Using Information on Clinical Conditions to Predict High-Cost Patients

John A. Fleishman and Joel W. Cohen

Objective. To compare the ability of different models to predict prospectively whether someone will incur high medical expenditures.

Data Source. Using nationally representative data from the Medical Expenditure Panel Survey (MEPS), prediction models were developed using cohorts initiated in 1996–1999 ($N = 52{,}918$), and validated using cohorts initiated in 2000–2003 ($N = 61{,}155$).

Study Design. We estimated logistic regression models to predict being in the upper expenditure decile in Year 2 of a cohort, based on data from Year 1. We compared a summary risk score based on diagnostic cost group (DCG) prospective risk scores to a count of chronic conditions and indicators for 10 specific high-prevalence chronic conditions. We examined whether self-rated health and functional limitations enhanced prediction, controlling for clinical conditions. Models were evaluated using the Bayesian information criterion and the *c*-statistic.

Principal Findings. Medical condition information substantially improved prediction of high expenditures beyond gender and age, with the DCG risk score providing the greatest improvement in prediction. The count of chronic conditions, self-reported health status, and functional limitations were significantly associated with future high expenditures, controlling for DCG score. A model including these variables had good discrimination ($c = 0.836$).

Conclusions. The number of chronic conditions merits consideration in future efforts to develop expenditure prediction models. While significant, self-rated health and indicators of functioning improved prediction only slightly.

Key Words. Health care expenditures, prediction models, DCG models, chronic conditions

Predicting health care expenditures is important for health care research, policy, and practice. In health care delivery settings, interest often focuses on identifying potential high-cost patients, who could be enrolled in case-management or disease-management programs. Moreover, when used as a means of risk adjustment to improve comparability of different groups, predictions of high-cost cases are useful for evaluating clinical interventions or the effects of policy changes.

Expenditure prediction models typically incorporate information on clinical conditions based on data from medical records or claims databases, in addition to demographic information (gender and age). The diagnostic cost group (DCG) system is one system for predicting health care expenditures based on ICD-9 codes obtained from insurance claims databases. "DCG models use age, sex, and diagnoses . . . to infer which medical problems are present for each individual and their likely effect on health care costs for a population" (Ash et al. 2000, p. 7).

Although software to estimate DCG models is available commercially, some potential users may not have the resources or expertise to run such software and may instead prefer simpler algorithms. One possible alternative algorithm is to focus on the presence of certain conditions that are prevalent and chronic; indicators of such conditions may provide sufficient prediction of expenditures (Farley, Harley, and Devine 2006; Baser, Palmer, and Stephenson 2008; Charlson et al. 2008). Another substitute could be a count of chronic conditions. Several studies have found that simple counts of diagnoses or conditions are associated with mortality or other outcomes (Melfi et al. 1995; Farley, Harley, and Devine 2006). To what extent does using such simpler algorithms result in a loss of predictive power, compared with DCG risk scores?

A second issue pertains to the utility of incorporating self-reported health status in risk-adjustment models. Several studies have examined the extent to which patient-reported health status information improves the performance of expenditure prediction models (Hornbrook and Goodman 1995; Pope et al. 1998; Pietz et al. 2004; Maciejewski et al. 2005). Adding scales from the SF-36 health status measure (Ware and Sherbourne 1992) to predictive models increased the adjusted R^2, but by small amounts (Hornbrook and Goodman 1995; Pietz et al. 2004). Generalization from these studies is limited by their use of nonrepresentative samples, such as VA patients, Medicare beneficiaries, or working-age members of a single HMO; the VA samples included relatively few women, and the HMO sample few elderly patients.

Fleishman et al. (2006) used a nationally representative sample to estimate predictive models that included the SF-12 health status measure

Address correspondence to John A. Fleishman, Ph.D., Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850; e-mail: john.fleishman@ahrq.hhs.gov. Joel W. Cohen, Ph.D., is with the Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality, Rockville, MD.

(Ware, Kosinski, and Keller 1996). SF-12 summary scores were significantly associated with expenditures, controlling for demographic characteristics and specific chronic conditions. A model including one general health status question instead of the full SF-12 performed nearly as well. However, these analyses included indicators for only six highly prevalent conditions. It is unclear whether the same results would be obtained if richer information on clinical conditions had been included in the predictive models.

The literature on risk adjustment is large (Iezzoni 2003). However, relatively few studies systematically compare the performance of different risk adjusters when predicting high-cost cases. Some studies compare the DCG risk score to the Charlson comorbidity index or the SF-36 but use mortality or some other clinical indicator as a criterion, not expenditures (Ash et al. 2003; Baldwin et al. 2006; Fan et al. 2006). Relative performance may differ when predicting different outcomes (Perkins et al. 2004). Other studies do examine expenditures as an outcome, but they rely on data from selected samples, such as VA patients with diabetes (Maciejewski, Liu, and Fihn 2009), HMO enrollees with hypertension (Farley, Harley, and Devine 2006), or patients with private insurance treated for migraine (Baser, Palmer, and Stephenson 2008).

