

Medication, Diagnostic, and Cost Information as Predictors of High-Risk Patients in Need of Care Management

Christopher B. Forrest, MD, PhD; Klaus W. Lemke, PhD;
David P. Bodycombe, ScD; and Jonathan P. Weiner, DrPH

Care management programs complement conventional health-care to improve the quality and coordination of services, optimize health, and reduce healthcare costs for individuals with chronic disease, multimorbidity, or other health risks.¹⁻³ These programs generally tier interventions such that patients deemed “low risk” receive usual care or low-intensity services, whereas “high-risk” individuals are given a more intensive package of services.

Segmenting populations by their need for care management requires clinically relevant and statistically valid screening tools. The screener must assign risk scores predictive of future events, identify patients across a continuum of clinically actionable health needs, and be practical to use. The most common data sources for these screening tools—commonly referred to as “predictive models” (PMs)—are healthcare claims and encounter data. Healthcare costs often are used as the outcome of interest for PMs.^{4,6} Cost information is easy to obtain from administrative data and has reasonable predictive accuracy.^{4,6} However, as a simple measure of resource consumption, cost information fails the clinical relevance criterion for a high-risk screening tool that would be useful for care management.

To overcome this limitation of prior cost, PMs based on morbidity, medication use, and other health risk measures have been developed to provide a potentially rich set of clinical predictors.⁷⁻¹² This manuscript contrasts a prior-cost PM with others formed using medication and diagnostic data as potential high-risk case identification screeners for care management programs. The empirical evaluation uses PMs from the Adjusted Clinical Group (ACG) risk adjustment system,¹³ a well-established tool for gaining insight into population morbidity and resource use.^{14,15} This is the first report to describe the development of the ACG predictive models. Although others have contrasted PMs formed with diagnosis versus medication codes,¹¹ clinically based PMs versus those based on prior cost,^{4,6} and combined diagnostic and medication markers in the same PM,^{7,11,16} this article extends this early work by using an approach organized specifically to inform the development of care management programs that seek to use healthcare claims or encounter data to segment a population by health risk.

Objective: To contrast the advantages and limitations of using medication, diagnostic, and cost data to prospectively identify candidates for care management programs.

Methods: Risk scores from prior-cost information and a set of clinically based predictive models (PMs) derived from diagnostic and medication data sources, as well as from a combination of all 3 data sources, were assigned to a national sample of commercially insured, nonelderly adults (n = 2,259,584). Clinical relevance of risk groups and statistical performance using future costs as the outcome were contrasted across the PMs.

Results: Compared with prior cost, diagnostic- and medication-based PMs identified high-risk groups with a higher burden of clinically actionable characteristics. Statistical performance was similar and in some cases better for the clinical PMs compared with prior cost. The best classification accuracy was obtained with a comprehensive model that united diagnostic, medication, and prior-cost risk factors.

Conclusions: Clinically based PMs are a better choice than prior cost alone for programs that seek to identify high-risk groups of patients who are amenable to care management services.

(*Am J Manag Care.* 2009;15(1):41-48)

In this issue
Take-away Points / p47

www.ajmc.com
Full text and PDF

Web exclusive
eAppendix Tables 1 and 2

METHODS

Data Source and Study Sample

Data were obtained from the PharMetrics Patient-Centric Data-

For author information and disclosures,
see end of text.

base (PharMetrics is a division of IMS Health Incorporated). This database represents the medical and pharmacy claims and enrollment records over the continuum of medical care for approximately 85 geographically diverse health plans (Midwest, 35%; Northeast, 21%; South, 31%; West, 13%).^{17,18} The data were from health plans that submit data in exchange for comparative benchmarks.

The initial sample from calendar year 2002 included 5,884,632 nonelderly patients. Individuals without pharmaceutical benefits (613,098, or 10% of the total) or with medical and pharmacy benefits for fewer than 6 months (1,461,471, or 25% of the total) were excluded. The sample was further restricted to enrollees with the same health plan for a minimum of 6 months in both 2001 and 2002, which resulted in a final sample of 2,259,584, randomly divided into a 60% development sample and 40% validation sample.

All diagnosis codes assigned in outpatient or inpatient settings, excluding claims for laboratory and imaging studies, were included as inputs for the PMs. Claims information from outpatient retail pharmacies was used for the medication-based PMs.

