Journal of Gerontology: MEDICAL SCIENCES
2005, Vol. 60A, No. 4, 491–498

Predicting Death in the Nursing Home: Development and Validation of the 6-Month Minimum Data Set Mortality Risk Index

Davina Porock,[1] Debra Parker Oliver,[2] Steve Zweig,[3] Marilyn Rantz,[4] David Mehr,[3] Richard Madsen,[5] and Greg Petroski[6]

[1]School of Nursing, University of Nottingham, United Kingdom.
[2]School of Social Work,
[3]Department of Family and Community Medicine, School of Medicine,
[4]Sinclair School of Nursing,
[5]Department of Statistics, and
[6]Health Science Center, University of Missouri-Columbia.

Background. Currently, 24% of all deaths nationally occur in nursing homes making this an important focus of care. However, many residents are not identified as dying and thus do not receive appropriate care in the last weeks and months of life. The aim of our study was to develop and validate a predictive model of 6-month mortality risk using functional, emotional, cognitive, and disease variables found in the Minimum Data Set.

Methods. This retrospective cohort study developed and validated a clinical prediction model using stepwise logistic regression analysis. Our study sample included all Missouri long-term-care residents (43,510) who had a full Minimum Data Set assessment transmitted to the Federal database in calendar year 1999. Death was confirmed by death certificate data.

Results. The validated predictive model with a c-statistic of .75 included the following predictors: a) demographics (age and male sex); b) diseases (cancer, congestive heart failure, renal failure, and dementia/Alzheimer's disease); c) clinical signs and symptoms (shortness of breath, deteriorating condition, weight loss, poor appetite, dehydration, increasing number of activities of daily living requiring assistance, and poor score on the cognitive performance scale); and d) adverse events (recent admission to the nursing home). A simple point system derived from the regression equation can be totaled to aid in predicting mortality.

Conclusions. A reasonably accurate, validated model has been produced, with clinical application through a scored point system, to assist clinicians, residents, and family members in defining good goals of care around end-of-life care.

TWENTY-FOUR percent of Americans die in nursing homes (1). Accordingly, dying in the nursing home has received an increasing amount of attention and scrutiny (2–4). Somewhat contradictory, the major focus of care in nursing homes has been restorative and rehabilitative to meet regulatory requirements and to generate greater reimbursement (4). Despite the number of deaths that occur in nursing homes, there is much research and clinical evidence to suggest that the care of residents at the end of their life is commonly unsatisfactory (5–7).

There are multiple potential benefits in recognizing nursing home residents at great risk of dying. This recognition should precipitate a thorough discussion of prognosis and goals of care. For those residents or family members choosing a palliative course, the focus of care might be on settling issues with family members and symptom management, perhaps foregoing surgical procedures or uncomfortable hospitalizations. For those choosing length of life as the highest priority, this serious prognosis recommends intensive investigation and attempted reversal of underlying problems. Knowing that a resident is at the end of life is fundamental to ensuring that their wishes are known and respected and that the quality of their life and death reflects their choices.

Every nursing home in the United States that receives Medicare or Medicaid funding for its residents is required to complete a full Minimum Data Set (MDS) assessment of functional, emotional, cognitive, and disease status on each resident a) within 14 days of admission, b) annually, and c) when any significant event or change in condition occurs. Further, a shortened assessment is completed every 90 days following admission. A recent study (8) found that just 4.5% of new admissions were designated as being at the end of life (expected to die within 6 months) as recorded on the MDS. Almost 1 in 5 residents not designated as near the end of life also died within 6 months of admission, thus demonstrating that (at least up to 6 months ahead of time) we do not recognize a substantial number of residents as dying.

The purpose of this study was to identify the MDS indicators that best predict 6-month mortality in nursing home residents to coincide with the Medicare hospice benefit timeframe. The predictive model was developed to inform research and practice with the goal of facilitating end-of-life planning and medical decision making.

Methods

The study was a secondary analysis involving linked MDS and death certificate data from the State of Missouri. Approval from the University of Missouri Health Sciences

Institutional Review Board at the University of Missouri-Columbia was obtained. The MDS provided the demographic and clinical variables to be considered as predictors of mortality, whereas the death certificates provided the most precise information about the date and place of death. The sample consisted of all Missouri long-term-care residents in nonhospital-based facilities who were over the age of 65 at the time of their first full assessment in 1999 and had a full MDS assessment transmitted to the Federal database in calendar year 1999.

Instrument

The MDS is a comprehensive standardized assessment instrument of more than 400 items for all long-term-care residents in facilities that receive Medicare or Medicaid funding (9). A full assessment is required within 14 days of admission, annually, and after significant change in resident status There is growing evidence in the literature of the reliability and validity of many of the items of the MDS instrument and data (9–15).