The current study uses nationally representative data from the Medical Expenditure Panel Survey (MEPS) to compare expenditure prediction models. We examine three approaches to incorporating clinical condition information: the prospective risk score generated by the DCG algorithm; indicators of specific prevalent chronic conditions; and a count of the number of chronic conditions. We focus on chronic conditions because their associated expenditures may be more predictable than those for acute conditions, and because a large proportion of total aggregate health care spending is associated with care of chronic diseases (Anderson and Horvath 2004; Machlin, Cohen, and Beauregard 2008). In addition, extending the results of Fleishman et al. (2006), we examine the additional predictive power contributed by a global health status question and two measures of functioning.

We focus on individuals who have high expenditures. Specifically, we examine the extent to which predictive models can identify individuals with "high future costs," that is, costs in the upper 10 percent of the subsequent year's expenditure distribution. While various thresholds for defining "high cost" have been used, the upper 10 percent is common (e.g., Farley, Harley, and Devine 2006).

METHODS

MEPS

The MEPS is a nationally representative survey of health care utilization and expenditures for the U.S. civilian noninstitutionalized population, sponsored by the Agency for Healthcare Research and Quality. MEPS is a panel survey, with an overlapping cohort design. A new cohort (panel) is initiated each year and provides information for a 2-year reference period. MEPS conducts five in-person interviews (Rounds) with one or more persons per household, who report on health care utilization, expenditures, insurance coverage, and medical conditions for each household member.

MEPS collects expenditure data in each Round of the survey. Expenditures in MEPS refer to direct payments for care, including out-of-pocket payments and payments from private insurance, Medicare, Medicaid, and other sources. Expenditures include prescription medications, hospital inpatient stays, home health visits, medical supplies (including vision and hearing aids), and visits to office-based providers, hospital clinics, emergency rooms, and dental providers. Payments for over-the-counter drugs are not included. MEPS expenditure data also reflect the utilization of health services by persons in capitated programs; studies based on claims data typically do not include this subgroup.

In addition to the household interview, MEPS includes a medical provider component (MPC), a follow-back survey that collects expenditure data from a sample of medical providers and pharmacies used by survey participants. MPC data capture expenditures for hospital inpatient stays, emergency room and outpatient hospital visits, office-based physician visits, home health care, and prescribed medicines. MPC expenditure data are considered to be more accurate than information reported by households and were given priority in expenditure estimation. For nonphysician visits, dental and vision services, other medical equipment and services, and home health care not provided by an agency, information on expenditures is obtained solely from household respondents (Cohen et al. 1996; Machlin and Taylor 2000).

The MEPS interview collects detailed data on sociodemographic characteristics and medical conditions of each sampled person. A household informant answers these questions for each household member. Age at January 1 of the second year of each panel was grouped into 15 categories that correspond to the age categories used in DCG models. To measure perceived health, respondents were asked: "In general, compared with other people (PERSON'S) age, would you say that (PERSON'S) health is excellent, very

good, good, fair, or poor?" Binary indicators represented female gender and, to capture possible nonlinear relationships, categories of age, and perceived health. To measure limitations in instrumental activities of daily living (IADLs), household members who received "help or supervision with using the telephone, paying bills, taking medication, preparing light meals, doing laundry, or going shopping" were identified. To measure functional limitations, household members who had difficulty "walking, climbing stairs, grasping objects, reaching overhead, lifting, bending or stooping, or standing for long periods of time" were identified. Analyses use responses obtained in Round 1 of each panel, predominantly in the first 6 months of the 2-year observation period.

A summary indicator of annual insurance coverage classified respondents based on the type of coverage they held for the greatest number of months during the first year of their observation period. Respondents with multiple types of coverage for the same number of months were classified based on the hierarchy: Medicare, private, Medicaid, or uninsured.

Medical Conditions

Information on specific medical conditions was obtained in the MEPS interview by asking which "health problems" had "bothered" each household member during the observation period. In addition, respondents reported the reasons for each medical event (outpatient visit, prescription medicine purchase, inpatient episode, ER visit, and home health visit). Reported medical conditions and procedures were recorded as verbatim text and then coded by professional coders to fully specified ICD-9-CM codes (AHRQ 2006). (For confidentiality, the MEPS public use files contain only three-digit ICD-9 codes, but the current analyses use five-digit codes.)

As the most comprehensive indicator of clinical status, we used the relative risk score (RRS) based on prospective DCG models. DCG models group over 15,000 ICD-9 codes into a few hundred hierarchically organized condition categories, which serve as the basis for estimating expected costs (Ash et al. 2000). DCG prospective models use hierarchical condition categories, age, and gender, to predict next year's expenditures from this year's data. These predictions are reported as RRSs. DCG estimation models have been developed from large claims databases, resulting in distinct predictive models for patients with commercial, Medicare, or Medicaid coverage. For comparability across respondents, we applied the DCG algorithm for commercial insurance, regardless of the actual insurance status of the respondent.

We used DxCG RiskSmart Software (version 2.2) to derive RRSs (DxCG Inc. 2007).

We classified ICD-9 codes as chronic or nonchronic; a chronic condition is one that lasts 12 months or longer and either (1) limits self-care or independent living or (2) results in a need for ongoing medical intervention (Hwang et al. 2001; Friedman et al. 2006). Software for performing this classification is publicly available (AHRQ 2009). For each person, we counted the number of chronic conditions reported in the first year of a panel, from one to nine or more conditions; we created binary indicators for each number of chronic conditions. Additional indicators identified respondents reporting no conditions and those reporting only acute conditions.

