Predicting Pharmacy Costs and Other Medical Costs Using Diagnoses and Drug Claims

Yang Zhao, PhD, Arlene S. Ash, PhD,*† Randall P. Ellis, PhD,*‡ John Z. Ayanian, MD, MPP,*§¶ Gregory C. Pope, MS,‖ Bruce Bowen, PhD,** and Lori Weyuker, ASA††*

Background: Predicting health care costs for individuals and populations is essential for managing care. However, the comparative power of diagnostic and drug data for predicting future costs has not been closely examined.

Objective: We sought to compare the predictive performance of claims-based models using diagnoses, drugs claims, and combined data to predict health care costs.

Subjects: More than 1 million commercially insured, nonelderly individuals in a national (MEDSTAT MarketScan®) research database comprised our sample.

Measures: We used 1997 and 1998 drug and diagnostic profiles to predict costs in 1998 and 1999, respectively. To assess model performance, we compared R^2 values and predictive ratios (predicted costs/actual costs) for important subgroups.

Results: Models using both drug and diagnostic data best predicted subsequent-year total health care costs (highest $R^2 = 0.168$ versus 0.116 and 0.146 for models based on drug or diagnostic data alone, respectively), with highly accurate predictive ratios (0.95–1.05) for subgroups of patients with major medical conditions. Models predicting pharmacy costs had substantially higher R^2 values than models predicting other medical costs (highest R^2 0.493 versus 0.124). Drug-based models predicted future pharmacy costs better than diagnosis-based models (highest $R^2 = 0.482$ versus 0.243), whereas diagnosis-based models predicted total costs (highest $R^2 = 0.146$ versus 0.116) and nonpharmacy costs (highest $R^2 = 0.116$ versus 0.071) more effectively than drug-based models. Newer models had markedly higher R^2 values than older ones, largely because of richer data rather than model refinements.

Conclusions: Combined drug and diagnostic data predicts total health care costs better than either type of data alone. Pharmacy spending is particularly predictable from drug data, whereas diagnoses are more useful than drugs for predicting other medical costs and total costs. Using even slightly more recent data can substantially boost model performance measures; thus, model comparisons should be conducted on the same dataset.

Key Words: risk adjustment, predictive models, population disease burden, pharmacy profile, actuarial prediction

(*Med Care* 2005;43: 34–43)

Medical and financial managers increasingly use current illness indicators to predict subsequent-year health care needs and costs.[1,2] Predictive models commonly rely on some combination of demographics (typically age and gender), diagnoses recorded during medical encounters, and prescription drug utilization data.

Models using diagnoses from claims to predict future health care costs were introduced in the 1980s.[3,4] In January 2004, to better link Medicare payments to health plans to the health status of their enrolled beneficiaries, the Centers for Medicare and Medicaid Services (CMS) introduced a claims-based Diagnostic Cost Group Hierarchical Condition Category (DCG/HCC) model to partially reimburse health plans that enroll Medicare beneficiaries. This CMS-HCC model uses demographics and a diagnosis-based medical profile captured during all clinician encounters—both inpatient and outpatient—to produce a health-based measure of future medical need.[5,6] Several states also make risk-adjusted Medicaid payments to providers.[7–9]

Many studies have confirmed the ability of diagnosis-based models to predict total health care costs in privately insured populations.[10–14] However, diagnoses from multiple sites of care often accrue slowly in centralized databases, whereas outpatient pharmacy claims are generated electron-

From*D_xCG, Inc., Boston, Massachusetts; the †Department of General Internal Medicine, Boston University, Boston, Massachusetts; the ‡Department of Economics, Boston University, Boston, Massachusetts; the §Division of General Medicine, Brigham and Women's Hospital, and the ¶Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts; the ‖Program on Health Economics and Financing, RTI International, Waltham, Massachusetts; **Multimetrics, Sisters, Oregon; and ††LW Consulting, Mill Valley, California.

Funded by DxCG, Inc, in Boston, Massachusetts, which develops and markets software to implement the predictive models described in this study.

Yang Zhao, Arlene S. Ash, Randall P. Ellis, and Gregory C. Pope are DxCG shareholders.

Dr. Zhao is now Senior Health Outcomes Scientist, Eli Lilly and Company, Indianapolis, Indiana.

Reprints: Yang Zhao, PhD, Eli Lilly and Company, Lilly Corporate Center, DC 1833, Indianapolis, IN 46285. E-mail: zhao_yang@lilly.com.

ISSN: 0025-7079/05/4301-0034

ically as prescriptions are filled. Pharmacy claims have been studied in United States privately insured, Medicaid, and veterans populations.[15–25] Since 2003, the Netherlands has used a drug-based model to reimburse sickness funds for their members.[26] Drug-based models predict future total costs well but less well than diagnosis-based models.[12,15,17,23] Accuracy may be improved by using both diagnoses and drug claims as predictors.[15,20] However, no studies have combined diagnoses from all sites with comprehensive outpatient drug claims to predict total health care costs. As more health plans have both diagnoses and drug claims available, it is important to assess how predictive accuracy is affected when both types of data are combined.

The extensive risk adjustment literature during the past 2 decades has largely been devoted to predicting future total health care costs. Other than modeling for the Medicare program, in which outpatient drug use is not reimbursed, few studies examined the predictability of medical costs excluding pharmacy costs[17] and pharmacy costs alone.[27–31] Predicting pharmacy costs is particularly important in light of the recent legislation authorizing Medicare prescription drug benefits.