Study Variables

The dependent variable in all analyses was death at 6 months following the first full assessment in 1999. The potential predictors of mortality were items from the MDS survey that represented factors from previous research and clinical experience associated with the dying process. The team, consisting of experienced researchers and clinicians, identified 50 individual MDS items as having a potential relationship with prognosis and/or mortality; these fell into four main categories: 1) demographics (e.g., age, sex), 2) diseases (e.g., cancer, chronic obstructive pulmonary disease, congestive heart failure), 3) clinical signs and symptoms (pain, shortness of breath, weight loss, activities of daily living [ADLs], cognitive function), and 4) adverse events (e.g., falls, infections, hospitalizations, loss of a spouse). The cognitive performance scale (CPS) was used to assess cognitive function as devised by Morris and colleagues (12). Independence in ADLs was assessed using a composite score of seven ADLs from the self-performance items from the MDS as devised by Morris and colleagues (16). These seven ADLs were bed mobility, transfer between surfaces (e.g., bed to chair), locomotion on unit, dressing, eating, personal hygiene, and toilet use.

Data Set Creation

The data from the MDS assessments from the 1999 calendar year were matched with Missouri death certificate data from January 1999 through December 2000 to definitively identify residents who died. Records from residents in hospital-based nursing facilities were excluded from the analyses, as were resident records with missing last name, sex, or Social Security Number. Details of this matching procedure can be found in the Appendix.

Data Analysis

There were 43,510 residents in the data set. Seventy-five percent of the data was randomly selected to become the developmental data set with the remaining 25% set aside for validation. From this developmental data set, 20 randomly selected independent subsamples of about 11,000 residents (one third of the developmental set) were created. One reason for doing this was to avoid having so much power that we were observing statistically significant differences that were so small as to be of no clinical relevance. A second reason for looking at multiple subsets of the developmental set was to avoid problems associated with using stepwise selection of predictor variables. Variables that appear to be significant in one subset of the data may not appear to be significant in other subsets. By looking for predictor variables that were consistently selected from one subset to another, it is more likely that a model based on these predictors will be predictive in the validation data.

Many of the 50 variables listed as potential predictors were simple dichotomous variables. For those variables that were not dichotomous, but were at least ordinal, we investigated the form of the relationship of the predictor, using residual plots from generalized additive models to help determine the best form (17). Next we considered all variables univariately to determine if any one, by itself, was a useful predictor of 6-month mortality. In view of the relatively large power when dealing with 11,000 residents, only variables significant at the .01 level were retained for further consideration in the multiple-predictor models. The remaining steps of the analysis are described with the resulting findings.

Of the 50 variables selected from the MDS for analysis, an initial screening showed that 26 had a significant relationship with 6-month mortality. Using all variables that passed the initial screening, we used a stepwise logistic regression procedure to find which variables would be retained in a multivariable predictor model. Due to sampling variation, a variable in one model might not be retained in a subsequent fitting of a model based on a different sample. For that reason, we tested the variables in the 20 randomly selected subsamples to find which variables were retained every time, all but one time, and so forth.

To determine which variables to include in a final model, we considered two factors: how often a variable was selected by the stepwise procedure, and the step at which the variable was selected. To this end, each variable received a score based on the frequency and order with which they entered each model, i.e., the first variable selected by stepwise regression received the score of 20 points, the second variable 19 points, and so on. A total score was the sum of points for each variable across the 20 models. Table 1 details the frequency with which each variable entered the models and the total points scored.

The cutoff point to determine whether a variable could be considered a reliable predictor was decided on the basis of the frequency that it appeared in the 20 subsamples as well as the total score. A break clearly appeared after the 14th variable, so the first 14 variables were kept for further fitting of the model.

After the set of variables to be kept had been determined, all possible two-way interaction terms were defined for possible inclusion in the final model. A stepwise procedure again was used on each developmental subsample with the condition that all main effects be forced into the model before the interactions were considered. Two interactions

Table 1. Frequency and Scores for Variable Entry Into 20 Subsamples of the Development Data Set