We used the health plans' allowed charges for each service (ie, the total amount the plan would pay for the medication or service, inclusive of patient cost sharing) to calculate prior-cost predictors and healthcare charge outcome variables. The total healthcare charge variable was the sum of pharmacy charges, inpatient facility and professional charges, outpatient facility and professional charges, and ancillary services.

Description of Predictive Models

Prior Cost. Although charges are an imperfect measure of actual costs, they are commonly referred to as “prior cost” in the literature and among analysts. We developed a prior-cost PM for total healthcare charges and another for pharmacy charges. Each model included age, sex, and prior-cost quantiles, grouped according to percentile ranking: ≤10, 11-25, 26-50, 51-75, 76-90, 91-93, 94-95, 96-97, and 98-99. Smaller quantiles were used at the top of the distribution to minimize within-group heterogeneity among the highest risk patients.

Dx-PM: A Morbidity-Based Predictive Model. Dx-PM provides a summary measure of disease burden, which is directly related to future healthcare use. (See [eAppendix Table 1](#), available at www.ajmc.com, for a complete list of all covariates in the model.) Furthermore, it provides a risk score that can be decomposed into a set of clinically actionable risk factors. Dx-PM uses age, sex, and diagnosis codes as data inputs. Specific risk factors are derived from the ACG case-mix system, version 7.1.¹³ The model incorporates the ACG total morbidity burden categories and several other categorical disease classification variables, which are described below.

The ACG categories are a morbidity taxonomy¹³⁻¹⁵ that assign individuals to 1 of approximately 100 categories. In Dx-PM, ACGs are organized into a 27-category variable that includes 3 groupings of ACG categories that represent low morbidity levels and 24 individual ACGs that represent the highest levels of medical need.

Expanded diagnostic clusters (EDCs) were used to identify conditions that we hypothesized would have a high impact on future health and resource consumption. EDCs cluster codes from *International Classification of Diseases, Ninth Revision, Clinical Modification* according to clinical similarity. The 98 EDCs used in Dx-PM represent uncommon diseases with high impact on both costs and health, and complications of chronic disease that signify high disease severity.

With research support from the Medicare Agency (now the Centers for Medicare & Medicaid Services), we developed the hosdom and frailty markers.^{19,20} The hosdom variable is a count of different morbidity types (ie, Aggregated Diagnostic Groups from the ACG case-mix system) with at least 1 diagnosis code that is associated with a >50% probability of future hospital admission among patients.¹³ Hosdom uses diagnoses that are “setting neutral” (ie, a person does not have to be hospitalized during the risk assessment period—which is consistent with the care management goal of reducing unnecessary hospitalizations among high-risk persons).

The frailty marker clusters diagnostic codes that indicate presence of at least 1 of 11 frail conditions—malnutrition, dementia and impaired vision, decubitus ulcer, urinary incontinence, loss of weight, fecal incontinence, morbid obesity, poverty, barriers to access to care, difficulty in walking, and falls.¹³ Among commercially insured populations, about 2% of individuals have at least 1 frail condition, and among elders this proportion is twice as high.

Rx-PM: A Medication-Based Predictive Model. Rx-PM is a predictive model constructed from demographic and medication data only. (See [eAppendix Table 2](#), available at www.ajmc.com, for a complete listing of all covariates in the model.) It was designed to produce a medication-defined assessment of morbidity burden that gives a summary measure of risk that can be decomposed into medication-derived drivers. This approach differs from the Chronic Disease Score and its variants, which use medication therapeutic classes to identify a limited set of diseases,^{7,16} and other approaches that simply categorize medications by functional or therapeutic class.¹¹ Rx-PM version 1.0 was used in this study.

Each generic medication/route of administration combination (n = 2800) was assigned to a single Rx-defined morbidity group (MG) (n = 60). Seven Rx-MGs included a single medication therapeutic class, and about one-third of all Rx-MGs included medications from 5 or more therapeutic classes. The

clinical criteria used to form Rx-MGs were primary anatomical-physiologic system, morbidity differentiation (ie, symptom vs disease), expected duration, and severity (ie, impact of the condition on the patient's physiologic stability or functional status). These 4 clinical dimensions not only characterize medications by morbidity type, but they also have major consequences for predictive modeling: higher levels of differentiation, chronicity, and greater severity would all be expected to increase resource use.