Compared with earlier reports,[1,3–4,7,9–11,13,14,16,17,19] the predictive performance of newer predictive models has improved markedly.[8,12,15,18,20,23,32] However, the data used are more recent and more complete. No study has systematically distinguished improvements as the result of more refined predictive models versus newer data.

In this study, we comprehensively characterized the predictive performance of different sources of data to predict various future health care costs and systematically explored whether observed improvements in predictive power were caused by newer data or more refined predictive models. Using a large, nationally representative cohort of commercially insured people younger than the age of 65, we examined 2 widely-used families of claim-based models: RxGroups® Releases 1 and 2, which rely on pharmacy data,[15] and the Diagnostic Cost Group Hierarchical Condition Category (DCG/HCC) models Releases 5 and 6, which use diagnoses.[10,32] Both models were developed by D_xCG, Inc. (Boston, MA). The Society of Actuaries recently has compared the performance of early versions of the RxGroup and DCG/HCC models with other drug and diagnosis models to predict next year's total health care costs.[12] Focusing on these 2 types of models, we systematically evaluated their predictive accuracy across 4 dimensions:

1. Model types (predictions based on drug, diagnosis or both kinds of data combined).
2. Cost outcomes (pharmacy costs, other medical costs, or total costs).
3. Model versions (previously published models[10,15] or newer, more clinically-refined models).
4. Sequential years (either 1997 data used to predict 1998 costs or 1998 data used to predict 1999 costs).

METHODS

Data Source

We used the Commercial Claims and Encounters Database of MEDSTAT's MarketScan® Research Database, which includes inpatient, outpatient, and pharmacy claims for individuals enrolled in more than 100 health plans contracting with large employers, state and local governments, and public organizations in the United States. Both fee-for-service and capitated health plans that submitted encounter data are included. The 1998–1999 data include 1.3 million individuals younger than the age of 65 who were enrolled in a participating health plan for at least 1 month in both 1998 (year 1) and 1999 (year 2), with pharmacy coverage throughout their enrollment period. The analogous 1997–1998 MarketScan data included 1.1 million individuals, and 874,000 people appeared in both the 1997–1998 and 1998–1999 data. The overlap of subjects between the 2 years is a strength of our study, enabling us to assess whether changes in predictive accuracy were related to more recent data for a relatively stable population.

All outcomes are year-2 costs: total, pharmacy (outpatient only), and nonpharmacy (total minus pharmacy). Costs include deductibles, copayments, coinsurance, and coordination-of-benefits payments. For partial-year enrollees, we annualized expenditures (actual spending divided by eligibility fraction) and used eligibility fractions as modeling weights. For example, a person leaving a health plan in June 1999 with $3000 in health care costs during the previous 6 months contributes half of an observation with annualized spending of $6000.

Pharmacy Categories

We classified all National Drug Codes (NDCs), mainly according to the drug's "function" (therapeutic indication). RxGroups differ from the "Chronic Disease Score" pharmacy model and its refinements, whose categorizations are keyed to "inferred medical condition."[16–20,25,26] The previous RxGroup model version 1.0 (developed in 2000) mapped 58,000 NDCs into 127 mutually exclusive categories (called "RxGroups") primarily based on therapeutic indication.[15] Version 2.0 classified the January 1, 2002, listing of more than 76,000 NDCs into 155 RxGroups. To better distinguish severity level and likely medical problems being treated, the newer system categorized NDCs along 4 dimensions: active ingredient(s), strength, route of administration, and dose. For example, the RxGroup "lipid-lowering agents" in RxGroup model version 1.0 was split into 2 RxGroups based on active ingredients: statin versus other. We distinguished 3 "routes" for asthma/chronic obstructive pulmonary disease drugs: injectable, oral, and inhaled; and the inhaled category split further into 3 ingredient-based groups: beta agonist, steroid, or other. The different uses of leuprolide for men and women yielded 2 distinct RxGroups: leuprolide in men (prostate cancer) and women (endometriosis).

A risk-adjustment model used to calculate payments or allocate resources should be minimally sensitive to discretionary practice patterns or coding idiosyncrasies. To increase our model's robustness to common variations in prescribing, we imposed hierarchies among RxGroups used to treat the same medical problem. For example, when the higher-ranked RxGroup "insulin," is present, the lower-ranked "oral diabetic agents" is ignored (Fig. 1A). The hierarchy for ophthalmic problems is more complex (1B), with 6 RxGroups in 3 tiers, with the model "recognizing" as many groups that may be present in the highest tier. For example, a person's predicted cost could be based on as many as all 3 ophthalmic categories 90 through 92, but only if no drugs in RxGroups 87 through 89 are recorded.

Diagnostic Categories

The Diagnostic Cost Group Hierarchical Condition Category (DCG/HCC) model uses age, sex, and diagnoses from both inpatient hospital admissions and outpatient professional medical services to record the presence of multiple medical problems and predict health care costs. More than 15,000 diagnoses map to clinically homogenous groups, called DxGroups, which further cluster into broader Condition Categories (CCs).[10,32] In updating the model from Release 5 to Release 6, we incorporated all recently introduced diagnoses and increased the number (and thus specificity) of DxGroups from 541 to 781 and of CCs from 118 to 184.