Variable Ranking	Variable Name	Frequency of Model Entry	Total Score of Ordered Entry
1	Activities of daily living (ADLs) requiring assistance (0–7)	20	379
2	Shortness of breath	20	338
3	Cancer diagnosis	20	328
4	Recent admission to nursing home	20	322
5	Poor appetite	20	287
6	Male sex	20	277
7	Deteriorating condition	20	274
8	Weight loss	20	249
9	Chronic heart failure	20	236
10	Age	20	190
11	Renal failure	20	180
12	Cognitive Performance Scale score (0–6)	19	119
13	Alzheimer's disease or dementia	18	97
14	Dehydrated, definition	17	129
15	Pain, moderate to severe nearly every day	13	76
16	Infection, pneumonia	13	56
17	Pain, excruciating every day	11	52
18	Sleep, definition	9	49
19	Parkinson's disease	6	30
20	Infection, tuberculosis	5	15
21	Affect change	4	14
22	No. of times hospitalized in the past 90 days	2	10
23	Infection, *Clostridium difficile*	2	7
24	Communication problems	1	7
25	Edema	1	6
26	Infection, antibiotic-resistant infection	1	5

consistently appeared in these analyses: ''cancer and age'' and ''admission to the nursing home and deterioration.'' With a diagnosis of cancer, the risk of dying was greater the younger the resident was. The interaction between admission and deterioration suggested that the effect of these two variables was not simply additive. Thus we had 14 variables and 2 interactions to fit the model.

After deciding on the variables to be entered into the predictive model, we used all of those variables with the entire developmental set (32,484 observations) to estimate the final parameters and validate the model. To account for possible dependence of outcomes within the same home, we used the Generalized Estimating Equations (GEE) (18) approach, and modeled the covariance using an exchangeable (or compound symmetry) model.

We compared the ordinary coefficients and the GEE coefficients and found them to be quite close. Table 2 shows the c-statistics for four cases. Using coefficients from the model found using the developmental data, we found the c-statistic when the model was fit to the developmental data and when the same model was used with the validation data. The c-statistic is a measure of the predictive value of the logistic regression model with values ranging from 0 to 1, with large values indicative of better predictive value. This comparison was repeated for the model using the ordinary coefficients and the model using GEE coefficients. The relatively small change when fitting the models to the validation data indicates that the model validates quite well.

Other measures of model fit related to measures of discrimination and calibration (19). Discrimination is the ability to separate the successes from the failures, i.e., for higher values of estimated probability, there should be

Table 2. Summary Table of Developmental Validation Data Sets

Data Set	Method	c-Statistic
Development	Ordinary	0.762
Development	Generalized Estimating Equations	0.762
Validation	Ordinary	0.753
Validation	Generalized Estimating Equations	0.753

Table 3. Comparison of the 1999 Missouri Minimum Data Set Data (Including Developmental and Validation Subsets) With National Data on Selected Demographics

Variable	Total Sample ($N = 43{,}510$)	Developmental Data ($N = 32{,}599$)	Validation Data ($N = 10{,}911$)	National Data*
Age, y				
65–74	13.00%	13.00%	12.77%	12.00%
75–84	36.48%	36.38%	36.79%	32.00%
85+	50.51%	50.53%	50.44%	46.00%
Sex				
Male	26.44%	26.31%	26.84%	28.00%
Female	73.56%	73.69%	73.16%	72.00%
Race				
White (non-Hispanic)	91.85%	91.86%	91.84%	87.10%
Black (non-Hispanic)	7.60%	7.58%	7.64%	10.40%
Other/unknown	0.50%	0.50%	0.50%	2.50%

Note: *Gabrel CS, Jones A. The National Nursing Home Survey: 1997 summary. Vital and Health Statistics-Series 13: Data from National Health Survey, 2000;147,1–21.

Table 4. Validated Logistic Regression Model of 6-Month Mortality in Nursing Home Residents

Variable	*df*	Estimate	Odds Ratio Estimates	95% Wald Confidence Limits	
Intercept	1	−5.8475			
Activities of daily living	1	0.2467	1.280	1.254	1.306
Shortness of birth	1	0.7849	2.192	2.019	2.381
Loss of appetite	1	0.4634	1.589	1.496	1.668
Sex	1	0.5885	1.801	1.689	1.921
Weight loss	1	0.4366	1.547	1.428	1.676
Chronic heart failure	1	0.3771	1.458	1.367	1.555
Renal disease/failure	1	0.6183	1.856	1.632	2.110
Cognitive performance scale	1	0.0907	1.095	1.073	1.117
Alzheimer's disease or dementia	1	−0.2399	0.787	0.737	0.840
Dehydrated	1	0.4603	1.585	1.416	1.774
Cancer†	1	5.2889			
Age†	1	0.0269			
Cancer * Age†	1	−0.0523			
Admission†	1	0.8379			
Deteriorated†	1	0.6904			
Admission * Deterioration†	1	−0.5057			

Note: †Odds ratios cannot be calculated for variables included in interaction terms.

*Interaction.

