Comparing Self-Reported Health Status and Diagnosis-Based Risk Adjustment to Predict 1- and 2 to 5-Year Mortality

Kenneth Pietz and Laura A. Petersen

Objectives. To compare the ability of two diagnosis-based risk adjustment systems and health self-report to predict short- and long-term mortality.

Data Sources/Study Setting. Data were obtained from the Department of Veterans Affairs (VA) administrative databases. The study population was 78,164 VA beneficiaries at eight medical centers during fiscal year (FY) 1998, 35,337 of whom completed an 36-Item Short Form Health Survey for veterans (SF-36V) survey.

Study Design. We tested the ability of Diagnostic Cost Groups (DCGs), Adjusted Clinical Groups (ACGs), SF-36V Physical Component score (PCS) and Mental Component Score (MCS), and eight SF-36V scales to predict 1- and 2–5 year all-cause mortality. The additional predictive value of adding PCS and MCS to ACGs and DCGs was also evaluated. Logistic regression models were compared using Akaike's information criterion, the c-statistic, and the Hosmer–Lemeshow test.

Principal Findings. The c-statistics for the eight scales combined with age and gender were 0.766 for 1-year mortality and 0.771 for 2–5-year mortality. For DCGs with age and gender the c-statistics for 1- and 2–5-year mortality were 0.778 and 0.771, respectively. Adding PCS and MCS to the DCG model increased the c-statistics to 0.798 for 1-year and 0.784 for 2–5-year mortality.

Conclusions. The DCG model showed slightly better performance than the eight-scale model in predicting 1-year mortality, but the two models showed similar performance for 2–5-year mortality. Health self-report may add health risk information in addition to age, gender, and diagnosis for predicting longer-term mortality.

Key Words. Risk adjustment, diagnostic cost groups, adjusted clinical groups, mortality, health self-report

Assessing the overall health status of a patient population is an important problem for medical plans and health care systems. Ensuring that the appropriate equipment, financial resources, and human resources are available for care of the patients requires careful planning. In this paper, we evaluate the ability of several types of patient-level risk adjustment data in a large database

to predict mortality. Because the majority of health care services are consumed in the last 6 months of life (McCall 1984; Gaumer and Stavins 1992; Lubitz and Riley 1993), mortality prediction can be an important feature of health care budgeting and planning.

Risk adjustment is widely used for predicting future costs and for other uses in medical plans and health care systems. Most risk adjustment schemes are based on patients' demographics and diagnoses. Diagnosis-based risk adjustment is based upon the fact that patients with certain groups of diagnoses have been found to have similar cost patterns. The ability of various diagnosis-based risk adjustment systems to predict cost has been evaluated in several articles (Lamers and van Vliet 1996; Rosen et al. 2001; Pietz, Byrne, and Petersen 2006).

Research on the ability of diagnosis-based risk adjustment systems to predict clinical outcomes such as mortality is less well developed. Recently, we compared two widely used such systems, Adjusted Clinical Groups (ACGs) and Diagnostic Cost Groups (DCGs), to predict nursing home care and mortality in the same year as the diagnoses used (Petersen et al. 2005). We found that DCGs outperformed ACGs in the Department of Veterans Affairs (VA) data.

Diagnosis-based risk adjustment systems capture only the provider's coding of the patient's medical condition and are dependent upon the patient having a health care encounter for which codes are recorded. Therefore, conditions that may significantly impact the risk of death may not be coded and thus not available for risk prediction.

Another potential source of information about patients' health status is provided by health self-report. Health self-report may contain important medical information about the patient unknown to the provider, and may therefore supplement diagnostic information (Hornbrook and Goodman 1996). Also, because some veterans use the VA for only part of their health care needs, diagnostic information found in VA data may not provide a complete picture of the patient's health (Byrne et al. 2006). The ability of health self-report data to predict mortality has been the subject of considerable

research (Idler and Angel 1990; Idler and Benyamini 1997; Inouye et al. 1998; Schoenman, Hayes, and Cheng 2001).