The more clinically specific DxGroups and CCs better support disease management, especially in the areas of diabetes,

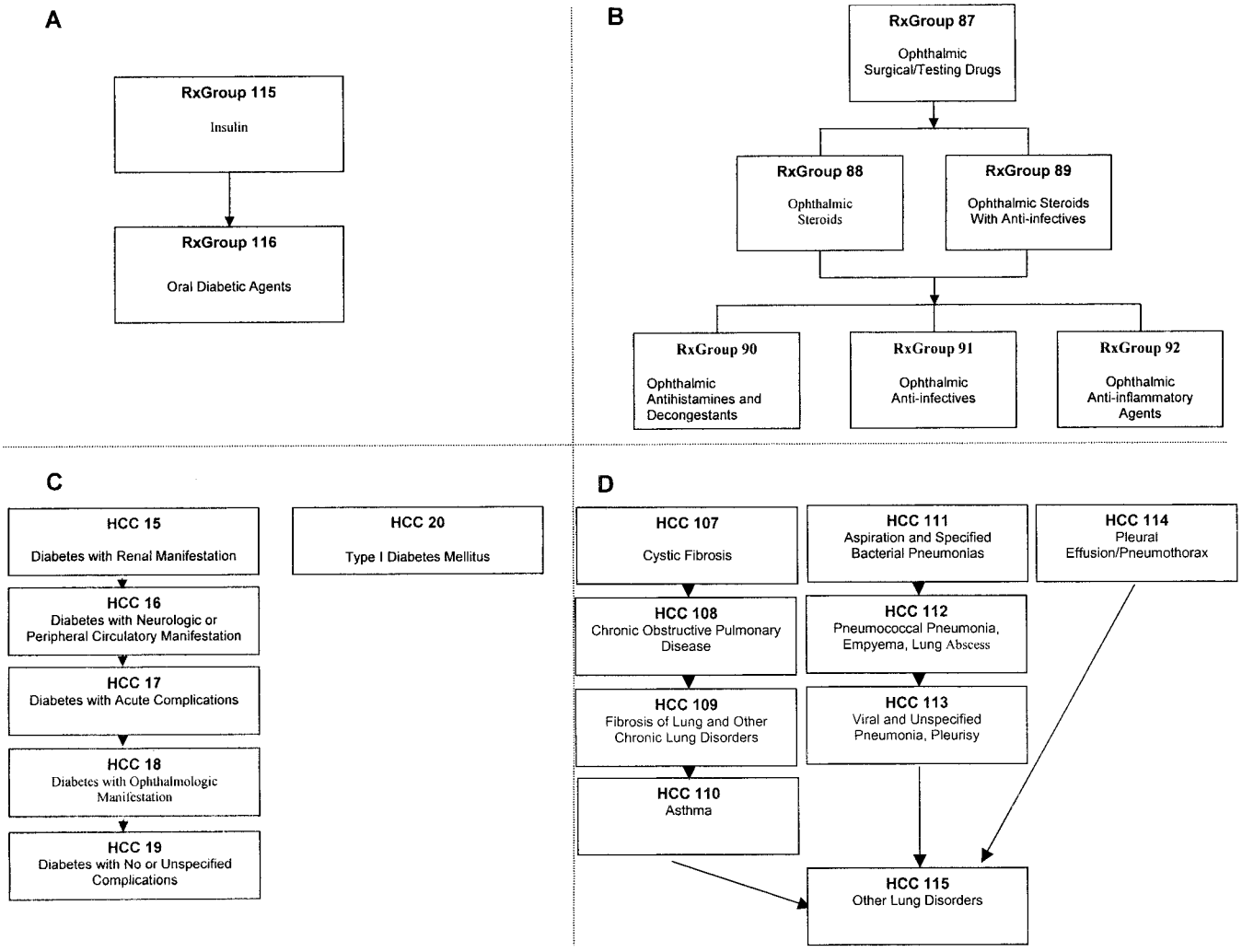

FIGURE 1. Sample RxGroups and condition category hierarchies. A, Diabetes drug hierarchy; B, ophthalmic drug hierarchy; C, diabetes condition hierarchy; D, pulmonary condition hierarchy.

TABLE 1. Demographics and Utilization Experience in Privately Insured Populations: 1997–1998 versus 1998–1999 samples*

	1997–1998	1998–1999	% Change†
Number of people	1,083,405	1,292,288	19.3
Percent female	50.2	50.7	1.0
Mean age in year 1	32.8	33.4	1.8
Age distribution			
0–17 years	26.2	25.9	−1.0
18–44 years	41.5	40.2	−3.0
45–64 years	32.3	33.8	4.7
Year-1 utilization statistics			
Percent with at least one diagnosis	71.9	73.9	2.8
Mean no. valid diagnoses per person	10.9	11.6	6.6
Mean no. distinct valid diagnoses per person	3.71	3.99	7.5
Mean no. HCCs per person	2.37	2.52	6.3
Percent with at least one prescription	63.9	66.4	3.9
Mean no. RxGroups per person	2.41	2.46	2.1
Year-2 total cost‡			
Mean	$1901	$2053	8.0
CV§	448	386	−13.9
Year-2 nonpharmacy cost			
Mean	$1531	$1601	4.6
CV§	539	471	−12.6
Year-2 pharmacy cost			
Mean	$370	$452	22.1
CV§	276	278	0.6

*For people with at least one month eligibility in both year 1 and year 2 in the MarketScan Research Database.
†All differences in means between years are significant at $P < 0.0001$ level after correcting for correlation induced by panel design.
‡Sum of inpatient, outpatient, and pharmacy costs.
§CV is the coefficient of variation defined as 100*standard deviation/mean.

liver disease, heart disease, and mental illness. We added a CC for type I Diabetes and mapped a few DxGroups, which were related to diabetes or congestive heart failure, to 2 CCs.