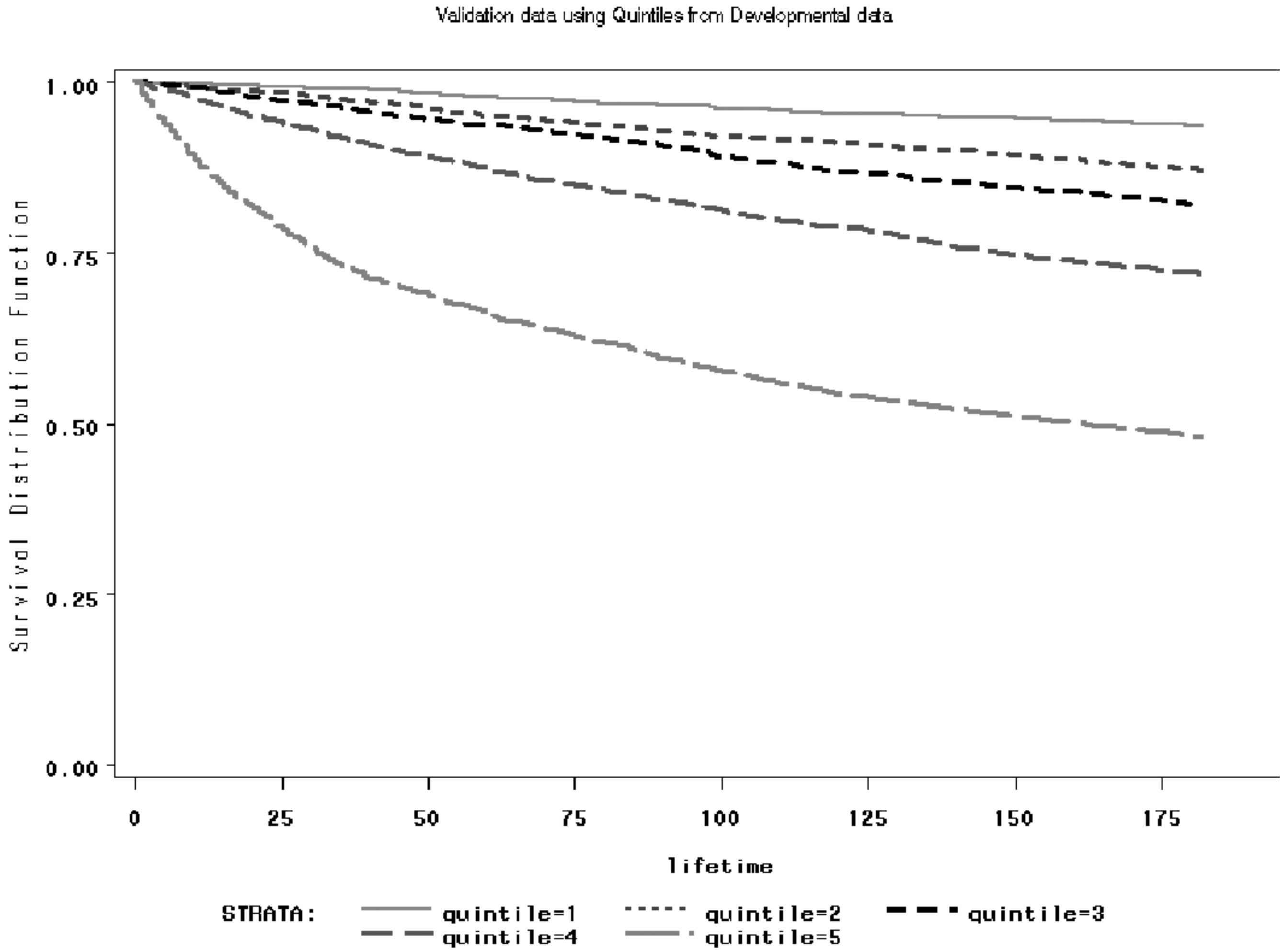

Figure 1. Kaplan–Meier survival curves for each quintile of risk of dying using the developmental data set.

a very high proportion of residents who die, whereas for lower values of estimated probability, a low proportion should die. We then compared the proportion of deaths in the highest quintile of the estimated probabilities relative to the proportion in the lowest quintile. For the developmental set, the ratio is 8.91; for the validation set, this drops only to 8.12. Calibration, which determines whether the predicted and observed mortality are similar over the range of predicted risk, was checked by looking at the observed and expected proportions of death within each decile of probability values. This was summarized by a Hosmer–Lemeshow statistic and, even with the large sample size and associated power, the results were satisfactory. For the developmental set, the p value for the Hosmer–Lemeshow test was .58 and for the validation set it was .16.

Results

Demographics

Demographics of the total sample of 43,510 residents and the development and validation subsets were compared with national statistics to determine the generalizability of the findings. Table 3 summarizes those findings. Overall, 23% of the residents died in the 6 months following their first full assessment of 1999/2000. The final validated 16-item model for predicting the risk of death within 6 months is presented in Table 4.