In this article, we compare the ability of ACGs, DCGs, and 36-Item Short Form Health Survey for veterans (SF-36V) summary measures combined with age and gender to predict short- and long-term all-cause mortality. The SF-36V is a slightly modified short form 36 (SF-36) for veterans, described in detail later. We also compare the added value of combining SF-36V summary measures and individual scales to ACGs and DCGs. The SF-36V instrument measures entirely different information, such as limitations in physical activities and social functioning that may provide additional predictive power. The null hypothesis is that the SF-36V data contribute no additional health information. We are not aware of any research that compares these three sets of variables for prediction of long-term mortality. The VA health care system provides an opportunity for this type of study because we are able to ascertain vital status even for patients who have left the VA health care system (Cowper et al. 2002).

METHODS

Study Population and Data Source

The data for this study were a population of 78,164 VA beneficiaries in one network of VA hospitals in the Pacific Northwest. The derivation of this population has been described elsewhere (Pietz et al. 2004). Briefly, the file contains all patients who had some medical service in the network during fiscal year (FY) 1998 (October 1, 1997–September 30, 1998) and who had a primary care provider assigned. The focus of this article will be on the 35,337 patients who voluntarily completed and returned an SF-36V form. They constituted a 45.2 percent response rate among patients who satisfied the other criteria.

The analysis was performed on two subpopulations. The first is the 77,473 patients who survived the baseline year (FY 1998), of whom 35,202 completed an SF-36V. The second is the 74,854 patients who survived FY 1999, of whom 34,043 completed an SF-36V.

Risk Adjustment

ACGs. The ACG methodology is one of several diagnosis-based risk adjustment systems developed to predict utilization of medical resources, using the fact that patients who have certain groups of diagnoses tend to have similar utilization patterns (Starfield et al. 1991). Originally developed to

predict outpatient care, the methodology now uses both inpatient and outpatient International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes for a period of time (FY 1998 in this case). The codes are grouped into 106 into mutually exclusive ACGs, which are designed to predict total cost. Version 5.0 of the ACG software was used in this study (ACG 2001). The Major Extended Diagnostic Groups (MEDCs) provided by the ACG software were also used to obtain general information about the patient population.

The ACGs have been used for many years to predict cost and evaluate provider performance (Reid et al. 2001; Pietz, O'Malley et al. 2002; Pietz, Ashton 2004). The ACG methodology is designed to classify patients in a general population, including pediatric and obstetrics patients, which are not found in the VA population. Of the 106 original ACGs, 53 were found to apply to this population of veterans. We added age and gender to ACGs for risk adjustment. For modeling, the ACGs with fewer than 0.1 percent of the population were combined with a closely related adjacent group, and all ACGs with no events were combined into a single group. The appendix contains tables of ACGs that were combined for the two outcomes.

DCGs. Another widely used diagnosis-based risk adjustment system is DCGs (Ash et al. 2000). The DCGs were originally developed to predict inpatient cost. Like ACGs, they also use 1 year of inpatient and outpatient ICD-9 CM codes to classify patients into groups with similar cost patterns. The diagnoses are classified into mutually exclusive DCGs, based on predicting future inpatient and outpatient cost. In contrast to ACGs, the DCGs are ranked according to cost, and thus higher numbered DCGs indicate increasing illness burden and severity. The software algorithm for DCG classifications has been developed using three different populations of patients: private insurance holders, Medicare beneficiaries, and Medicaid beneficiaries. We used the Medicare-derived concurrent DCG groupings, as the Medicare population most closely resembles the VA population. The DCGs were assigned using DCGs version 4 (DxCG 1999). The appendix also contains DCGs that were combined due to low event counts or low patient counts. Age and gender were added to DCGs as well.

In addition to the DCGs, a single continuous measure called the relative risk score (RRS) is assigned by the DCG software. The RRS for an individual is calculated based on the individual's total predicted cost relative to the population mean (DxCG 1999). Individuals with a higher RRS have a higher predicted cost, based on diagnoses.

Self-Reported Health Status

The health self-report data used for this study were generated as part of two projects: the Veterans' Health Study (Kazis et al. 1998, 1999) and the Ambulatory Care Quality Improvement (ACQUIP) project (Fan et al. 2002). The instrument was a modification of the Medical Outcomes Study (MOS) SF-36, designed to be more appropriate for veterans. The modified instrument, known as SF-36V, differs from the MOS SF-36 in that two of the items, role limitations due to physical and emotional problems, are dichotomized (i.e., "yes"/"no") in the MOS SF-36 whereas five-point ordinal scales ("no, none of the time" to "yes, all of the time") are used for the SF-36V (Kazis et al. 1999).