The CCs categorize a person's recorded medical conditions but may contain inconsistent or redundant information. Hierarchies are imposed on the CCs (leading to an HCC profile) to address clinical and statistical concerns. A hierarchy may replace a provisional diagnosis with a more definitive one (such as a specific cancer over "uncertain neoplasm" or asthma over "persistent cough"); mark the presence of or progression to a more severe disease state (such as metastatic versus nonmetastatic cancer or myocardial infarction versus angina); or resolve an inconsistency (such as, when a single individual receives codes indicating both moderate and severe developmental deficiencies). Thus, individual HCC markers are more informative than CC markers, and models based on HCC profiles are less apt to interpret redundant coding as evidence of greater medical need. Figure 1C illustrates the diabetes hierarchy; except for HCC 20 (type I diabetes mellitus), each CC dominates the next lower one; any individual can be assigned to at most one of HCCs 15 through 19. In contrast, the lung hierarchy has 3 subhierarchies (Fig. 1D); thus, a single person can be classified into as many as 3 lung HCCs. The complete CCs and hierarchies are listed elsewhere.[32]

Model Development

Using each classification system, we estimated models to predict each of our 3 cost outcomes (total, nonpharmacy and pharmacy only) in the subsequent year. We used weighted ordinary least squares regression; an observation's weight is the fraction of year 2 during which the person is eligible to incur costs. In all models, we added indicator variables to distinguish among 16 age/sex categories. In updating the RxGroups models, we included (interaction) terms when the joint effect of combinations of drugs on next year's costs was not additive. We considered all interactions explored in the earlier models,[15] and added additional interactions that clinical advisors deemed important. Twenty-seven interactions (2-, 3-, and 4-way) retained in the model pertained to at least 100 people (of 1.3 million). All were statistically significant at the 0.05 level and were judged to

TABLE 2. R^2 Values for Predicting Year-2 Total Costs in Privately Insured Populations: 1997–1998 versus 1998–1999 Samples*

	R^2 Values		% Change Attributable to	
	1997–1998	1998–1999	New Data	Both†
RxGroup model				
Old classification‡	0.084	0.115	38.2	
New classification§	0.084	0.116	38.3	
% Change attributable to new classification	0.7	0.8		39.3
DCG/HCC model				
Old classification¶	0.113	0.137	21.6	
New classification‖	0.117	0.146	25.0	
% Change attributable to new classification	3.6	6.6		29.6
Combined (Rx+DCG) model				
Old classifications‡¶	0.126	0.160	26.6	
New classifications§‖	0.131	0.168	28.9	
% Change attributable to new classification	3.5	5.3		33.4
% Change from (new) RxGroup to combined model	55.2	44.6		
% Change from (new) DCG/HCC to combined model	11.4	14.9		

*For people with at least 1 month of eligibility in both year 1 and year 2 in the MarketScan Research Database (1997–1998 sample: n = 1,083,405; 1998–1999 sample: n = 1,292,288). Total costs are the sum of inpatient, outpatient, and pharmacy costs.
†Better data and new classification(s).
‡Old drug-based model (RxGroup) predicts from 127 RxGroup drug categories.[15]
§Old diagnosis-based (DCG/HCC) model predicts from 118 hierarchical condition categories (HCCs).[10]
¶New drug-based model (RxGroup) predicts from 155 RxGroup drug categories, as described in the Methods section.
‖New diagnosis-based (DCG/HCC) model predicts from 184 hierarchical condition categories (HCCs), as described in the Methods section.

be clinically sensible. We also included interactions among coexisting medical conditions (HCCs) in the updated DCG/HCC models. We had previously examined interactions among 6 common chronic medical conditions for the Medicare program[32]: diabetes, cerebrovascular disease, congestive heart failure, vascular disease, coronary artery disease, and chronic obstructive pulmonary disease. In updating the DCG/HCC model, we explored 2- and 3-way interactions among 8 medical conditions: the original 6 plus uncompleted pregnancy and respiratory disease. To explore meaningful cost differences by age, we further tested interactions between age (younger than age 18) and all HCCs. As above, all 28 interactions retained in the model had at least 100 people and were both clinically reasonable and statistically significant ($P < 0.05$).

We excluded some RxGroups or HCCs from the models if they had a negative coefficient (although typically only slightly so), which would reduce predictions for people with such prescriptions or diagnoses. RxGroups that do not predict future costs (based on clinical judgment) also were dropped. These groups included drugs for "major diagnostic testing," drugs that are available in over-the-counter forms (ie, "OTC drugs"), and drugs commonly used for a range of typically minor medical problems such as fungal skin infection ("miscellaneous, recognized NDCs").

We enforced monotonicity in the models so that no HCC had a smaller coefficient than a less clinically severe HCC. This avoids "paying less" for a potentially more serious medical problem. For example, because respiratory arrest would otherwise have had a lower coefficient than cardiorespiratory failure and shock, we forced the model to calculate the same coefficient for both groups.