Combined Diagnostic/Medication Predictive Models. The combined DxRx-PM model includes all the diagnostic and medication risk factors from Dx-PM and Rx-PM. A secondary model added the prior-cost quantiles. Combining these models was done to examine the added value of each type of risk factor and is justified clinically because the morbidity information obtained from diagnostic codes complements that obtained from medication codes.

Data Analysis

Multivariable linear regression was used to assign risk scores by regressing year 2 charges on the PM risk factors using the development sample. Beta coefficients served as the risk factor weights. We included persons with 6 to 12 months of enrollment in each of 2 consecutive years, all persons included in our database had equal weight in the analysis, and no adjustments were made for nonusers. An individual's PM risk score was the sum of all relevant risk factor weights. The mean of the sample's PM risk scores was centered at 1.0, such that a score of 1.5 indicates future charges 50% higher than expected compared with the sample average, and 0.5 indicates future charges 50% lower than expected compared with the sample average. We estimated PM risk score standard deviations, medians, skewness, and ranges, and used the Spearman rank correlation coefficient to assess risk score correlations.

We hypothesized that the clinical PMs would identify more patients with chronic conditions in the highest risk groups than a prior-cost model. To test this hypothesis, we examined the clinical characteristics (age, sex, and presence of specific chronic conditions) of patients in a top 1% risk group.

Using the same top 1% risk groups, we evaluated group-level measures of medication use, physician ambulatory visits, hospitalization, and mean pharmacy and total charges. For this set of analyses, we hypothesized that the Rx-PM risk groups would have the highest future medication use, whereas the risk group formed with the combined DxRx-PM model with the prior-cost quantiles would have the highest healthcare charges.

We contrasted the PMs on predictive accuracy measures in which pharmacy and total healthcare charges served as the

response variables. The proportion of variation in the dependent variable explained by the predictive models was assessed with the r^2 statistic. To minimize the effect of extreme outliers, we capped the response variables at maximums of \$20,000 and \$50,000 for pharmacy and total charges, respectively. Individuals with expenditures higher than these thresholds were assigned one of these top values. Truncation of expenditures provides more robust and stable estimates than using the raw dollars, and it reflects common re-insurance practices of health plans. Classification accuracy was examined using logistic regression in which the dependent variable was assignment to the top 1% of the population in terms of 2001 (year 2) charges—yes or no. Model fit is presented as the C statistic, which ranges from 0.5 (model no better than the flip of a coin) to 1.0 (perfect true positive and true negative classification).

We did multisample validation using 200 samples of 225,000 randomly drawn cases with replacement from the validation sample. Findings from these bootstrapping analyses showed tight intervals between the 2.5 and 97.5 percentiles, demonstrating good stability of predictive accuracy statistics.

RESULTS

The development and validation data sets had the same distribution of age, sex, geographic residence, chronic conditions, utilization, and healthcare charges (**Table 1**). The prior-cost PM risk scores had 3 times more variability and 6 times more positive skewness than the other models, which is consistent with the well-known distribution properties of healthcare costs in a population. The levels of variability and positive skew among the clinical PMs were similar and less than those of prior cost.

Prior cost was moderately correlated with each of the clinical PMs, and Dx-PM was moderately correlated with Rx-PM (**Table 2**). The correlation between the comprehensive model (DxRx-PM) with and without prior cost was nearly 1.0, suggesting that the addition of prior cost to the clinical markers has minimal impact on rankings by risk score.

Risk groups formed from the clinical PMs were more likely to include patients with chronic disease compared with the prior-cost model (**Table 3**). The mean numbers of chronic conditions was 2.9 for prior cost, 3.6 for Dx-PM, 3.6 for Rx-PM, 3.7 for DxRx-PM, and 3.7 for DxRx-PM with prior cost. ($P < .01$ for all comparisons with prior cost.) The highest proportions of most of the individual chronic diseases were found in the Rx-PM risk group. Addition of the prior-cost variable to the combined DxRx-PM model did not alter the distribution of personal characteristics of that risk group.