The individual items in the SF-36V (veterans' SF-36) scores are combined into eight summary scores: physical functioning, role limitations due to physical problems, bodily pain, general health perceptions, energy/vitality, social functioning, role limitations due to emotional problems, and mental health. These eight scores are further processed into two summary measures, the Physical Component Score (PCS) and the Mental Component Score (MCS) (Ware, Kosonski, and Keller 1994). Kazis et al. (1999) found that 90 percent of the reliable variation in the eight scores in a population of VA beneficiaries was explained by PCS and MCS. In both the eight scores and the two summary measures, higher scores indicate better perceived health. Low PCS scores indicate bodily pain and limitations in daily activities; higher scores indicate the absence of disabilities or physical limitations. Low MCS scores indicate psychological distress and social and role limitations due to emotional problems; high scores indicate the absence of these conditions. The mean and standard deviation for the general U.S. population are 50 and 10, respectively (Ware, Kosonski, and Keller 1994). The mean and standard deviation for PCS in the Veterans Health study are 37.13 and 11.85, respectively, while the mean and standard deviation for MCS are 47.81 and 12.23, respectively (Kazis et al. 1999). An editorial by Ware discusses how PCS and MCS add value to diagnosis information (Ware 2000). The changes to the original MOS SF-36 incorporated in SF-36V described above were found to add precision and discriminant validity to PCS and MCS and the eight scales (Kazis 2000).

Outcomes

The diagnosis-based variables (ACGs and DCGs) were calculated using all inpatient and outpatient diagnoses that each patient had in FY 1998 (October 1, 1997–September 30, 1998). The 45.2 percent of patients in our study who completed and returned SF-36V forms also did so during FY 1998. To

investigate differences in predictive ability over time, we evaluated all-cause death for two subpopulations during two time periods. For the patients ($N = 77,473$) who were alive on October 1, 1998, we evaluated death between October 1, 1998 and September 30, 1999 (1-year mortality). For those patients who were alive on October 1, 1999 ($N = 74,854$), we evaluated death between October 1, 1999 and September 30, 2003 (2–5-year mortality). Restricting 1-year mortality to those patients who survived the baseline year (FY 1998) eliminates some patients who were too ill to complete the forms. Five-year mortality was restricted to patients who had survived the first year after the baseline year (FY 1999) in order to provide some assurance that the results for long-term mortality were not affected by short-term deaths. Death was determined using the VA National Patient Care Database and the VA Beneficiary Identification and Record Location Subsystem (BIRLS) death file. Studies have shown that the combination of these two source files has a high sensitivity for death in the VA patient population (Cowper et al. 2002).

Statistical Analysis

The basic tool was binary logistic regression. We evaluated numerous logistic models with different combinations of the variables as covariates. Seven models have been selected for presentation for the SF-36V responders. As age and gender are known to have a relatively small but significant effect on mortality (Federal Interagency Forum on Aging-Related Statistics [as of 2000] 2000) model 1 used these variables alone as covariates. All other models also contain age and gender and are compared with model 1 as well as to each other. Model 2 used age, gender, and ACGs as covariates. Model 3 used age, gender, and DCGs. Model 4 used age, gender, PCS, and MCS, and model 5 used age, gender, and the eight individual scales. An important goal of the study was to determine whether health self-report data contained health information in addition to diagnosis-based risk adjustment. Accordingly, model 6 used age, gender, ACGs, PCS, and MCS. Finally, model 7 had age, gender, DCGs, PCS, and MCS as independent variables. For the nonresponders, we used ACGs and DCGs in addition to age and gender for comparison. We also checked for interactions among the variables.

Model fit was assessed using the Akaike information criterion (AIC), a penalized log likelihood (Harrell 2001). The AIC is a commonly used measure of fit in logistic models that takes into account the different numbers of variables in different models. Smaller values indicate better model performance. The AIC is an effective way to compare models with different numbers of covariates on the same data with the same outcome. The AIC cannot be used

to compare models with different outcomes (e.g., 1-year and 2–5-year mortality), however. We also tested the models for goodness of fit using the Hosmer–Lemeshow test (Hosmer and Lemeshow 2000). All processing was done using *SAS* version 9.1. Models were run using SAS proc Logistic.