To illustrate how well the quintiles of the estimated probability of dying (or the risk of dying) relate to survival, we made Kaplan–Meier survival curves for each quintile within the validation set. The plot shown in Figure 1 illustrates how the estimated survival curves are successively lower as the quintiles of risk get higher.

Implications for Practice–The 6-Month MDS Mortality Risk Index

Having identified an optimal set of predictors, we derived a 6-month mortality risk index from the final logistic model. The 6-month MDS Mortality Risk Index (MMRI) is a simple algorithm that assists in using selected MDS items to determine a resident's risk of dying within the next 6 months. The algorithm was guided by the results of the logistic regression analysis but is not identical to the regression model, and is an additive scale with weights

6-Month MDS Mortality Risk Index Point System

Age Without Cancer		Age With Cancer			
≤69	1 pt				
70–78	2 pts	≤74	8 pts		
79–88	3 pts	75–84	7 pts		
89–98	4 pts	85–94	6 pts		
99+	5 pts	≥95	5 pts		
				Age points	_______

CPS score			
0–1	0 pts		
2–4	1 pt		
5–6	2 pts		
		CPS points	_______

Admission and/or deteriorating score			
Admission only	3 pts		
Deteriorating	3 pts		
Both admission and deteriorating	4 pts		
		Admission/deteriorating points	_______
		ADL score	_______
		Shortness of breath (3 pts)	_______
		Poor appetite (2 pts)	_______
		Male (2 pts)	_______
		Weight loss (2 pts)	_______
		Chronic heart failure (2 pts)	_______
		Renal failure (2 pts)	_______
		Dehydrated (2 pts)	_______
Subtotal number of points			_______
		Alzheimer's (subtract 1)	_______
Grand total number of points			_______

Figure 2. Algorithm for calculating individual risk based on points derived from the logistic regression analysis–Minimum Data Set Mortality Risk Index.

being assigned by rounding the regression coefficients from the final logistic model. The MMRI point system can be found in Figure 2. Figure 3 compares the actual deaths with the predicted deaths using the point system, and demonstrates the validity of the system. To illustrate the utility of the MMRI, Table 5 presents the mean proportion of deaths that occurred in the 6 months following assessment.

The following examples illustrate how the MMRI could work. A 90-year-old man with Alzheimer's disease who has a score of 6 on the CPS (reflecting advanced cognitive impairment) and a score of 5 on the ADL scale would have a total 14 points. Table 5 indicates that, in this point range, about 20% of residents would be expected to die in the following 6 months. Whereas, an 82-year-old man with a CPS score of 6, poor appetite, weight loss, a "totally dependent" score on the ADL scale, and assessed to be deteriorating would receive a total score of 22, and would have a nearly 75% chance of dying within the next 6 months. If these conditions were judged irreversible, it would certainly be appropriate to plan for end-of-life care.

Discussion

The validated predictive model of 6-month mortality in nursing home residents included variables that are not surprising to those working in this area. It is possible that some predictors are potentially reversible, for example,