We used Clinical Classification Software to aggregate clinically similar ICD-9 codes (including V-codes) into 263 mutually exclusive, clinical condition categories (Elixhauser et al. 2000). To ensure adequate sample size, categories with clinically similar conditions were further collapsed (Cohen and Krauss, 2003). We selected the 10 most prevalent categories of chronic conditions as "key" conditions. Categories that included mostly acute or short-term conditions (e.g., influenza, injuries) were excluded. Key conditions included hypertension, diabetes, asthma, cancer (excluding skin cancers), mood disorders (depression), anxiety disorders, heart disease (including AMI, CHF, but excluding nonspecific chest pain), nontraumatic joint disorders, and diseases of arteries, veins, and lymphatics (Appendix Table SA2).

Analyses

Analyses used data from MEPS Panels 1 (begun in 1996) through 8 (begun in 2003). Each panel is independent of the others. Respondents who provided data for only 1 year of a panel were excluded from analyses. Of a total of 134,503 respondents in Panels 1–8, 120,327 provided data for both years. Of this group, 116,727 had a positive longitudinal analysis weight; nonpositive weights were primarily due to some period of ineligibility (i.e., not in the civilian noninstitutionalized population). After deleting respondents with missing data, the final analytic sample included 114,073 observations with a positive analytic weight. No observation was excluded on an a priori basis due to a particular characteristic, such as age or pregnancy status.

Within each panel, we derived a binary indicator for being in the top 10 percent of expenditures in Year 2. Defining indicators separately for each panel implicitly adjusts for inflation. As a sensitivity analysis, we repeated analyses using the upper 5 percent as a cutoff for high-cost cases. Using more

extreme cutoffs (e.g., 1 percent) would have resulted in samples that were too small to provide stable estimates in large regression models.

Using data from MEPS Panels 1–4 (unweighted $N = 52,918$), we estimated logistic regression models predicting whether respondents were in the top 10 percent of the Year 2 expenditure distribution, using information from Year 1. A baseline model included gender, age categories, insurance, and panel indicators. To examine different specifications of clinical conditions, subsequent models added (1) indicators for 10 specific "key" chronic conditions, (2) indicators for the number of chronic conditions (key and not key), and (3) the DCG risk score, which was categorized on the basis of percentiles as 0–25, 26–50, 51–75, 76–90, and 91–100 percent in order to capture nonlinearity in the association with expenditures. Another model added self-rated health, IADL help, and functional limitations to the baseline model. Models with combinations of two of the three sets of condition-related variables were examined next. To gauge their additional contribution, the functional status and perceived health variables were added to the best of these models. All analyses incorporated MEPS analytic weights.

We evaluated different models by comparing differences in the Bayesian information criterion (BIC; Raftery 1995), with lower values indicating a better-fitting model. In addition, we also examined the Hosmer–Lemeshow goodness-of-fit statistic adapted for data from complex sample surveys (Archer and Lemeshow 2006); the pseudo-R^2 statistic (which compares the likelihoods of the current model and the model with only an intercept); the Pearson correlation between predicted probabilities and the dependent variable, which has been recommended as an indicator of goodness-of-fit (Ash and Shwartz 1999; Zheng and Agresti 2000); and the *c*-statistic, which assesses a model's ability to discriminate high from nonhigh expenditure cases and equals the area under the receiver operating curve (Hosmer and Lemeshow 2000). Unlike the other goodness-of-fit criteria, which must improve as more terms are included in a model, the BIC penalizes for the number of parameters and can show worse fit if unnecessary variables are included.

To validate the performance of the regression models, we applied the estimated coefficients to data from MEPS Panels 5–8 (unweighted $N = 61,155$) to predict being in the top 10 percent of expenditures in Year 2 of these panels, based on data from Year 1. We examined *c*-statistics for different models. For the best-performing models, we compared predictions with actual Year 2 expenditures and examined sensitivity and positive predictive value, using a predicted probability > 0.1 as a criterion.

RESULTS

Descriptive Analyses

Panels 1–4 and 5–8 had similar demographic and clinical characteristics, although inducted in different years (Table SA3). Over all Panels, 14 percent reported no clinical conditions, 39 percent only acute conditions, and 47 percent identified one or more chronic conditions in Year 1. Among key conditions, hypertension was most common (12 percent), and 4 percent or more reported each of depression, diabetes, or cancer. Nine percent had some functional limitation, but only 2 percent reported receiving help for IADL tasks. Most respondents were eligible for the full observation period in both years; across all panels, the proportion with fewer than 365 days of eligibility was 0.8 percent in Year 1 and 3.6 percent in Year 2 (results not shown).

The mean inflation-adjusted expenditure was U.S.$2,662 in Year 1 and U.S.$2,770 in Year 2. In Year 1, 14 percent had zero expenditures; 17 percent had no expenditures in Year 2. The maximum expenditure was U.S.$550,273 in Year 1 and U.S.$1,005,935 in Year 2.