The predictive ability of the models was assessed by the probability of concordance, or *c*-statistic. This statistic is computed by taking all possible pairs of patients for which one died and the other survived. The *c*-statistic is the proportion of these pairs for which the individual who died had a greater probability of doing so than the one who survived. In contrast to the AIC, a larger value of the *c*-statistic indicates better predictive ability. A value of 0.5 indicates random predictions and a value of 1 indicates perfect predictions (Harrell 2001). We tested whether *c*-statistics for the different models were significantly different using an algorithm due to Hanley and McNeil (1982). As there are multiple models, all statistical tests were conducted at the 0.01 level.

We conducted an additional analysis to gain insight into the relationship between the risk adjusters and age for the male SF-36V responders. The population was divided into age quintiles and the *c*-statistics for 2–5-year mortality were compared for DCG-based RRS, PCS, and MCS as well as the three variables combined. We used the continuous variable RRS because not all of the ACGs and DCGs contained enough decedents within the quintiles. This analysis was limited to men because so few women were in the individual quintiles.

RESULTS

Patient Characteristics

Calculation using the MEDCs showed that the most common diagnostic group among the SF-36V responders was cardiovascular disorders ($N=$ 23,059; 65.3 percent) followed by musculoskeletal conditions ($N=$ 15,306; 43.3 percent) and psychosocial disorders ($N=$ 12,149; 34.4 percent). The MEDCs are not mutually exclusive.

Table 1 shows the differences between patients who responded to SF-36V and those who did not, for both subpopulations (those who survived FY 1998 and those who survived FY 1999). In each case, the responders were older but slightly less sick than the nonresponders. In the first subpopulation, the responders and nonresponders had similar mortality rates, but in the second subpopulation the responders had a higher 2–5-year mortality rate. The age difference may have had a more significant effect on long-term mortality.

Table 2 provides information about the SF-36V responders who survived through FY 1999 and the differences between those who survived for 5

Table 1:   Differences between SF-36V Responders and Nonresponders

	Responders (n = 35,202)	Nonresponders (n = 42,271)	p-Value
Patients surviving FY 1998			
Number (%) women	1,908 (5.4)	2,772 (6.6)	<.0001
Mean (SD) age	62.0 (13.3)	55.6 (15.0)	<.0001
Mean (SD) DCG relative risk	0.81 (1.15)	0.87 (1.23)	<.0001
Number (%) FY 1999 deaths	1,159 (3.3)	1,460 (3.5)	NS
	Responders (n = 34,043)	Nonresponders (n = 40,811)	p-Value
Patients surviving FY 1999			
Number (%) women	1,875 (5.5)	2,732 (6.7)	<.0001
Mean (SD) age	61.7 (13.3)	55.1 (14.8)	<.0001
Mean (SD) DCG relative risk	0.77 (1.08)	0.82 (1.15)	<.0001
Number (%) FY 2000–2003 deaths	4,928 (14.5)	4,907 (12.0)	<.0001

NS, not significant; FY, fiscal year; DCG, Diagnostic Cost Group.

years and those who did not. The patients who died were approximately 10 years older than those who survived. A much lower percentage of the patients who died were women than men. The mean RRS is included as an indication of the illness burden as determined by the DCG methodology. As expected, the survivors have lower RRS values, indicating they were less severely ill. The difference in PCSs between survivors and decedents was more than twice the difference in MCSs. This effect is also indicated by the fact that among the eight individual scales, the greatest differences are in the physical scales.

The Hosmer–Lemeshow goodness-of-fit test showed no lack of fit for any of these models at the 0.01 level. Table 3 shows a comparison of c-statistics and AIC for the different models for 1-year and 2–5-year mortality. The c-statistic for the age–sex only model was found to be significantly lower than all others for both outcomes using the Hanley–MacNeil algorithm. The DCGs performed better than either ACGs or SF-36V measures, but these differences were not significant for either outcome. Adding PCS and MCS to ACGs and DCGs increased the c-statistic but the only significant increase was for ACGs for 2–5-year mortality. For 2–5-year mortality, DCGs combined with PCS and MCS showed significantly better predictive power than PCS and MCS alone. The relationship among the AIC values was similar to that among the c-statistics.