■ **Table 1.** Characteristics of Development and Validation Samples

Characteristic	Development (n = 1,355,577) 60%	Validation (n = 904,007) 40%
Baseline (year 1)		
Age, y		
0-17	29.0%	29.0%
18-44	42.4%	42.5%
45-64	28.6%	28.5%
Mean age, y	31.5	31.5
Female	51.2%	51.3%
Geographic residence		
Northeast	33.9%	33.8%
Midwest	28.8%	28.9%
South	20.4%	20.5%
West	16.9%	16.8%
Chronic conditions		
None	68.1%	68.2%
1	17.4%	17.3%
≥2	14.4%	14.5%
Mean number of chronic conditions	0.61	0.60
Resource use (year 2)		
Utilization		
Physician visit	76.2%	76.1%
Mean number of physician visits	3.7	3.7
More than 1 hospitalization	4.5%	4.4%
Mean number of hospital days per 1000 population	254	244
Healthcare charges		
Pharmacy	\$467	\$468
Total (pharmacy + outpatient + inpatient)	\$2024	\$2026

Compared with prior cost and the other clinical models, the Rx-PM identified a high-risk group that used more medications (Table 4). The mean numbers of medications per person in the risk groups were 11.9 for prior cost, 9.6 for Dx-PM, 13.7 for Rx-PM, 11.2 for DxRx-PM, and 11.5 for DxRx-PM with prior cost. ($P < .01$ for all comparisons with prior cost.) Thus, the Rx-PM risk group used 15.1% more medications than the prior-cost risk group in the year after the risk assessment period.

The alternative PMs did not differ substantively with respect to future use of physician visits, each identifying a risk group that had a 95% to 97% chance of making a physician visit with a mean that ranged from 12 to 14 visits per annum (Table 4). Regarding hospital utilization, however, the mod-

els showed more distinct differences. The DxRx-PM model with prior-cost risk group had the highest probability of hospitalization (30.3%) and the largest number of hospital days per 1000, both of which were significantly higher than the values observed with the prior-cost model. Furthermore, the DxRx-PM with prior-cost risk group had 58.9% more hospital days than the Rx-PM risk group.

The DxRx-PM with prior-cost risk group had the highest healthcare charges among all the PM risk groups (Table 4). However, the Rx-PM risk group had the highest pharmacy charges, which were 42.6% higher than those with the prior-cost model.

Across all the PMs, the DxRx-PM with prior-cost model explained the most variance (Table 5) in pharmacy charges (59%) and total charges (26%). The DxRx-PM (without prior cost) model explained more variance in total healthcare charges than prior cost. Classification accuracy measured using the area under the receiver operating characteristic curve was comparable across all the PMs.

DISCUSSION

The purpose of this study was to conduct empirical assessments of alternative PMs that identify high-risk patients to make recommendations on

what type of model best suits the needs of a given care management program. Compared with prior cost alone, the clinically based PMs derived from demographic, diagnostic, and medication data identify high-risk groups with more chronic disease, greater medication usage and polypharmacy, and higher healthcare costs. Another study using Dx-PM among elders found that it identified a high-risk group that had a greater burden of chronic disease, more functional limitations, and poorer perceived health than the lower-risk counterparts.²¹

Clinically based PMs are likely to provide greater yield per resources expended in care management programs than prior cost alone, because they are more accurate at identifying high-cost risk groups and selecting high-risk individuals who are more amenable to clinical interventions. Although

Predictors of High-Risk Patients

■ **Table 2.** Spearman's Rank Correlation Matrix for Risk Scores Produced by Alternative Predictive Models

Predictive Model ^a	Prior Cost	Dx-PM	Rx-PM	DxRx-PM	DxRx-PM With Prior Cost
Prior cost	—	—	—	—	—
Dx-PM	0.57	—	—	—	—
Rx-PM	0.42	0.61	—	—	—
DxRx-PM	0.59	0.93	0.82	—	—
DxRx-PM with prior cost	0.65	0.91	0.80	0.97	—

^aData inputs for these models were: Dx-PM—age, sex, and *International Classification of Diseases* codes; Rx-PM—age, sex, and National Drug Code codes; DxRx-PM—age, sex, *International Classification of Diseases* codes, and National Drug Code codes; DxRx-PM—age, sex, *International Classification of Diseases* codes, National Drug Code codes, and healthcare charges.