Table 1a displays the unadjusted association between number of chronic conditions and subsequent expenditures for Panels 1–4. The probability of being in the top 10 percent and mean expenditures in Year 2 both rose with increasing numbers of chronic conditions. Very few people (1.6 percent) with no chronic conditions in Year 1 were in the top 10 percent of the next year's expenditure distribution. People reporting no conditions in Year 1 averaged U.S.$579 in expenditures in Year 2, presumably due to incident conditions or routine medical care. More than 50 percent of people with six or more chronic conditions had high future costs. However, <2 percent of the overall population had so many chronic conditions. The prospective DCG risk score was also strongly related both to the probability of being in the top 10 percent and to mean expenditures in the subsequent year (Table 1b).

Multivariate Analyses: Panels 1–4

All logistic regression models included age, gender, insurance, and indicators for each MEPS panel. This baseline model did not fit, using the Hosmer–Lemeshow test. Except for a model including key conditions (Table 2, Model 4), subsequent models each had nonsignificant Hosmer–Lemeshow goodness-of-fit statistics, indicating satisfactory overall fit. Adding the number of chronic conditions (Model 2) or the DCG score categories (Model 3) each markedly improved the model's performance, as shown by several goodness-of-fit indices in Table 2. Adding key condition indicators (Model 4) or perceived

Table 1a: Expenditures in Year 2 as a Function of Chronic Conditions in Year 1, MEPS Panels 1–4

Number of Chronic Conditions	Unweighted N	Weighted Proportion	Proportion in Top 10% Next Year	Mean Expenditures Next Year
No conditions	7,784	0.138	0.016 (0.002)	579 (59)
Acute conditions only	20,876	0.394	0.045 (0.002)	1,178 (37)
1	12,204	0.237	0.084 (0.003)	2,216 (83)
2	5,833	0.114	0.165 (0.006)	4,056 (149)
3	2,972	0.057	0.269 (0.009)	6,377 (342)
4	1,576	0.029	0.361 (0.014)	9,450 (924)
5	864	0.016	0.485 (0.021)	10,925 (859)
6	390	0.007	0.526 (0.031)	11,518 (1,030)
7	218	0.004	0.642 (0.034)	16,061 (1,601)
8	106	0.002	0.701 (0.050)	11,259 (1,068)
9+	95	0.001	0.811 (0.054)	23,824 (4,176)

MEPS, Medical Expenditure Panel Survey.

Table 1b: Expenditures in Year 2 as a Function of DCG Risk Score Category in Year 1, MEPS Panels 1–4

DCG Prospective Risk Score (Private Insurance Model) (%)	Unweighted N	Proportion in Top 10% Next Year	Mean Expenditures Next Year	Mean DCG Risk Score
Lowest 25	12,696	0.015 (0.001)	524 (28.94)	0.18
26–50	13,705	0.031 (0.002)	989 (48.63)	0.42
51–75	13,523	0.074 (0.003)	2,028 (59.37)	0.91
76–90	7,728	0.199 (0.005)	4,540 (138.05)	1.71
Top 10	5,266	0.425 (0.009)	10,028 (445.10)	3.37

DCG, diagnostic cost group; MEPS, Medical Expenditure Panel Survey.

Table 2:   Goodness-of-Fit Indicators for Several Logistic Models Predicting High (Upper 10%) Expenditures in Year 2, MEPS Panels 1–4

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Log-likelihood	−15,297	−14,093	−13,804	−14,556	−14,625	−13,596	−14,017	−13,694	−13,483	−13,845
Hosmer–Lemeshow test	4.29*	0.62	0.55	4.46*	1.62	0.80	0.65	0.29	0.74	0.58
Pseudo-R^2	0.118	0.188	0.204	0.161	0.157	0.216	0.192	0.211	0.223	0.202
Correlation	0.310	0.414	0.419	0.385	0.380	0.441	0.420	0.433	0.452	0.436
c-statistic	0.765	0.813	0.827	0.796	0.793	0.833	0.815	0.830	0.836	0.821
BIC	−3,865	−6,165	−6,809	−5,238	−5,144	−7,115	−6,208	−6,920	−7,276	−6,487
Change in BIC	—	−2,300	−2,944	−1,373	−1,279	−306	+601	−111	−161	+628
Comparison model	—	1	1	1	1	3	3	3	6	6

*$p < 0.05$.

Model 1 (baseline): age, gender, insurance, panel.
Model 2: Model 1+number of chronic conditions.
Model 3: Model 1+DCG categories.
Model 4: Model 1+key condition indicators.
Model 5: Model 1+perceived health and functioning.
Model 6: Model 1+number of chronic conditions+DCG categories.
Model 7: Model 1+number of chronic conditions+key condition indicators.
Model 8: Model 1+DCG categories+key condition indicators.
Model 9: Model 1+number of chronic conditions+DCG categories+perceived health and functioning.
Model 10: Model 1+number of chronic conditions+key condition indicators+perceived health and functioning.
BIC, Bayesian information criterion; DCG, diagnostic cost group; MEPS, Medical Expenditure Panel Survey.

health and functioning (Model 5) to the baseline model also improved performance, but not as much as the prior two variables. Among Models 2–5, the DCG model had the lowest BIC value (−6,809) and highest c-statistic (0.827).

Combining Models 2 and 3 (Model 6) resulted in further improvements in fit, compared with Model 3. In contrast, Model 8 (DCG scores and key conditions) resulted in less improvement. Model 7 resulted in worsening of the BIC, relative to Model 3.