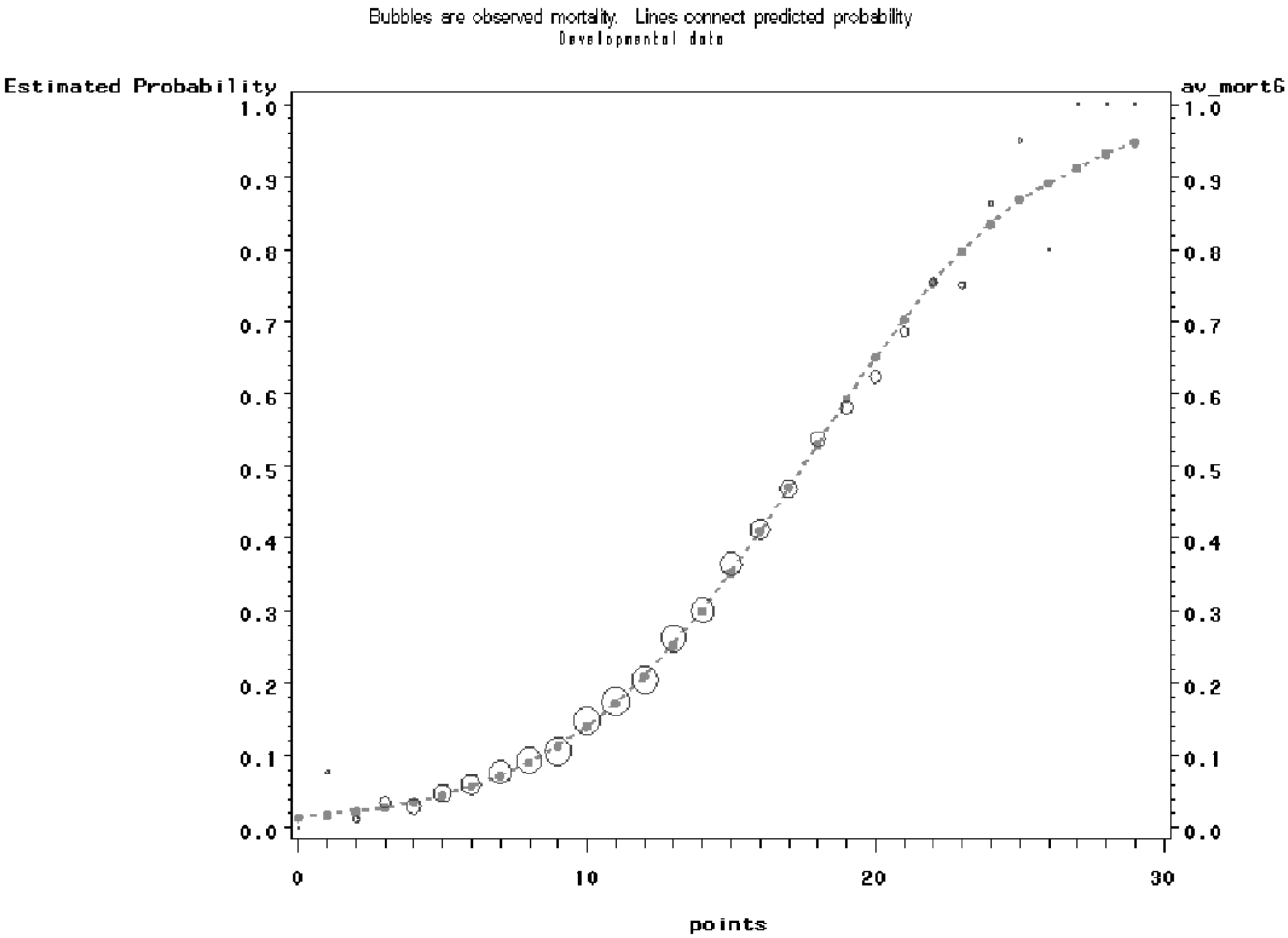

Figure 3. Comparison between actual mortality and predicted mortality using the point system.

artificial feeding and/or fluids for weight loss and hydration. However, the potential for reversibility must be considered in the light of whether such intervention would in fact be futile, what the risk of causing harm (e.g., to dignity or risk of infection) might be, and the wishes of the resident (20). The finding that Alzheimer's disease was protective in terms of the risk for death in 6 months at first seems counter-intuitive. However, people with dementia are more likely to enter a nursing home because of problems in behavior, wandering, and incontinence rather than through loss of function due to serious medical illness such as cancer or heart disease. Therefore, on admission or at any given time during the course of their stay in the nursing home, they might be less likely to die of an immediately life-threatening illness. That being said, those persons with advanced dementia will eventually die of diseases that result from that chronic neurodegenerative condition—failure to thrive with resulting malnutrition, falls, fractures, and infections.

Table 5. Proportion of Mortality Within 6 Months Based on Minimum Data Set Mortality Risk Index Point Score Ranges

Set	Points	*N*	% Died	Sensitivity, %*	Specificity, %	False-Positive	False-Negative
Development	0–10	14,239	9.1	100.0	0.0	76.9	—
	11–14	10,743	23.0	82.6	51.8	66.1	9.1
	15–18	5438	42.8	49.7	84.9	50.4	15.1
	19–21	1466	61.9	18.6	97.3	32.5	20.0
	22+	598	81.3	6.5	99.6	18.7	22.0
Validation	0–10	4710	10.1	100.0	0.0	76.3	—
	11–14	3589	22.9	81.6	51.0	66.0	10.1
	15–18	1854	43.4	49.6	84.3	50.6	15.6
	19–21	520	58.1	18.4	96.9	35.1	20.7
	22+	209	81.8	6.6	99.5	18.2	22.5

Note: *Sensitivity, specificity, false-positive, and false-negative rates are given for a rule that ''predicts'' that a person with a point score in this category or higher would die within 6 months. For example, if a person was categorized based on a score of 11 or higher, the sensitivity would be 82.6%.

Several attempts using the MDS to predict mortality in nursing home residents have been published in recent years. Abicht-Swensen and Debner (21) conducted a retrospective study of 199 residents who had been referred to hospice from 24 Minnesota nursing homes. The main finding of their study was the strength of the relationship between short-term mortality and a decline in functional status in the areas of cognitive functioning, communication, ADLs, incontinence, and nutrition. These findings corroborate with ours but, unfortunately, their study focused on residents already referred to hospice and, therefore, already recognized as dying.