■ **Table 3.** Clinical Characteristics of Top 1% Risk Groups Formed by Alternative Predictive Models^a

Characteristic	Top 1% Risk Group ^b					
	Total Sample	Prior Cost	Dx-PM	Rx-PM	DxRx-PM	DxRx-PM With Prior Cost
Mean age, y	31.5	43.6	45.6 ^c	49.7 ^c	46.4 ^c	46.2 ^c
Female	51.2	52.9	53.7	60.9 ^c	55.4 ^c	56.1 ^c
Any chronic condition	31.8	84.2	91.0 ^c	91.2 ^c	92.6 ^c	93.1 ^c
Diabetes	3.0	15.2	23.3 ^c	29.1 ^c	23.7 ^c	23.4 ^c
Asthma	3.8	7.8	9.0 ^c	10.1 ^c	9.5 ^c	9.5 ^c
Ischemic heart disease	1.6	15.9	15.1	17.8 ^c	15.4	15.5
COPD	0.9	5.5	7.7 ^c	7.6 ^c	7.8 ^c	7.8 ^c
CHF	0.3	4.8	7.6 ^c	7.4 ^c	7.8 ^c	7.2 ^c
Osteoarthritis	2.1	9.5	9.0	11.1 ^c	9.4	9.8
Low back pain	8.4	21.7	19.3 ^c	22.9 ^c	21.1	21.5
Chronic renal failure	0.1	4.2	8.3 ^c	5.4 ^c	8.0 ^c	7.5 ^c
Cancer	0.6	2.5	3.0 ^c	2.5	3.0 ^c	3.1 ^c

CHF indicates congestive heart failure; COPD, chronic obstructive pulmonary disease.

^aClinical data are from year 2 and predictive modeling risk factors are from year 1.

^bValues are percentages unless otherwise indicated.

^cComparison with prior-cost mean or proportion was significantly different at $P < .01$.

beneficial effects of care management have not been consistently demonstrated,²² preliminary evidence from an intensive nurse-based intervention for high-risk elders appears to show great promise in terms of cost reduction²³ and better quality of care.²⁴

Multimorbidity, another clinically actionable characteristic, was extremely common in the high-risk groups identified. The mean number of chronic conditions among individuals in the top 1% high-risk groups formed by the clinical PMs ranged from 3 to 4. Compared with the general population, individuals in the high-risk groups had approximately 6 times the burden of chronic disease. This level of multimorbidity suggests that the most effective care management programs will need to address service provision for multiple conditions

simultaneously, focusing on care for an individual rather than a single disease.^{1,22}

Limitations

The Centers for Medicare & Medicaid Services has several ongoing pilot projects evaluating the impact of care management for Medicare beneficiaries. Our analyses did not include individuals older than age 65 years, although other work using Dx-PM has focused specifically on this age group.^{21,23,24} The direction and relative magnitude of effects that we observed are likely to apply to older populations, which have higher burdens of chronic disease than the working population in our sample.

The PharMetrics database we used for these analyses has had limited use by researchers. Recognizing this limitation,

■ **Table 4.** Resource Use for Top 1% Risk Groups Formed by Alternative Predictive Models^a

Resource Use Measure	Top 1% Risk Groups				
	Prior Cost	Dx-PM	Rx-PM	DxRx-PM	DxRx-PM With Prior Cost
Medication use, %					
Any medication	93.9	88.6 ^b	98.8 ^b	93.0	93.7
≥10 unique medications	56.7	43.3 ^b	68.4 ^b	52.9 ^b	54.4 ^b
Physician ambulatory visits					
Any visit, %	95.1	96.1 ^b	96.3 ^b	96.6 ^b	97.0 ^b
Mean number of visits	12.2	12.7	12.6	13.4 ^b	13.7 ^b
Hospitalization					
≥1 hospitalizations, %	27.3	28.4	25.1 ^b	29.0	30.3 ^b
Mean number of hospital days per 1000	475	492	316 ^b	486	502
Healthcare charges, mean					
Pharmacy	\$4035	\$3798	\$5753 ^b	\$4792 ^b	\$5134 ^b
Total	\$22,124	\$21,452	\$18,726 ^b	\$22,797	\$23,824 ^b

^aResource use data are from year 2 and predictive modeling risk factors are from year 1.
^bThe reference category for statistical comparisons was prior cost. Statistical differences in proportions were assessed using logistic regression, and the differences in means used multivariable linear regression. A critical value of .01 was used to indicate significance.