Adding functioning variables to Model 6 resulted in slightly improved fit (Model 9). Although the change in the c-statistic was minimal, the difference in c-statistics between Models 6 and 9 was significant based on a test of areas under correlated ROC curves ($\chi^2 = 51.14$, df = 1). The final model thus included the baseline predictors, the DCG score categories, the functioning variables, and the number of chronic conditions.

For Model 9, all the coefficients for the number of chronic conditions and for the DCG score categories were significant ($p < .001$; Table SA4). The odds of being in the top 10 percent of expenditures rose monotonically as the number of chronic conditions rose and as the DCG score category rose. The increase was nonlinear, with notable jumps in the coefficients for seven or more chronic conditions and for the highest decile of DCG scores. In addition, respondents with ratings of "good," "fair," and "poor" health were significantly more likely to have high expenditures than those reporting "excellent" health (respective adjusted odds ratios [AOR] = 1.23, 1.37, 1.90). IADL and functional limitations were each significantly associated with high future expenditures (respective AORs = 1.59, 1.23), controlling for self-rated health and clinical variables.

As a final comparison, Model 10 included the number of chronic conditions, the key condition indicators, and the functioning variables. This model represents the level of prediction attainable without the complexity of DCG scores. This model fit worse than Models 6 and 9, with drops in the BIC, c-statistic, and other fit indicators.

Sensitivity Analyses. Relative model performance using the upper 5 percent as a cutoff for high-cost cases was consistent with that obtained using the upper 10 percent (Table SA5).

It is possible that prediction models might perform differently in subsets of the population. In particular, chronic diseases are more prevalent among the elderly, and the predictive power of key chronic condition indicators might improve in an older sample. We repeated the analyses

above, among nonelderly adults (ages 18–64) and among those aged 65 or older (Tables SA6 and SA7). In both subgroups, Model 9 had the best goodness-of-fit indicators (c-statistic = 0.791 for nonelderly adults, and 0.748 for the elderly).

Predicting Expenditures: Panels 5–8

As external validation, we applied the parameter estimates from Models 1–10, based on Panels 1–4, to predict "future high cost" cases for MEPS Panels 5–8. Results for Panels 5–8 were similar to those for Panels 1–4 (Table 3). The c-statistics for Models 1–10 were higher in magnitude than for Panels 1–4, but their relative magnitudes were the same, with the highest value for Model 9.

The mean of the predicted probabilities in the validation sample was slightly > 0.1, ranging from 0.103 to 0.114, depending on the model. Using a cutoff of a predicted probability of 0.1 to define a likely future high-cost case, sensitivities varied between 0.72 and 0.78, and positive predicted values between 0.22 and 0.29. Model 9 was not superior to Model 6 using these criteria.

To further evaluate Models 6 and 9, we compared the number of cases in each decile of predicted probability who were subsequently in the top 10 percent with the predicted number of top 10 percent cases, based on the sum of predicted probabilities in each decile. Both models tended to over-predict for deciles 1–9, but Model 9 under-predicted for decile 10. For Model 6, the predicted numbers of high-cost cases were 1,434 and 2,929 in deciles 9 and 10, while the actual numbers of high-cost cases were 1,395 and 2,929 in these deciles, respectively. For Model 9, the predicted numbers of high-cost cases were 1,389 and 2,892 in deciles 9 and 10, while the actual numbers of high-cost cases were 1,413 and 2,921 in these deciles, respectively.

We also compared the proportion of high-cost dollars that each model predicted, by forming the ratio of total expenditures for people in the top decile of predicted probability to the total expenditures for all actual high-cost cases (e.g., Meenan et al. 2003). Models 6 and 9 had the highest proportions, with Model 9 being slightly higher (0.622 versus 0.629).

DISCUSSION

Results of this study extend prior comparisons of risk-adjustment models to nationally representative data, rather than data from a specific provider or

Table 3: Validation of Logistic Models Predicting High (Upper 10%) Expenditures in Year 2, MEPS Panels 5–8

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Correlation	0.347	0.446	0.454	0.423	0.405	0.473	0.454	0.466	0.480	0.464
c-statistic	0.791	0.835	0.849	0.824	0.816	0.854	0.837	0.852	0.857	0.842
Sensitivity ($p>.1$)	0.72	0.75	0.76	0.74	0.74	0.78	0.76	0.77	0.78	0.76
Positive predictive value	0.22	0.26	0.29	0.26	0.24	0.29	0.26	0.29	0.29	0.26
BIC	−5,813	−8,748	−9,702	−7,911	−7,296	−10,051	−9,895	−8,884	−10,187	−9,119

Model 1 (baseline): age, gender, insurance, panel.
Model 2: Model 1+number of chronic conditions.
Model 3: Model 1+DCG categories.
Model 4: Model 1+key condition indicators.
Model 5: Model 1+perceived health and functioning.
Model 6: Model 1+number of chronic conditions+DCG categories.
Model 7: Model 1+number of chronic conditions+key condition indicators.
Model 8: Model 1+DCG categories+key condition indicators.
Model 9: Model 1+number of chronic conditions+DCG categories+perceived health and functioning.
Model 10: Model 1+number of chronic conditions+key condition indicators+perceived health and functioning.
BIC, Bayesian information criterion; DCG, diagnostic cost group; MEPS, Medical Expenditure Panel Survey.

provider system, spanning all age groups. Incorporating information on medical conditions substantially improved prediction of high expenditures, compared with using only data on gender and age. The DCG risk score provided the greatest improvement in prediction among the sets of variables considered. Unlike some other studies, we also examined the extent to which combinations of potential risk adjustors improved prediction (rather than only comparing each risk adjustor by itself). Counts of the number of chronic conditions and indicators of poor health status and functional limitations were significant, even after controlling for the DCG risk score.