We also estimated a simple, additive "combination" model (Rx+DCG) for each outcome, using (in addition to age/sex indicators) both drug and disease predictors: RxGroups and their interactions, HCCs and their interactions, and interactions between age and HCCs. We did not explore interactions between RxGroups and HCCs for this model.

Data Analysis

For each population (1997–1998 and 1998–1999), we described its demographic distribution, year-1 diagnoses and drug use, and year-2 costs. We used *t* statistics for the statistical significance of the differences of means. To correct for the nonindependence of the 2 sets of MarketScan data, all tests were adjusted for the correlation between observations for the same person in the 2 data sets (see Appendix).

To measure the predictive power for each model type (RxGroup, DCG/HCC, and Rx+DCG), we assessed the mod-

TABLE 3. R^2 Values for Predicting Year-2 Total Costs Without Pharmacy in Privately Insured Populations: 1997–1998 versus 1998–1999 samples*

	R^2 Values		% Change Attributable to	
	1997–1998	1998–1999	New Data	Both†
RxGroup model				
Old classification‡	0.053	0.070	32.6	
New classification§	0.053	0.071	33.0	
% change attributable to new classification	1.1	1.4		34.5
DCG/HCC model				
Old classification¶	0.089	0.107	19.6	
New classification‖	0.094	0.116	23.1	
% Change attributable to new classification	5.5	8.5		29.8
Combined (Rx+DCG) model				
Old classification‡¶	0.096	0.115	20.6	
New classifications§‖	0.100	0.124	23.9	
% Change attributable to new classification	4.6	7.5		29.6
% Change from (new) RxGroup to combined models	88.1	75.4		
% Change from (new) DCG/HCC to combined models	6.2	6.9		

*For people with at least 1 month of eligibility in both year 1 and year 2 in the MarketScan Research Database (1997–1998 sample: n = 1,083,405; 1998–1999 sample: n = 1,292,288).
†Better data and new classification(s).
‡Old drug-based model (RxGroup) predicts from 127 RxGroup drug categories.[15]
§Old diagnosis-based (DCG/HCC) model predicts from 118 hierarchical condition categories (HCCs).[10]
¶New drug-based model (RxGroup) predicts from 155 RxGroup drug categories, as described in the Methods section.
‖New diagnosis-based (DCG/HCC) model predicts from 184 hierarchical condition categories (HCCs), as described in the Methods section.

els' R^2 values (percentage of variation in costs explained). To disentangle the effects of older and newer models and data, we examined the performance of both old and new models, for each model type, on both older and newer data. That is, for each model type (drug-based, diagnosis-based, and combined) and each of the 3 cost outcomes we applied both the older and newer versions of the model to both the older and newer datasets.

The ratio of predicted costs to actual costs within selected disease cohorts is widely used to assess model accuracy.[10–12] When a model predicts well for a group, this "predictive ratio," or PR, approximately equals 1.00; when it underpredicts, the PR is less than 1; PRs greater than 1 indicate overprediction. We used 1998–1999 as the validation sample, and applied models estimated from 1997–1998 to generate predictions for each person and PRs for 3 kinds of subgroups as identified in 1998: cohorts defined by the presence of a relevant diagnosis from either inpatient or outpatient settings; cohorts defined by the presence of a relevant drug claim; and cohorts defined by total health care costs in the initial year.

RESULTS

Demographic distributions and medical care costs changed little between 1997–1998 and 1998–1999 (Table 1). The later cohort was slightly older (mean, 33.4 versus 32.8 years). Total health care costs increased by 8.0%, reflecting a substantial increase in pharmacy costs (22.1%), and a smaller increase in other medical costs (4.6%). Relative variation, as measured by the coefficient of variation (CV, equal to 100 times the standard deviation divided by the mean) declined for total costs (−13.9%) and nonpharmacy costs (−12.6%) but not for pharmacy costs (0.6%). In the newer data, more people had at least one diagnosis; also, there were more valid diagnoses, distinct valid diagnoses, and HCCs per person. The proportion of people with any prescription and the number of distinct types of drugs (RxGroups) per person were also larger in the newer data.

The combined model (Rx+DCG) predicted total costs best, followed closely by the diagnosis model and more distantly by the drug model (Table 2). For example, with the new models and data, the respective R^2 values were 0.168 (Rx+DCG), 0.146 (DCG/HCC), and 0.116 (RxGroup). That is, moving from the RxGroup to the DCG/HCC to the combined model increased the R^2 first by 26% and then by an additional 15%. Moving from older to newer data (while holding the model fixed) also produced striking improvements. Among new models to predict total cost, the switch to new data always increased R^2 values by at least 25% and often by much more; the new drug model's R^2 was 0.084 in

the old data and 0.116 in the new, a 38% increase. In contrast, when holding the data set fixed and moving from older to newer predictive models, R^2 values always improved, although only minimally for the RxGroup models, and always modestly (the largest improvement was from 0.137 to 0.146, a 7% increase, for the DCG/HCC models in the new data).