■ **Table 5.** Predictive Accuracy of Alternative Predictive Models: Pharmacy and Total Healthcare Costs

Predictive Accuracy Measure	Alternative Predictive Models				
	Prior Cost	Dx-PM	Rx-PM	DxRx-PM	DxRx-PM With Prior Cost
Variance explained, R²					
Pharmacy charges (capped at \$20,000)	54%	29%	45%	47%	59%
Total healthcare charges (capped at \$50,000)	21%	21%	20%	25%	26%
Area under ROC curve (top 1% year 2 cost group = outcome)					
Pharmacy charges	.94	.91	.94	.95	.96
Total healthcare charges	.86	.84	.8	.84	.85

ROC indicates receiver operating characteristic.

we performed extensive evaluations of data quality (ie, completeness, valid data ranges, known group differences). These analyses supported the validity of the data source. Furthermore, the predictive accuracy statistics that we found are similar in magnitude to those reported in objective evaluations of alternative PMs.¹²

Over the last 2 decades there has been an increasing reliance on diagnosis information from claims submitted for hospital and physician payment. Prior research²⁵⁻²⁷ has indicated that the specificity of diagnoses found in claims is good, even if the sensitivity is moderate. In other words, not all persons with a disease are detected by diagnoses recorded in claims; however, when a diagnosis is present, the chances that the

person truly has the disorder are high. More accurate capture of information about patients' disorders from electronic health records will improve the classification accuracy of diagnostic information, which will improve the predictive performance of PMs. Additional data types from electronic health records such as disease severity, laboratory values, imaging results, social history, and genomics will usher in a new generation of PMs that will further refine the ability of care managers and health professionals to identify proactively those patients with the greatest healthcare needs.

Validation analyses have been performed with pharmacy insurance claims as well. The degree of accuracy of the types of drugs reimbursed by the insurance system is believed to be

high based on comparisons to date,^{28,29} although actual adherence to medication regimens cannot be assessed with retail pharmacy medication claims.

Implications for Care Management Programs

The Rx-PM, which is based on the National Drug Code (NDC), defined a risk group that was associated with the highest rates of chronic disease, more medication use, and more polypharmacy than either the prior-cost DM or the Dx-PM. Model performance was very strong for pharmacy costs but weaker than the alternative models for total healthcare costs, which is a pattern found by others.¹¹ NDC-based models are particularly well suited for healthcare organizations (eg, pharmacy benefit management firms) or specific programs that target care management of individuals' medication usage.

Another advantage of NDC-based models is that pharmacy data accrue rapidly and are available shortly after patients enroll in a plan, whereas 6 to 12 months may be needed before sufficient information is available from diagnostic codes to assign risk scores. A medication-only model can be used for care management interventions among these new enrollees. Once a medical interview is done, either by a health professional or online with a health risk appraisal, and the full list of medications is obtained, an NDC-derived risk score can be calculated and used to enroll individuals in care management programs. All that is needed to assign a risk score using Rx-PM is the list of generic medication names along with their route of administration; dosage and administration frequency are not needed.

Compared with prior cost, a diagnosis-based model (Dx-PM) identified patients with a higher burden of chronic disease and similar rates of hospitalization, and has comparable predictive accuracy for total healthcare costs, findings supported by other investigators.⁶ When the target of care management is use of medical services, particularly reductions in the rates of hospitalization, a diagnosis-based model such as Dx-PM appears to be superior to prior cost.

Other investigators have combined diagnostic and medication code markers within the same predictive model, and their findings are consistent with those we report in this article. The addition of medication data to diagnostic data has a small to negligible impact on predictive performance for total healthcare costs^{7,11} and healthcare utilization^{7,16}; however, adding diagnostic markers to medication data improves prediction of total resource use.^{7,11,16} Adding diagnostic markers to medication data does not appear to improve predictions

Take-away Points

Care management is used to improve the quality and outcomes of care, while reducing costs for individuals with complex health needs. The best screening tools for selecting candidates for these programs remain unclear.

- Prior cost is commonly used and is a good predictor of future cost.
- Screening tools based on diagnostic or medication data are just as accurate as prior cost, while identifying groups that are more amenable to clinical interventions.
- Clinically based "predictive models" are a superior choice to prior-cost information alone as a screening tool for case management.

for pharmacy costs,¹¹ which tend to be a more stable set of expenditures.