Comparing the c-statistics for 1- and 2–5-year mortality shows that the predictive ability of the DCGs tended to decrease over time whereas the predictive ability of health self report combined with age increased slightly. This makes sense, as most of the health care costs (diagnoses) are accumulated

Table 2:   Patient Characteristics (SF-36 Responders Who Survived FY 1998 and FY 1999)

Demographic Characteristics	Overall Group* (n = 34,043)	2–5 Year Survivors* (n = 29,115)	2–5 Year Decedents* (n = 4,298)	p-Value
Mean (SD) age, years	61.7 (13.3)	60.2 (13.2)	70.3 (10.0)	<.0001
% Women	5.5	6.0	2.5	<.0001
Mean (SD) DCG relative risk	0.77 (1.08)	0.68 (0.94)	1.32 (1.57)	<.0001
Mean (SD) SF-36 physical component score (PCS)	33.7 (11.3)	34.5 (11.3)	28.9 (9.9)	<.0001
Mean (SD) SF-36 mental component score (MCS)	44.7 (13.9)	45.0 (14.0)	43.2 (13.1)	<.0001
Mean (SD) SF-36 body pain (BP)	44.7 (25.3)	45.4 (25.3)	40.8 (25.1)	<.0001
Mean (SD) SF-36 general health perception (GH)	45.9 (23.9)	47.2 (23.9)	37.8 (22.2)	<.0001
Mean (SD) SF-36 mental health (MH)	64.4 (23.5)	64.7 (23.6)	63.0 (22.8)	<.0001
Mean (SD) SF-36 physical functioning (PF)	50.4 (28.6)	52.8 (28.3)	36.3 (26.5)	<.0001
Mean (SD) SF-36 role emotional (RE)	50.4 (48.8)	52.7 (48.7)	37.1 (47.0)	<.0001
Mean (SD) SF-36 role physical (RP)	28.7 (38.2)	31.1 (38.9)	14.3 (29.7)	<.0001
Mean (SD) SF-36 social functioning (SF)	57.6 (30.6)	58.8 (30.5)	50.1 (30.1)	<.0001
Mean (SD) SF-36 vitality (VT)	40.6 (23.8)	41.6 (23.9)	34.2 (22.0)	<.0001

*The overall group includes all SF-36V responders. The 5-year survivors and decedents include only the responders who survived FY 1998 and FY 1999.
SF-36, 36-Item Short Form Health Survey; FY, fiscal year; DCG, Diagnostic Cost Group.

in the last 6 months of life (McCall 1984; Gaumer and Stavins 1992; Lubitz and Riley 1993). Therefore, relatively more of the diagnostic information needed to predict the 2–5-year mortality outcome is missing compared with the 1-year mortality outcome. There is a slight increase in predictive ability of ACGs from 1-year mortality to 2–5 year mortality, but this is probably due to the fact that fewer groups needed to be combined for longer-term mortality because there were more events (see Appendix). Table 4 shows the modeling results for nonresponders. Again, the ACG and DCG models added significant predictive ability to age and sex alone for both outcomes, but there was no significant difference between ACGs and DCGs for either outcome.

Table 5 shows the *c*-statistics for DCG-based RRS, PCS, and MCS, and the improvement provided by adding PCS and MCS to RRS, for the age quintiles, using 2–5-year mortality as outcome. All models show a gradual decrease in predictive ability with increasing age after age 68, but the RRS model decreases at a higher rate. The Hosmer–Lemeshow test showed lack of fit for the RRS

Table 3:   Model Results for Responders

Model	Independent Variables	Akaike Information Criterion (AIC): 1-Year Mortality	c-Statistic: 1-Year Mortality	Akaike Information Criterion (AIC): 2–5-Year mortality	c-Statistic: 2–5-Year Mortality
1	Age, gender	9,614	0.700	25,311	0.727
2	Age, gender, ACG	9,294	0.753	24,509	0.758
3	Age, gender, DCGs	8,992	0.778	24,056	0.771
4	Age, gender, PCS, MCS	9,231	0.754	24,413	0.759
5	Age, gender, 8 SF-36V scales	9,087	0.766	24,031	0.771
6	Age, sex, ACGs, PCS, MCS	9,035	0.782	23,936	0.775
7	Age, sex, DCGs, PCS, MCS	8,812	0.798	23,581	0.784

SF-36, 36-Item Short Form Health Survey; MCS, Mental Component Score; PCS, Physical Component Score; DCG, Diagnostic Cost Group; ACG, Adjusted Clinical Group.

models at the 0.01 level for several quantiles, but the other models showed no lack of fit. We ran models with age, sex, RRS, PCS, MCS, and the interactions of age with RRS, PCS, and MCS, for the two outcomes. The interaction of age with RRS was significant for 2–5-year mortality but not for 1-year mortality. The interaction of age with PCS and MCS was not significant.