Hirdes, Frijters, and Teare (22) created the MDS-CHESS (Changes in Health, End-stage and Symptoms and Signs) Scale. Their scale included items from three sections of the MDS: declining health status, end-stage disease, and symptoms and signs of medical problems. Many similarities are found between this model and ours even though the population studied was Complex Continuing Care hospital patients rather than long-term-care residents. The main limitation noted by Hirdes and colleagues was their inability to verify death after discharge from the Complex Continuing Care hospital.

The linking of the MDS and death certificate data is a particular strength of our study. Furthermore, the transformation of the logistic regression model to a point system provides greater clinical utility for decision making. Unlike Hirdes and colleagues, we decided not to use the end-stage disease item of the MDS although it has excellent prognostic value for those who are so designated (8). What our analysis found was that, despite the validity of the item when it is used, it was not used reliably in the MDS. It is fair to say that there are many complex and varied reasons why a physician or an MDS nurse would not choose to document that a resident has "six or fewer months to live," even if it were suspected.

In a study of 1-year survival in nursing home residents (2003), Flacker and Kiely (23) linked MDS data with the National Death Index to overcome the problems in tracking deaths, and also used developmental and validation data sets. The principal difference (other than time) between Flacker and Kiely's model and ours was the stratification of long-stay residents and new admissions producing two models, whereas we incorporated the predictor "recent admission" into our model.

We chose a 6-month timeframe to calculate risk for mortality because it has clinically useful application in identifying residents who may benefit from specialist palliative care or hospice services. In our study, we found that many residents were at high risk of dying in 6 months. Overall, 23% of the residents died within 6 months of their first full assessment in 1999–2000. Included in that group were many residents who were most at risk. Identifying those most at risk of death—in other words, making the diagnosis of dying—is the first step in ensuring that the goals of care are appropriate and the wishes of the resident are known, documented, and respected.

Several aspects of our work support the validity of the MMRI. First is the use of state death certificate data to confirm the outcome variable of death and the strong linkage of these outcomes with the MDS data. Second, the model development was rigorous with the use of multiple development data sets and reserved data for validation of the final model, thus producing a reliable and valid method of prediction.

One particular limitation of our study is the lack of ethnic diversity in the sample; specifically, the proportion of African American and Hispanic elderly persons in Missouri nursing homes is not as high as in national statistics. Bearing in mind these strengths and limitations, future research needs to focus on multi-state studies using the MMRI, the transferability of the MMRI to non-MDS settings, the inclusion of predictors not currently found on the MDS (for example, social cues), and the impact that prediction makes on decision making and goal setting in the nursing home.

High quality end-of-life care cannot be achieved if the diagnosis of dying occurs only hours or days before death. Therefore, the ability to predict accurately the transition to the end of life is vital. The particular significance of this work was that it focused on MDS data that are routinely collected by nursing homes and are, therefore, already part of the workload, not an additional imposed expectation. The heightened awareness of a resident's transition to the end of their life may in itself create the impetus for a change in goals of care.

Acknowledgments

The members of the University of Missouri MDS and Nursing Home Quality Research Team gratefully acknowledge the support of the Missouri Department of Health and Senior Services. Research activities were partially supported by contract #C-5-31455 from the Missouri Division of Aging to the Sinclair School of Nursing and Biostatistics Group of the School of Medicine, University of Missouri-Columbia. Opinions are those of the authors and do not represent the Missouri Department of Health and Senior Services.

Address correspondence to: Davina Porock, PhD, RN, School of Nursing, Faculty of Medicine and Health Sciences, University of Nottingham, A Floor, Queen's Medical Centre, Nottingham, England NG7 2HA. E-mail: Davina.Porock@nottingham.ac.uk