These empirical and conceptual differences provide a rationale for combining the predictors from each model into a comprehensive PM. The correlation of nearly 1.0 between DxRx-PM (diagnostic and medication risk factors only) and DxRx-PM with prior cost, coupled with the superior predictive performance of the latter, argues for use of the DxRx-PM with prior cost when a comprehensive model is desired for high-risk case identification. The primary disadvantage of this comprehensive model is the relatively high level of data processing required. This disadvantage is counterbalanced by the model's superior predictive performance and its inclusion of a rich set of clinically actionable predictors.

The predictive models presented in this article are a first step toward mining existing clinical information to forecast future health and resource consumption. Ultimately, the best high-risk screening tools will be those that identify prodromal conditions that can be targeted for intervention before a morbidity event occurs or flares to high severity, thereby forgoing the need for costly interventions and continuing care.

Author Affiliations: From the Department of Pediatrics (CBF), Children's Hospital of Philadelphia, Philadelphia, PA; and the Department of Health Policy and Management (KWL, DPB, JPW), Johns Hopkins University, Baltimore, MD.

Funding Source: The ACG System is commercially distributed in North America by DST Health Solutions, Inc, under a license with The Johns Hopkins University, which holds the copyright to the ACG System, including the predictive models used in this research. Johns Hopkins University benefits financially from the sale of this software. A portion of these royalties is used toward supporting ongoing development work of the system, including the research presented within this article.

Author Disclosure: Drs Forrest, Lemke, Bodycombe, and Weiner receive an inventors' share of the royalties from the ACG software. The authors also benefit from free access to the ACG System software.

Authorship Information: Concept and design (CBF, DPB, JPW); acquisition of data (CBF, KWL, DPB, JPW); analysis and interpretation of data (CBF, KWL, DPB, JPW); drafting of the manuscript (CBF, KWL, DPB); critical revision of the manuscript for important intellectual content (CBF, DPB, JPW); statistical analysis (CBF, KWL, DPB); provision of study materials or patients (CBF, KWL); obtaining funding (CBF, JPW); administrative, technical, or logistic support (CBF, DPB, JPW); and supervision (CBF, JPW).

Address correspondence to: Christopher B. Forrest, MD, PhD, Department of Pediatrics, Children's Hospital of Philadelphia, 3400 Civic Center Blvd, ABR 1335, Philadelphia, PA 19104. E-mail: forrestc@email.chop.edu.