These results were robust. Models developed using four large independent samples successfully predicted high expenditures using validation data from a different set of four large independent samples. The c-statistic for Model 9 was 0.836, which represents excellent discrimination. This value compares favorably with results of a study that predicted high-cost cases using data from several HMOs; for high-cost cases defined as the upper 1 percent of expenditures, c-statistics ranged from 0.81 to 0.85, depending on the specific risk-adjustment model (Meenan et al. 2003). DeSalvo et al. (2009) obtained a comparable c-statistic (0.85) for predicting the upper 10 percent of total expenditures with a model including DCG scores and self-rated health.

We expected that the DCG score would overlap with indicators of the number of chronic conditions, and that these indicators would not be significant if the DCG scores were controlled. However, the number of chronic conditions also significantly predicted high-cost cases, controlling for DCG score category. A simple count of the number of chronic conditions is useful in predicting future high-cost cases and may be sensitive to information on severity of conditions that the DCG score is not picking up.

In contrast, separate indicators of specific prevalent diagnoses were significant predictors of high-costs when the model excluded the DCG score, but the set of key condition indicators provided less improvement than the count of chronic conditions when the DCG score was controlled. The redundancy between the DCG score and specific condition indicators is great, and thus this result is not surprising. The model including the relatively straightforward count of chronic conditions and key condition indicators did not perform as well as models including the DCG scores, suggesting that the improved predictive power may compensate for the cost and complexity of using the DCG scores.

Prior studies suggest that measures of health status, such as the SF-36 or self-rated health, significantly contribute to prediction of expenditures, but the improvement in prediction is slight if condition-based indices are controlled

(Hornbrook and Goodman 1995; Pope et al. 1998; Pietz et al. 2004; Maciejewski et al. 2005; Fleishman et al. 2006; DeSalvo et al. 2009). The current results are consistent with prior studies; although statistically significant, the health status and functioning variables, when added to a model that included DCG score and number of chronic conditions, increased the *c*-statistic from 0.833 to 0.836 in the derivation sample and from 0.854 to 0.857 in the validation sample. Adding the functioning variables did not notably improve sensitivity or positive predictive value in the validation sample.

The utility of a predictive model depends on the specific context in which it will be used, the cost of data collection, and the ease of interpreting the results. If it is used to select individual patients for more intensive case management, one must also consider the costs of treating false positives and not treating false negatives. Self-rated health and measures of IADL and functional limitations provided some additional predictive power, but less than DCG scores and chronic condition counts. While not as elaborate as the SF-12 or SF-36, these three questions are simple to ask. Whether they provide sufficient information to justify a data collection effort may depend on local costs and staff availability and expertise. (The SF-12 was first incorporated in the MEPS in Panel 5 and not available for model development using Panels 1–4.)

Some studies predicting high-cost cases have generated predictions of the full distribution of expenditures and then compared predicted to actual expenditures for those with predicted expenditures above a prespecified cutoff (Ash et al. 2001; Meenan et al. 2003). Instead of attempting to predict actual health care expenditures, we developed models to predict being in the upper 10 percent of the distribution of expenditures. Meenan et al. (1999) compared the two approaches, finding that they produced nearly equivalent sensitivities. Rakovski et al. (2002) found that linear and logistic regression models performed similarly in predicting utilization, but they did not examine costs.

Data on clinical conditions came from interviews with nonmedically trained respondents, not from medical records or claims databases. Thus, the potential for errors of omission or misclassification exists. It is possible that, had diagnoses been based on medical records data, the predictive power of the DCG risk score or the count of chronic conditions would have been even stronger. The sensitivity of household-reported medical conditions in the MEPS varies from 0.92 for diabetes and 0.82 for hypertension, to 0.58 for cerebrovascular disease (Machlin et al. 2009). Although agreement between self-report and medical records varies by condition and is often only fair, models using self-report comorbidity indices perform nearly as well as models using administrative data (Chaudhry, Jin, and Meltzer 2005; Susser,

McCusker, and Belzile 2008). Comprehensive medical record data are not available for all MEPS respondents, and examination of the performance of DCG scores based on such data could not be performed.

In sum, an unweighted count of the number of chronic conditions merits consideration in future efforts to develop expenditure prediction models. Studies in disease-specific groups of patients also point to the utility of counts of conditions (Melfi et al. 1995; Perkins et al. 2004; Farley, Harley, and Devine 2006). The current findings also support Baser, Palmer, and Stephenson's (2008) contention that combining several measures in the same model can be more informative than just selecting one, as condition counts improved prediction over DCG score categories alone.