DISCUSSION

We found that the predictive ability of the eight-scale SF-36V model was less than the diagnosis-based variables for 1-year mortality but equal to the DCG

Table 4:   Model Results for Nonresponders

Model	Independent Variables	Akaike Information Criterion (AIC): 1-Year Mortality	c-Statistic: 1-Year Mortality	Akaike Information Criterion (AIC): 2–5-Year Mortality	c-Statistic: 1–5-Year Mortality
1	Age, gender*	11,237	0.776	25,549	0.775
2	Age, gender, ACGs	10,735	0.818	24,610	0.802
3	Age, gender, DCGs	10,451	0.831	24,143	0.812

*Hosmer–Lemeshow test showed significant lack of fit for the 2–5 year model at the 0.01 level. DCG, Diagnostic Cost Group; ACG, Adjusted Clinical Group.

Table 5:   *c*-Statistics for Age Quintiles (2–5 Year Mortality; Men Only)

Independent Variables	Age < 51 (N = 6,885)	52 ≤ Age < 61 (N = 6,274)	61 ≤ Age < 68 (N = 6,115)	68 ≤ Age < 75 (N = 6,773)	Age ≥ 75 (N = 6,121)
DCG relative risk	0.726*	0.665	0.675*	0.643*	0.624*
PCS, MCS	0.656	0.645	0.654	0.640	0.622
DCG relative risk, PCS, MCS	0.742	0.701	0.702	0.675	0.651
Added value of PCS, MCS (%)	2.2	5.4	4.0	5.0	4.3

*Hosmer–Lemeschow test indicates lack of fit at the 0.01 level.

DCG, Diagnostic Cost Group; MCS, Mental Component Score; PCS, Physical Component Score.

model for 2–5-year mortality. A slight increase in the predictive ability of the health self-report variables over time contrasted with a slight decrease in the predictive ability of DCGs when both were combined with age and sex. The health self-report data provided consistently high predictive ability for both short- and long-term mortality. This suggests that survey-based measures may pick up the more permanent/chronic aspects of a patient's health whereas diagnosis-based measures contain more accurate information on more acute and/or transitory aspects of health.

The health self-report variables PCS and MCS also showed significant predictive ability, although less than ACGs or DCGs. Health self-report was found to add predictive power when added to diagnosis-based risk adjustment variables. The best performing model was age, sex, PCS, MCS, and DCGs, with *c*-statistics of 0.798 for 1-year mortality and 0.784 for 2–5-year mortality. The predictive ability of both the diagnosis-based and SF-36V-based risk adjusters gradually decreased with advancing age, after age 68.

The patients who responded to the SF-36V survey had a slightly lower mortality rate that those who did not. The relationship between the ACG and DCG models was about the same for responders and nonresponders. The DCGs generally had higher predictive power than ACGs. These results are consistent with our earlier findings assessing the predictive ability of diagnosis-based risk adjusters for clinical outcomes (Petersen et al. 2005).

Of course, neither ACGs, DCGs nor SF-36 measures were designed to predict mortality. However, it is well known that a large part of medical expenses for a patient occur as a result of end-of-life care. Therefore, knowledge of what factors predict long-term mortality can be important in resource planning (Pietz , Byrne, and Petersen 2006). Obtaining diagnosis information on patients

requires that they present for care at a VA medical facility. Current diagnosis information may not be available for patients with limited access to care. Survey information, on the other hand, can be obtained by mass mailing, as was done in this study. Although not all patients will respond, the results will provide an indication of expenses beyond what can be expected for the next several years. Our results suggest that for 1-year planning, diagnosis-based information may be more appropriate. For 5-year planning, a medical facility may want to consider obtaining SF-36 information on potential users by a mass mailing.

The results of a large-scale SF-36 survey could also be used to look for differences in perceived health across networks or geographic areas. The relationship between perceived health and mortality rates could be examined to look for patterns. Differences in the relationship could be targets for further investigation. Of course, the potential value of the information gained would have to be weighed against the cost of collecting and processing this information.