Models predicting nonpharmacy costs (the most variable outcome) had lower R^2 values than the models predicting total costs (Table 3). The diagnosis models predicted these costs better than the drug models, and the combined models provided modestly better predictions than the diagnosis models. Models predicting pharmacy costs (the least variable outcome) had the highest R^2 values, between 0.47 and 0.49 for all models relying on drug data and between 0.21 and 0.24 for the diagnosis-based models alone (Table 4). Adding diagnoses to drug claims increased the predictive accuracy only minimally (between 1 and 3%). In contrast to the findings for predicting total and nonpharmacy costs, newer data did not yield consistently higher R^2 values (maximum increase was 0.7%), with the diagnosis-based models' R^2 values actually decreasing by 7 to 8%. Switching from older to newer models yielded only modest improvements in R^2 values (2 to 9%).

Table 5 shows means and predictive ratios for 1999 total costs (new models only) for 15 groups, 5 each based on specific kinds of information from the prior year: medical conditions, drug use, and costs. The most expensive of these groups (those who had an acute myocardial infarction or who were in the top 5% of spending during 1998) incurred costs more than 5 times as high as the 1999 average of $2053, whereas spending in the least expensive group examined (those with below median 1998 spending) had costs in 1999 that were about one third of this average.

All 3 models predicted these large cost differences with reasonable accuracy, with the largest deviations occurring for the RxGroup model in the most extreme cost-based groups. Specifically, the mean RxGroup prediction was $955 (= 1.36 times $702) for the below-median cost group, and $7407 (= 0.69 times $10,735) for the group with the highest 5% of prior-year costs. The DCG/HCC predictive ratios were 1.12 and 0.80, and the combined model predictive ratios were 1.00 and 0.88 for these same groups, respectively. The diagnosis-based model predicted costs very well for the diagnosis-identified groups (predictive ratios between 0.98 and 1.02), but somewhat underpredicted the groups defined by the use of drugs (predictive ratios of 0.81–0.90). Analogously, the drug-based model did well with the groups defined by the use of drugs (predictive ratios of 0.95–1.01) whereas underpredicting the costs of the medical condition cohorts (predictive ratios of 0.81–0.90). Only the combined model predicted group averages within 5% of actual costs for all diagnosis and drug-based subgroups.

TABLE 4. R^2 Values for Predicting Year-2 Pharmacy Costs in Privately Insured Populations: 1997–1998 versus 1998–1999 Samples*

	R^2 Values		% Change Attributable to	
	1997–1998	1998–1999	New Data	Both†
RxGroup model				
Old classification‡	0.472	0.474	0.4	
New classification§	0.479	0.482	0.7	
% Change attributable to new classification	1.6	1.9		2.3
DCG/HCC model				
Old classification¶	0.225	0.207	−8.2	
New classification‖	0.243	0.225	−7.3	
% Change attributable to new classification	7.9	8.9		0.0
Combined (Rx+DCG) model				
Old classifications‡¶	0.478	0.479	0.1	
New classifications§‖	0.491	0.493	0.3	
% Change attributable to new classification	2.8	3.0		3.1
% Change from (new) RxGroup to combined models	2.6	2.2		
% Change from (new) DCG/HCC to combined models	102.6	119.2		

*For people with at least 1 month of eligibility in both year 1 and year 2 in the MarketScan Research Database (1997–1998 sample: n = 1,083,405; 1998–1999 sample: n = 1,292,288).

†Better data and new classification(s).

‡Old drug-based model (RxGroup) predicts from 127 RxGroup drug categories.[15]

§Old diagnosis-based (DCG/HCC) model predicts from 118 hierarchical condition categories (HCCs).[10]

¶New drug-based model (RxGroup) predicts from 155 RxGroup drug categories, as described in the Methods section.

‖New diagnosis-based (DCG/HCC) model predicts from 184 hierarchical condition categories (HCCs), as described in the Methods section.

TABLE 5. Predictive Ratios for Next Year's Total Costs for Disease-, Drug Use- and Cost-Defined Groups

	n	Mean Total 1999 Cost	Predictive Ratios for New Models		
			RxGroup*	DCG/HCC†	Rx+DCG*†
Medical condition groups					
Acute myocardial infarction	2571	10,949	0.86	1.00	1.05
Asthma	38,361	3921	0.90	0.98	1.00
Chronic obstructive pulmonary disease (COPD)	35,603	5080	0.81	0.98	1.00
Depression	48,611	5167	0.85	1.01	1.01
Diabetes	33,083	7613	0.84	1.02	1.03
Drug utilization groups					
Antidepressants	90,335	5888	0.98	0.82	0.99
Asthma/COPD	83,877	3756	0.95	0.86	0.95
Diabetes	23,391	7450	1.01	0.90	1.03
Lipid-lowering	60,864	5933	1.01	0.86	1.01
Ulcer/gastroesophageal reflux disease (GERD)	80,239	6553	1.00	0.81	1.00
1998 spending percentiles					
Lowest 50%	646,144	702	1.36	1.12	1.00
Next highest 30%	387,686	1872	1.09	1.12	1.08
Next highest 10%	129,229	3638	0.99	1.01	1.04
Second highest 5%	64,615	4809	0.97	0.96	1.03
Highest 5%	64,614	10,735	0.69	0.80	0.88

Models were fit to MarketScan Research Database 1997–1998 data (n = 1,083,405) and validated on analogous 1998–1999 data (n = 1,292,288). Predictive ratios equal model-predicted 1999 costs for the specified group divided by actual costs in 1999. Medical condition groups consist of people with at least 1 relevant diagnosis in 1998 from any inpatient or outpatient setting; drug utilization groups, those with at least one relevant pharmacy fill in 1998.