ACKNOWLEDGMENTS

REFERENCES

AHRQ PUF Data Files. MEPS HC-087: 2004 Medical Conditions. Rockville MD: Agency for Healthcare Research and Quality, November, 2006 [accessed on April 14, 2009]. Available at http://www.meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h87/h87doc.pdf

AHRQ. Chronic Condition Indicator for ICD-9-CM. Agency for Healthcare Research and Quality [accessed on September 30, 2009]. Available at http://www.hcup-us.ahrq.gov/toolssoftware/chronic/chronic.jsp#download

Anderson, G. F., and J. Horvath. 2004. "The Growing Burden of Chronic Disease in America." *Public Health Reports* 119: 263–70.

Archer, K. J., and S. Lemeshow. 2006. "Goodness-of-Fit Test for a Logistic Regression Model Fitted Using Survey Sample Data." *The Stata Journal* 6: 97–105.

Ash, A. S., R. P. Ellis, G. C. Pope, J. Z. Ayanian, D. W. Bates, H. Burstin, L. I. Iezzoni, E. McKay, and W. Yu. 2000. "Using Diagnoses to Describe Populations and Predict Costs." *Health Care Financing Review* 21: 7–28.

Ash, A. S., M. A. Posner, J. Speckman, S. Franco, A. C. Yacht, and L. Bramwell. 2003. "Using Claims Data to Examine Mortality Trends Following Hospitalization for Heart Attack in Medicare." *Health Services Research* 38: 1253–62.

Ash, A. S., and M. Shwartz. 1999. "R^2: A Useful Measure of Model Performance When Predicting a Dichotomous Outcome." *Statistics in Medicine* 18: 375–84.

Ash, A. S., Y. Zhao, R. P. Ellis, and M. S. Kramer. 2001. "Finding Future High-Cost Cases: Comparing Prior Cost versus Diagnosis-Based Methods." *Health Services Research* 36 (part II): 194–206.

Baldwin, L.-M., C. N. Klabunde, P. Green, W. Barlow, and G. Wright. 2006. "In Search of the Perfect Comorbidity Measure for Use with Administrative Claims Data: Does It Exist?" *Medical Care* 44: 745–53.

Baser, O., L. Palmer, and J. Stephenson. 2008. "The Estimation Power of Alternative Comorbidity Indices." *Value in Health* 11: 946–55.

Charlson, M. E., R. E. Charlson, J. C. Peterson, S. S. Marinopoulos, W. M. Briggs, and J. P. Hollenberg. 2008. "The Charlson Comorbidity Index Is Adapted to Predict Costs of Chronic Disease in Primary Care Patients." *Journal of Clinical Epidemiology* 61: 1234–40.

Chaudhry, S., L. Jin, and D. Meltzer. 2005. "Use of a Self-Generated Charlson Morbidity Index for Predicting Mortality." *Medical Care* 43: 607–15.

Cohen, J. W., and N. A. Krauss. 2003. "Spending and Service Use among People with the Fifteen Most Costly Medical Conditions, 1997." *Health Affairs* 22: 129–38.

Cohen, J. W., A. Monheit, K. Beauregard, S. B. Cohen, D. C. Lefkowitz, D. E. Potter, J. P. Sommers, A. K. Taylor, and R. H. Arnett. 1996. "The Medical Expenditure Panel Survey: A National Health Information Resource." *Inquiry* 33: 373–89.

DeSalvo, K. B., T. M. Jones, J. Peabody, J. McDonald, S. Fihn, V. Fan, J. He, and P. Munther. 2009. "Health Care Expenditure Prediction with a Single Item, Self-Rated Health Measure." *Medical Care* 47: 440–7.

DxCG Inc. 2007. *DxCG Risk Smart™ Stand Alone V2.2 User Guide.* Boston, MA: DxCG Inc.

Elixhauser, A., C. A. Steiner, C. A. Whittington, and E. McCarthy. 1998. *Clinical Classifications for Health Policy Research: Hospital Inpatient Statistics, 1995.* Healthcare Cost and Utilization Project, HCUP-3 Research Note, AHRQ PUB 98-0049, Rockville, MD: Agency for Healthcare Research and Quality.

Fan, V. S., M. L. Maciejewski, C.-F. Liu, M. B. McDonnell, and S. D. Fihn. 2006. "Comparison of Risk Adjustment Measures Based on Self-Report, Administrative Data, and Pharmacy Records to Predict Clinical Outcomes." *Health Services and Outcomes Research Methodology* 6: 21–36.

Farley, J. F., C. R. Harley, and J. W. Devine. 2006. "A Comparison of Comorbidity Measurements to Predict Healthcare Expenditures." *American Journal of Managed Care* 12: 110–7.

Fleishman, J. A., J. W. Cohen, W. G. Manning, and M. Kosinski. 2006. "Using the SF-12 Health Status Measure to Improve Predictions of Medical Expenditures." *Medical Care* 44 (suppl): I-54–63.

Friedman, B., H. J. Jiang, A. Elixhauser, and A. Segal. 2006. "Hospital Costs for Adults with Multiple Chronic Conditions." *Medical Care Research and Review* 63: 327–46.

Hornbrook, M. C., and M. J. Goodman. 1995. "Assessing Relative Health Plan Risk with the RAND-36 Health Survey." *Inquiry* 32: 56–74.

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

Hwang, W., W. Helle, H. Ireys, and G. Anderson. 2001. "Out-of-Pocket Medical Spending for Care of Chronic Conditions." *Health Affairs* 20: 267–78.