In Pietz et al. (2004), we found that SF-36V data added little to the ability of ACG-derived variables to predict total medical care cost. Taken together, our results show that predicting mortality is a much different problem than predicting short-term cost. The ACGs and DCGs were developed to predict medical care cost. Factors that influence a patient's propensity to die may not always be those that require the most costly care. For example, a patient with metastatic lung cancer may elect to have palliative care only, rather than invasive treatments.

Survey methodology seeks to estimate observables in a specific population (Rubin 1987). The results should not be considered estimates of results that might be obtained in a hypothetical larger population. Our response rate was 45.2 percent and the missing values are not missing at random (Pietz et al. 2004). However, the actual number of patients who did respond is substantial. Rather than attempting to impute the large percentage of missing values, we have presented an investigation into how the data could be used to predict mortality on the body of patients for whom the data was obtained.

Many veterans use the VA for only part of their medical care. We did not have information on the Medicare utilization of our population, for example, although we were able to assess mortality for these patients. Some of the patients' medical diagnoses may not be entered into VA databases (Byrne et al. 2006). It is possible that with more complete coding information, the added value of the SF-36V data would have been less. Also, this study used all-cause mortality. It can be assumed that some deaths were not the result of medical conditions. Finally, the VA population is older and largely male, in contrast to many other patient populations.

The relationship between diagnosis-based health information and self-reported health information will be a subject of continuing research. Determining better ways of assessing a patient's true state of health can lead to better medical care. Methods of determining a patient's state of health are still poorly understood. A physician assesses a patient's health status based on physical examination, test results, and information obtained from the patient. Our research indicates that the results of a well-validated health self-report instrument, such as the SF-36V, contain unique medical information not found in diagnosis-based risk adjustment variables, especially for long-term outcomes.

ACKNOWLEDGMENTS

REFERENCES

Ash, A. S., R. P. Ellis, G. C. Pope, J. Z. Ayanian, D. W. Bates, H. Burstin, L. I. Iezzoni, E. MacKay, and W. Yu. 2000. "Using Diagnoses to Describe Populations and Predict Costs." *Health Care Financing Review* 2: 7–28.

Byrne, M. M., M. Kuebeler, K. Pietz, and L. A. Petersen. 2006. "Effect of Using Information from Only One System for Dually Eligible Health Care Users." *Medical Care* 44(8): forthcoming.

Cowper, D. C., J. D. Kubal, C. Maynard, and D. M. Hynes. 2002. "A Primer and Comparative Review of Major US Mortality Databases." *Annals of Epidemiology* 12: 462–8.

DxCG, I. (1999). *Guide to the Diagnostic Cost Groups (DCGs) and DCG Software, Release 4.* Waltham, MA: DxCG Inc.

Fan, V. S., D. Au, P. Heagerty, R. A. Deyo, M. B. McDonell, and S. D. Fihn. 2002. "Validation of Case-Mix Measures Derived from Self-Reports of Diagnoses and Health." *Journal of Clinical Epidemiology* 55: 371–80.

Federal Interagency Forum on Aging-Related Statistics (as of 2000). 2000. *Older Americans 2000: Key Indicators of Well-Being.*

Gaumer, G. L., and J. Stavins. 1992. "Medicare Use in the Last Ninety Days of Life." *Health Services Research* 26: 725–42.

Hanley, J. A., and B. J. McNeil. 1982. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve." *Radiology* 143: 29–36.

Harrell, F. E., Jr. 2001. *Regression Modeling Strategies, with Applications to Linear Models, Logistic Regression, and Survival Analysis. 247.* New York: Springer.

Hornbrook, M. C., and M. J. Goodman. 1996. "Chronic Disease, Functional Health Status, and Demographics: A Multi-Dimensional Approach to Risk Adjustment." *Health Services Research* 31: 283–307.

Hosmer, D. W. Jr., and S. Lemeshow. 2000. *Applied Logistic Regression.* 2d Edition. New York: Wiley.

Idler, E. L., and R. J. Angel. 1990. "Self-Rated Health and Mortality in the NHANES-I Epidemiologic Follow-up Study." *American Journal of Public Health* 80: 446–52.

Idler, E. L., and Y. Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior* 38: 21–37.