*New drug-based model (RxGroup), using 155 RxGroup drug categories, as described in the Methods section.

†New diagnosis-based (DCG/HCC) model, using 184 hierarchical condition categories (HCCs), as described in the Methods section.

DISCUSSION

Predictive modeling based on claims data is an important tool for managing the financing and delivery of health care. Both the drug and diagnosis-based classification systems evaluated in this study are more clinically detailed than their precursors, making them more useful for cost profiling and disease management. Each model can identify and predict costs for clinically important subgroups. Using a recent, large, and nationally representative research database for privately insured individuals younger than the age of 65, this study compared the performance of models using different types of data (diagnoses, drugs, or both) and model versions (previously published models versus updated models) to predict different components of future health care costs (total, nonpharmacy, and pharmacy costs) in older and newer data. Combining diagnoses with drug claims substantially improved predictions of future total costs but only marginally improved on the ability of diagnosis-only models to predict nonpharmacy costs or drug-only models to predict pharmacy costs.

Drug costs were far more predictable than total or nonpharmacy cost. Relying only on the list of drugs ever used in a year (and not number of scripts, number of refills, dosages, or strength), RxGroup models explained nearly 50% of the variation in pharmacy cost during the subsequent year. Although it makes sense that current pharmaceutical costs and drug codes (NDCs) for a person are highly correlated with future pharmacy spending, our study is the first to quantify how accurately drug models can predict next year's pharmacy costs. It is likely that an RxGroup model that additionally tracks the number of prescriptions, or other volume measures, will provide even more robust predictions, especially among people who use drugs for chronic medical conditions.

Although we studied a privately insured population that was younger than the age of 65, this finding also has important implications for recently enacted pharmacy benefits in the Medicare program. Because elderly adults often have many chronic illnesses treated with drugs, their pharmacy costs may be even more predictable than what we found in a younger population. Therefore, stakeholders who have access to current patterns of drug use will be able to identify future high-cost users of pharmacy benefits. To address the potential problem of biased selection or "cherry picking" of Medicare beneficiaries with low predicted drug costs, CMS will need

access to the same drug utilization information that private health plans providing drug benefits have.

Although updated models were more predictive than earlier models for total or nonpharmacy costs, more improvement came from newer data (with more diagnoses and prescriptions) than from more clinically refined classifications. This finding is important because models developed on more recent data typically perform better than models developed from older data. Thus, credible comparisons of the performance of different models require evaluation on the same data. The more recent, richer diagnostic data provided more accurate predictions, but we cannot readily determine how much of the increased coding of diagnoses reflected a true shift in illness burden or better data collection. Increases in drug utilization can also be affected by changes in clinical guidelines. When coding and prescribing practices change rapidly, diagnosis- and drug-based models may not reliably identify true changes in need.

Current diagnosis- and drug-based models are powerful predictors of future cost. Each model captures population disease burden reasonably well and can be used to monitor or allocate use of health care resources. Drug and diagnosis models explain 12% and 15% of the variation in total cost, respectively, and the model combining both types of data explains 17% of this variation. Drug data are far superior for predicting pharmacy costs, whereas the diagnosis-based risk adjustment model better characterizes the population and more accurately predicts total and nonpharmacy cost. When more timely predictions are important, drug-based predictive models can provide an attractive alternative for predicting even total and nonpharmacy costs. As claims data become richer and more informative, the previously anticipated boundary of 20% for the explanatory power of claims-based models to predict total costs in general populations[4,33] may soon be surpassed.

ACKNOWLEDGMENTS

We are grateful to Marilyn Kramer of DxCG for helpful comments on an earlier draft of the manuscript.