Iezzoni, L. I. 2003. *Risk Adjustment for Measuring Health Care Outcomes.* Chicago, IL: Health Administration Press.

Machlin, S., J. W. Cohen, and K. Beauregard. 2008. *Health Care Expenses for Adults with Chronic Conditions, 2005.* Statistical brief #203, May, 2008, Rockville, MD, Agency for Healthcare Research and Quality [accessed April 14, 2009]. Available at http://www.meps.ahrw.gov/mepsweb/data_files/publications/st203/stat203.pdf

Machlin, S. R., and A. K. Taylor. 2000. *Design, Methods, and Field Results of the 1996 Medical Expenditure Panel Survey Medical Provider Component*, MEPS Methodology Report No. 9. Rockville, MD, Agency for Healthcare Research and Quality, AHRQ Pub No. 00-0028

Machlin, S. W., J. Cohen, A. Elixhauser, K. Beauregard, and C. Steiner. 2009. "Sensitivity of Household Reported Medical Conditions in the Medical Expenditure Panel Survey." *Medical Care* 47: 618–25.

Maciejewski, M. C., C.-F. Liu, A. Derleth, M. McDonnell, S. Anderson, and S. D. Fihn. 2005. "The Performance of Administrative and Self-Reported Measures for Risk Adjustment of Veterans Affairs Expenditures." *Health Services Research* 40: 887–904.

Maciejewski, M. C., C.-F. Liu, and S. D. Fihn. 2009. "Performance of Comorbidity, Risk Adjustment, and Functional Status Measures in Expenditure Prediction for Patients with Diabetes." *Diabetes Care* 32: 75–80.

Meenan, R. T., M. J. Goodman, P. A. Fishman, M. C. Hornbrook, M. C. O'Keeffe-Rosetti, and D. J. Bachman. 2003. "Using Risk-Adjustment Models to Identify High-Cost Risks." *Medical Care* 41: 1301–12.

Meenan, R. T., M. C. O'Keefe-Rosetti, M. C. Hornbrook, D. J. Bachman, M. J. Goodman, P. A. Fishman, and A. V. Hurtado. 1999. "The Sensitivity and Specificity of Forecasting High-Cost Users of Medical Care." *Medical Care* 37: 815–23.

Melfi, C., E. Holleman, D. Arthur, and B. Katz. 1995. "Selecting a Patient Characteristics Index for the Prediction of Medical Outcomes Using Administrative Claims Data." *Journal of Clinical Epidemiology* 48: 917–26.

Perkins, A. J., K. Kroenke, J. Unutzer, W. Katon, J. W. Williams, C. Hope, and C. M. Callahan, et al. 2004. "Common Comorbidity Scales Were Similar in Their Ability to Predict Health Care Costs and Mortality." *Journal of Clinical Epidemiology* 57: 1040–8.

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

Pope, G. C., K. W. Adamache, E. G. Walsh, and R. K. Khandker. 1998. "Evaluating Alternative Risk Adjusters for Medicare." *Health Care Financing Review* 20: 109–2.

Raftery, A. E. 1995. "Bayesian Model Selection in Social Research." In *Sociological Methodology*, edited by A. E. Raftery, pp. 111–163, Oxford, England: Blackwell.

Rakovski, C. C., A. K. Rosen, F. Wang, and D. R. Berlowitz. 2002. "Predicting Elderly at Risk of Increased Future Healthcare Use: How Much Does Diagnostic Information Add to Prior Utilization?" *Health Services and Outcomes Research Methodology* 3: 267–77.

Susser, S. R., J. McCusker, and E. Belzile. 2008. "Comorbidity Information in Older Patients at an Emergency Visit: Self-Report vs. Administrative Data Have Poor Agreement but Similar Predictive Validity." *Journal of Clinical Epidemiology* 61: 511–5.

Ware, J. E., M. Kosinski, and S. D. Keller. 1996. "A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity." *Medical Care* 34: 220–33.

Ware, J. E., and C. D. Sherbourne. 1992. "The MOS 36-Item Short-Form Health Survey (SF-36): I Conceptual Framework and Item Selection." *Medical Care* 30: 473–8.

Zheng, B., and A. Agresti. 2000. "Summarizing the Predictive Power of a Generalized Linear Model." *Statistics in Medicine* 19: 1771–8.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix Table SA2: Key Conditions and Associated Clinical Condition Categories.

Appendix Table SA3. Descriptive Statistics for Two Sets of MEPS Panels.

Appendix Table SA4. Adjusted Odds-Ratios of Being in the Top 10% of the Expenditure Distribution, Year 2, Panels 1–4, Model 9.

Appendix Table SA5. Goodness-of-Fit Indicators for Several Logistic Models Predicting High (Upper 5%) Expenditures in Year 2, MEPS Panels 1–4.

Appendix Table SA6. Goodness-of-Fit Indicators for Several Logistic Models Predicting High (Upper 10%) Expenditures in Year 2, MEPS Panels 1–4. Ages 18–64 Only ($n = 31,424$).

Appendix Table SA7. Goodness-of-Fit Indicators for Several Logistic Models Predicting High (Upper 10%) Expenditures in Year 2, MEPS Panels 1–4. Ages 65 and Older ($n = 6,491$).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.