References

1. Brown, U. S. o. M. (2000). *Brown Atlas Site of Death 1989-1997* [url]. Brown University. Retrieved February 2000, from the World Wide Web: http://www.chcr.brown.edu/dying/2001DATA.HTM
2. Keay TH, Taler GH, Redman L, Levenson SA. Assessing medical care of dying residents in nursing homes. *Am J Med Qual.* 1997;12:151–156.
3. Engle VF. Care of the living, care of the dying: reconceptualizing nursing home care. *J. Am Geriatr Soc.* 1998;45:1772–1774.
4. Lawhorne LW. End-of-life care in nursing homes—is a good death compatible with regulatory compliance? *Bioethics Forum.* 1999;15:23–28.
5. Keay TH, Fredman L, Taler GA, Datta S, Levenson SA. Indicators of quality medical care for the terminally ill in nursing homes. *J Am Geriatr Soc.* 1994;42:853–860.
6. Fowler FJ, Coppola KM, Teno JM. Methodological challenges for meshing quality of care at end of life. *J Pain Sympt Manage.* 1999;17: 114–119.
7. Hanson LC, Danis M, Garrett J. What is wrong with end-of-life care? *J Am Geriatr Soc.* 1997;45:1339–1344.
8. Porock D, Parker-Oliver D, Zweig S, Rantz M, Petroski G. A Profile of residents admitted to long-term-care facilities for end-of-life care. *JAMDA.* 2003;4:16–22.
9. Morris JN, Hawes C, Fries B, Phillips C, Mor V, Katz S, et al. Designing the national resident assessment instrument for nursing homes. *Gerontologist.* 1990;30:293–307.
10. Hawes C, Phillips CD, Mor V, Fries B, Morris JN. MDS data set should be used for research. *Gerontologist.* 1992;32:563–564.
11. Hawes C, Morris JN, Phillips CD, Mor V, Fries BE, Nonemaker S. Reliability estimates for the minimum data set for nursing home assessment and care screening (MDS). *Gerontologist.* 1995;35:172–178.
12. Morris JN, Fries BE, Mehr DR, Hawes C, Phillips C, Mor V, et al. MDS cognitive performance scale. *J Gerontol A Biol Sci Med Sci.* 1994; 49A:M174–M182.
13. Phillips CD, Chu CW, Morris JN, Hawes C. Effects of cognitive impairment on the reliability of geriatric assessments in nursing homes. *J Am Geriatr Soc.* 1993;41:136–142.

14. Phillips CD, Hawes C, Mor V, Fries BE, Morris JN. *Evaluation of the nursing home resident assessment instrument*: Executive summary. Report to health care financing administration by the Research Triangle Institute 1996.
15. Morris JN, Nonemaker S, Murphy K, Hawes C, Fries BE, Mor V, et al. A commitment to change: revision of HCFA's RAI. *J Am Geriatr Soc*. 1997;45:1011–1016.
16. Morris JN, Fries BE, Morris SA. Scaling ADLs within the MDS. *J Gerontol A Biol Sci Med Sci*. 1999;54:M546–M553.
17. Hastie TJ, Tibshirani RJ. *Generalized additive models*. New York: Chapman and Hall; 1990.
18. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:13–22.
19. D'Agostino RB Sr, Griffith JL, Schmid CH, Terrin NT. *Measures for Evaluating Model Performance*. Proceedings of the Biometrics Section, American Statistical Association; 1997.
20. Finucane TE, Christmas C, Travis K. Tube feeding in patients with advanced dementia: a review of the evidence *JAMA*. 1999;282: 1365–1370.
21. Abicht-Swensen LM, Debner LK. The minimum data set 2.0: a functional assessment to predict mortality in nursing home residents. *Am J Hosp Palliat Care*. 1999;16:527–532.
22. Hirdes JP, Frijters DH, Teare GF. The MDS-CHESS scale: a new measure to predict mortality in institutionalized older people. *J Am Geriatr Soc*. 2003;51:96–100.
23. Flacker JM, Kiely D. Mortality-related factors and 1-year survival in nursing home residents. *J Am Geriatr Soc*. 2003;51:213–221.

Received September 3, 2003
Accepted November 14, 2003
Decision Editor: John E. Morley, MB, BCh

Appendix

Details of Linking Missouri Minimum Data Set (MDS) Records With Death Certificate Data

Three methods were used to determine whether residents were alive or dead at 6 months after the first full assessment in 1999. First, we determined that deaths recorded in the MDS were accurate, and assigned those residents as dead. Second, we assigned residents with continued MDS assessments as alive. Finally, we linked the remaining MDS assessments with Missouri death certificate data for 1999 and 2000. The overwhelming majority of MDS records and the death certificate data contain the individual's Social Security Number (SSN), date of birth, sex, and first and last names. Thus, the SSN provided the primary link between the two data sets. To simplify the problem of positively matching long-term care residents to the death certificates, we excluded resident records with missing or invalid SSN, missing or incorrectly coded sex variable, or missing last name. This process resulted in the exclusion of 128 residents from the analysis. In the 1999 and 2000 data, 0.2% of death certificates had missing or invalid SSNs. The need to positively determine the date of death dictated that these records could not be excluded. Thus, when the death certificate SSN was missing, matching was based on name, date of birth, and sex. All potential computer matches for these cases were finally reviewed ''by hand.'' From the MDS–death certificate record linkage, each resident's date of death and, therefore, survival time from their first full assessment in 1999 was determined. When the linkage failed, we deduced that the resident was still alive at the close of calendar year 2000. Thus, we determined whether each of these residents survived at least 6 months beyond the assessment date. An additional 552 long-term-care residents were excluded from the analysis because the first full assessment of 1999 coincided with their date of death, thus yielding zero survival times.