Inouye, S. K., P. N. Peduzzi, J. T. Robison, J. S. Hughes, R. I. Horwitz, and J. Concato. 1998. "Importance of Functional Measures in Predicting Mortality among Older Hospitalized Patients." *Journal of the American Medical Association* 279: 1187–93.

Johns Hopkins. 2001. *Johns Hopkins ACG Case-Mix System: Documentation and Application Manual.* Baltimore: The Johns Hopkins University.

Kazis, L. E. 2000. "The Veterans SF-36 Health Status Questionnaire: Development and Application in the Veterans Health Administration." *Monitor* 5: 1–14.

Kazis, L. E., D. R. Miller, J. Clark, K. Skinner, A. Lee, W. Rogers, A. Spiro, S. Payne, G. Fincke, A. Selim, and M. Linzer. 1998. "Health-Related Quality of Life in Patients Served by the Department of Veterans Affairs: Results from the Veterans Health Study." *Archives of Internal Medicine* 158: 626–32.

Kazis, L. E., S. R. Xinhua, A. Lee, K. Skinner, W. Rogers, J. A. Clark, and D. R. Miller. 1999. "Health Status in VA Patients: Results from the Veterans Health Study." *American Journal of Medical Quality* 14: 28–38.

Lamers, L. M., and R. C. van Vliet. 1996. "Multiyear Diagnostic Information from Prior Hospitalization as a Risk-Adjuster for Capitation Payments." *Medical Care* 34: 549–61.

Lubitz, J. D., and G. F. Riley. 1993. "Trends in Medicare Payments in the Last Year of Life." *New England Journal of Medicine* 328: 1092–6.

McCall, N. 1984. "Utilization and Costs of Medicare Services by Beneficiaries in Their Last Year of Life." *Medical Care* 22: 329–42.

Petersen, L. A., K. Pietz, L. D. Woodard, and M. Byrne. 2005. "Comparison of the Predictive Validity of Diagnosis-Based Risk Adjusters for Clinical Outcomes." *Medical Care* 43: 61–67.

Pietz, K., C. M. Ashton, M. McDonell and N. P. Wray. 2004. "Predicting Future Health Care Costs in a Population of VA Beneficiaries Using Diagnosis-Based Risk Adjustment and Self-Reported Health Status." *Medical Care* 42: 1027–35.

Pietz, K., M. M. Byrne, and L. A. Petersen. 2006. "A Decision-Theoretic Approach to Identifying Future High-Cost Patients." *Medical Care* 44(9): forthcoming.

Pietz, K., K. J. O'Malley, M. Byrne, J. Souchek, N. Petersen, and C. M. Ashton. 2002. "Physician-Level Variation in Practice Patterns in the VA Healthcare System." *Health Services and Outcomes Research Methodology* 3: 95–106.

Reid, R. J., L. MacWilliam, L. Verhulst, N. Roos, and M. Atkinson. 2001. "Performance of the ACG Case-Mix System in Two Canadian Provinces." *Medical Care* 39: 86–99.

Rosen, A. K., S. Loveland, J. J. Anderson, J. A. Rothendler, C. S. Hankin, C. C. Rakovski, M. A. Moskowitz, and D. R. Berlowitz. 2001. "Evaluating Diagnosis-Based Case-Mix Measures: How Well Do They Apply to the VA Population?" *Medical Care* 39: 692–704.

Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys.* Hoboken, NJ: John Wiley & Sons.

Schoenman, J. A., K. J. Hayes, and C. M. Cheng. 2001. "Medicare Physician Payment Changes: Impact on Physicians and Beneficiaries." *Health Affairs (Millwood.)* 20: 263–73.

Starfield, B., J. Weiner, L. Mumford, and D. Steinwachs. 1991. "Ambulatory Care Groups: A Categorization of Diagnoses for Research and Management." *Health Services Research* 26: 53–74.

Ware, J. E. Jr. 2000. "Editorial: Using Generic Measures of Functional Health and Well-Being to Increase Understanding of Disease Burden." *Spine* 25(12): 1467.

Ware, J. E. Jr., M. Kosonski, and S. D. Keller. 1994. *SF-36 Physical and Mental Health Summary Scales: A User's Manual.* Boston: The Health Institute, New England Medical Center.

SUPPLEMENTARY MATERIAL

The following supplementary material for this article is available online:
    APPENDIX: ACGs and DCGs Combined for the Different Models.