, G. G., & Branch, L. (1993). Educational status and active life expectancy among older blacks and whites. *New England Journal of Medicine*, *329*, 110–116.
- Iezzoni, L. I. (1997). Dimensions in risk. In L. I. Iezzoni (Ed.), *Risk adjustment for measuring healthcare outcomes* (pp. 43–168). Chicago: Health Administration Press.
- Justice, A. C., Covinsky, K., & Berlin, J. (1999). Assessing the generalizability of prognostic information. *Annals of Internal Medicine*, *130*, 515–524.
- Katz, S., Ford, A., Moskowitz, R., Jackson, B., & Jaffee, M. (1963). Studies of illness in the aged: The index of ADL, a standardized measure of biological and psychosocial functioning. *Journal of the American Medical Association*, *185*, 94–101.
- Kieszak S., Flanders, W. D., Kosinski, A., Shipp, C., & Karp, H. (1999). A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. *Journal of Clinical Epidemiology*, *52*, 137–142.
- Lawton, M., & Brody, E. (1969). Assessment of older people: Self-sustaining and instrumental activities of daily living. *The Gerontologist*, *9*(3), 179–186.
- Mangione, C., Marcantonio, E., Goldman, L., Cook, E., Donaldson, M., Sugarbaker, D., et al. (1993). Influence of age on measurement of health status in patients undergoing elective surgery. *Journal of the American Geriatrics Society*, *41*, 377–383.
- McCarty, H., Roth, D., Goode, K., Owen, J., Harrell, L., Donovan, K., et al. (2000). Longitudinal course of behavioral problems during Alzheimer's Disease: Linear versus curvilinear patterns of decline. *Journal of Gerontology: Medical Sciences*, *55A*, M200–M206.
- Miller, R., Clay, T., Fox, P., & Newcomer, R. (1999). Effects of the Medicare Alzheimer's disease demonstration on nursing home entry. *Health Services Research*, *34*, 691–713.
- Morrison, R., & Siu, A. L. (2000). Survival in end-stage dementia following acute illness. *Journal of the American Medical Association*, *284*, 1856–1860.
- Newcomer, R., Clay, T., Luxenberg, J., & Miller, R. (1999). Misclassification and selection bias when identifying Alzheimer's disease solely from Medicare claims records. *Journal of the American Geriatrics Society*, *47*, 215–219.
- Newcomer, R., Yordi, C., DuNah, R., Fox, P., & Wilkinson, A. (1999). Effects of the Medicare Alzheimer's disease demonstration on caregiver burden and depression. *Health Services Research*, *34*, 669–689.
- Walter, L. C., Brand, R. J., Counsell, S. R., Palmer, R. M., Landefeld, C. S., Fortinsky, R. H., et al. (2001). Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. *Journal of the American Medical Association*, *285*, 2987–2997.
- Yesavage, J., Rink, T., Rose, T., & Aday, M. (1983). Geriatric depression rating scale: Comparison with other self-report and psychiatric rating scales. In T. Crook, S. Ferris, & R. Bartusleds (Eds.), *Assessment in geriatric psychopharmacology* (pp. 153–167). New Canaan, CT: Mark Powley and Associates.
- Zarit, S., Reeves, K., & Bach-Peterson, J. (1980). Relatives of the impaired elderly: Correlates of feelings of burden. *The Gerontologist*, *20*, 260–266.

Received July 31, 2002

Accepted December 2, 2002

Decision Editor: Charles F. Longino, Jr., PhD