REFERENCES

1. Pope GC, Ellis RP, Ash AS, et al. Principal inpatient diagnostic cost group model for Medicare risk adjustment. *Health Care Financ Rev*. 2000;21:93–118.
2. Tuohy C. A "new buzz" infects health care management. *Risk Insur*. 2004;Jan 1: http://www.riskandinsurance.com/040101_tech_1. asp.
3. Ash AS, Porell F, Gruenberg L, et al. Adjusting Medicare capitation payments using prior hospitalization. *Health Care Financ Rev*. 1989;10: 17–29.
4. Newhouse JP, Manning WG, Keeler EB, et al. Adjusting capitation rates using objective health measurers and prior utilization. *Health Care Financ Rev*. 1989;10:41–54.
5. Centers for Medicare and Medicaid Services. "Cover letter. Announcement of calendar year (CY) 2004 Medicare+Choice payment rates." Available at: http://cms.hhs.gov/healthplans/rates/2004/cover.asp.
6. Pope GC, Kautter J, Ellis RP, et al. Risk Adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev*. 2004;25:119–141.
7. Kronick R, Dreyfus T, Lee L, et al. Diagnostic risk adjustment for Medicaid: the disability payment system. *Health Care Financ Rev*. 1996;17:7–33.
8. Kronick R, Gilmer T, Dreyfus T, et al. Improving health-based payment for Medicaid beneficiaries: CDPS. *Health Care Financ Rev*. 2000;21: 29–64.
9. Weiner JP, Tucker AM, Collins AM, et al. The development of a risk-adjusted capitation payment system: the Maryland Medicaid model. *J Ambul Care Manage*. 1998;21:29–52.
10. Ash AS, Ellis RP, Pope GC, et al. Using diagnoses to describe populations and predict costs. *Health Care Financ Rev*. 2000;21:7–28.
11. Ash AS, Byrne-Logan S. How well do models work? predicting health care costs. Proceedings of the Section on Statistics in Epidemiology of the American Statistical Association, Dallas, 1998.
12. Cumming RB, Knutson D, Cameron BA, et al. A comparative analysis of claims-based methods of health risk assessment for commercial populations. Final report to the Society of Actuaries. Available at: http://www.soa.org/sections/riskadjfinalreport1. pdf. 2002.
13. Starfield B, Weiner JP, Mumford L, et al. Ambulatory care groups: a categorization of diagnoses for research and management. *Health Serv Res*. 1991;26:53–74.
14. Weiner JP, Starfield BH, Steinwachs DM, et al. Development and application of a population-oriented measure of ambulatory case-mix. *Med Care*. 1991;29:452–472.
15. Zhao Y, Ellis RP, Ash AS, et al. Measuring population health risks using inpatient diagnoses and outpatient pharmacy data. *Health Serv Res*. 2001;36:180–193.
16. Clark DO, Von Korff M, Saunders K, et al. A chronic disease score with empirically derived weights. *Med Care*. 1995;33:783–795.
17. Fishman PA, Shay DK. Development and estimation of a pediatric Chronic Disease Score using automated pharmacy data. *Med Care*. 1999;37:874–883.
18. Fishman PA, Goodman M, Hornbrook M, et al. Risk adjustment using automated pharmacy data: the RxRisk model. *Med Care*. 2003;41:84–99.
19. Johnson RE, Hornbrook MC, Nichols GA. Replicating the Chronic Disease Score (CDS) from automated pharmacy data. *J Clin Epidemiol*. 1994;47:1191–1199.
20. Gilmer T, Kronick R, Fishman P, et al. The Medicaid Rx model: drug-based risk adjustment for public programs. *Med Care*. 2001;39: 1189–1202.
21. Malone DC, Billups SJ, Valuck RJ, et al. Development of a chronic disease indicator score using a veterans affairs medical center medication database. *J Clin Epidemiol*. 1999;52:551–557.
22. Roblin DW. Physician profiling using outpatient pharmacy data as a source for case mix measurement and risk adjustment. *J Ambul Care Manage*. 1998;21:68–84.
23. Sales AE, Liu CF, Sloan KL, et al. Predicting costs of care using a pharmacy based measure risk adjustment in a veteran population. *Med Care*. 2003;41:753–760.
24. Sloan KL, Sales AE, Liu CF, et al. Construction and characteristics of the risk-V: a VA-adapted drug-based case-mix instrument. *Med Care*. 2003;41:761–774.
25. Von Korff M, Wagner EH, Saunders K. A Chronic Disease Score from automated pharmacy data. *J Clin Epidemiol*. 1992;45:197–203.
26. Lamers LM. Pharmacy Costs Groups: a risk-adjuster for capitation payment based on the use of prescribed drugs. *Med Care*. 1999;37:824–830.
27. Coulsen NE, Stuart BC. Persistence in the use of pharmaceuticals by the elderly: evidence from annual claims. *J Health Econ*. 1992;11:315–328.
28. Pauly M, Zeng Y. Adverse selection and the challenges to stand-alone prescription drug insurance. National Bureau of Economic Research working paper 9919, Cambridge, MA: NBER; August 2003.
29. Van Vliet RCJA. Predictability of individual health care expenditures. *J Risk Insur*. 1992;59:443–460.
30. Wouters AV. Disaggregated annual health care expenditures: their predictability and role as predictors. *Health Serv Res*. 1991;26:247–272.
31. Wrobel MV, Doshi J, Stuart BC, et al. Predictability of prescription drug expenditures for Medicare beneficiaries. *Health Care Financ Rev* 2003–2004;25:37–46.

32. Pope GC, Ellis RP, Ash AS, et al. Diagnostic Cost Group Hierarchical Condition Category models for Medicare risk adjustment. Final Report to the Health Care Financing Administration under Contract No. 500-95-048. Waltham, MA: Health Economics Research, Inc.; December 2000.
33. Newhouse JP. Reimbursing health plans and health providers: efficiency in production versus selection, *J Econ Lit*. 1996;34:1236–1263.

APPENDIX

Because our sample contains 874,000 people who occur in both the 1997–1998 and 1998–1999 cohorts, tests of statistical significance for differences in means between the 2 populations in Table 1 were calculated using the following formula, which takes into account the fact that measures are correlated between the 2 samples.

$$t = (\bar{X}_{98} - \bar{X}_{99}) \Big/ \sqrt{\left[\frac{s_{98}^2}{N_{98}} + \frac{s_{99}^2}{N_{99}} + \frac{874000^* s_{98} s_{99} \rho_{98,99}}{N_{98} N_{99}}\right]}$$

where $\bar{X}_i$, s_i, and N_i are the mean, standard deviation, and sample size, respectively, of some variable of interest for year $i = 98$ or 99 and $\rho_{98,99}$ is the correlation coefficient between the 1998 and 1999 samples for people who appear in both years. Because the sample sizes are large and the correlation coefficients are mostly small, tests remain powerful even after this correction. For our regression analysis developing the predictive models, we did not explicitly correct standard errors for this correlation, but instead used a higher significance threshold for deciding which variables to include.