REFERENCES

1. **Boyd CM, Boulton C, Shadmi E, et al.** Guided care for multimorbid older adults. *Gerontologist*. 2007;47(5):697-704.
2. **Counsell SR, Callahan CM, Clark DO, et al.** Geriatric care management for low-income seniors: a randomized controlled trial. *JAMA*. 2007;298(22):2623-2633.
3. **Morisky DE, Kominski GF, Afifi AA, Kotlerman JB.** The effects of a disease management program on self-reported health behaviors and health outcomes: evidence from the "Florida: A Healthy State (FAHS)" Medicaid program. *Health Educ Behav*. February 21, 2008. Epub ahead of print.
4. **Meenan RT, Goodman MJ, Fishman PA, Hornbrook MC, O'Keefe-Rosetti MC, Bachman DJ.** Using risk-adjustment models to identify high-cost risk. *Med Care*. 2003;41(11):1301-1312.
5. **Pietz K, Ashton CM, McDonnell MJ, Wray NP.** Predicting healthcare costs in a population of Veterans Affairs beneficiaries using diagnosis-based risk adjustment and self-reported health status. *Med Care*. 2004;42(10):1027-1035.
6. **Zhao Y, Ash AS, Haughton J, McMillan B.** Identifying future high-cost cases through predictive modeling. *Dis Manag Health Outcomes*. 2003;11(6):389-397.
7. **Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE.** A chronic disease score with empirically derived weights. *Med Care*. 1995;33(8):783-795.
8. **Fishman PA, Goodman MJ, Hornbrook MC, et al.** Risk adjustment using automated ambulatory pharmacy data: the RxRisk model. *Med Care*. 2003;41(1):84-99.
9. **Kuhlthau K, Ferris TG, Davis RB, Perrin JM, Iezzoni LI.** Pharmacy- and diagnosis-based risk adjustment for children with Medicaid. *Med Care*. 2005;43(11):1155-1159.
10. **Powers CA, Meyer CM, Roebuck MC, Vaziri B.** Predictive modeling of total healthcare costs using pharmacy claims data: a comparison of alternative econometric cost modeling techniques. *Med Care*. 2005;43(11):1065-1072.
11. **Zhao Y, Ash AS, Ellis RP, Ayanian JZ, Pope GC, Bowen B, Weyuker L.** Predicting pharmacy costs and other medical costs using diagnoses and drug claims. *Med Care*. 2005;43(1):34-43.
12. **Winkelman R, Meymud S.** A comparative analysis of claims-based tools for health risk assessment. Society of Actuaries sponsored research project, April 20, 2007. <http://www.soa.org/research/health/hlth-risk-assessment.aspx>. Accessed February 27, 2008.
13. **The Johns Hopkins University ACG Case-Mix System.** <http://acg.jhsph.edu>. Accessed February 27, 2008.
14. **Starfield B, Weiner JP, Mumford L, Steinwachs D.** Ambulatory care groups: a categorization of diagnoses for research and management. *Health Serv Res*. 1991;26(1):53-74.
15. **Weiner JP, Starfield B, Steinwachs DM, Mumford LM.** Development and application of a population-oriented measure of ambulatory care case-mix. *Med Care*. 1991;29(5):452-472.
16. **Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ.** Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol*. 2001;154(9):854-864.
17. **Berger A, Dukes E, Martin S, Edelsberg J, Oster G.** Characteristics and healthcare costs of patients with fibromyalgia syndrome. *Int J Clin Pract*. 2007;61(9):1498-1508.
18. **International Society for Pharmacoeconomics and Outcomes Research.** PharMetrics Patient-Centric Database. <http://www.ispor.org/DigestOfIntDB/Default.aspx?rcd=426>. Accessed February 27, 2008.
19. **Weiner JP, Dobson A, Maxwell SL, et al.** Risk adjusted Medicare capitation using ambulatory and inpatient diagnoses. *Health Care Financ Rev*. 1996;17(3):77-99.
20. **Weiner JP.** Updating and calibrating the Johns Hopkins ACG risk adjustment methods for application to Medicare risk contracting. Johns Hopkins University research contract #500-98-0002 to the Health Care Financing Administration, February 2000.
21. **Sylvia ML, Shadmi E, Hsiao CJ, Boyd CM, Schuster AB, Boulton C.** Clinical features of high-risk older persons identified by predictive modeling. *Dis Manag*. 2006;9(1):56-62.
22. **Luck J, Parkerton P, Hagigi F.** What is the business case for improving care for patients with complex conditions? *J Gen Intern Med*. 2007;22(suppl 3):396-402.
23. **Sylvia ML, Griswold M, Dunbar L, Boyd CM, Park M, Boulton C.** Guide care: cost and utilization outcomes in a pilot study. *Dis Manag*. 2008;11(1):29-36.
24. **Boyd CM, Shadmi E, Conwell LA, et al.** A pilot test of the effect of guided care on the quality of primary care experiences for multimorbid older adults. *J Gen Intern Med*. 2008;23(5):536-542.
25. **Maclean JR, Fick DM, Hoffman WK, et al.** Comparison of 2 years for clinical practice profiling in diabetic care: medical records versus claims and administrative data. *Am J Manag Care*. 2002;8(2):175-179.
26. **Steinwachs DM, Stuart ME, Scholle S, et al.** A comparison of ambulatory Medicaid claims to medical records: a reliability assessment. *Am J Med Qual*. 1998;13(2):63-69.
27. **Fowles JB, Lawthers AG, Weiner JP, et al.** Agreement between physicians' office records and Medicare Part B claims data. *Health Care Financ Rev*. 1995;16(4):189-199.
28. **Tamblyn R, Lavoie G, Petrella L, Monette J.** The use of prescription claims databases in pharmacoepidemiological research: the accuracy and comprehensiveness of the prescription claims database in Quebec. *J Clin Epidemiol*. 1995;48(8):999-1009.
29. **Levy AR, O'Brien BJ, Sellors C, et al.** Coding accuracy of administrative drug claims in the Ontario drug benefit database. *Can J Clin Pharmacol*. 2003;10(2):67-71. ■