Health Care Management Science 5, 211–219, 2002

Disease Burden Profiles: An Emerging Tool for Managing Managed Care

YANG ZHAO *
DxCG, Inc., 25 Kingston Street, Suite 200, Boston, MA 02111, USA
E-mail: yang.zhao@dxcg.com

ARLENE S. ASH
Department of General Internal Medicine, Boston University, and DxCG, Inc., Boston, MA, USA

RANDALL P. ELLIS
Department of Economics, Boston University, and DxCG, Inc., Boston, MA, USA

JAMES P. SLAUGHTER
Medical Records Group, Northern California Kaiser Permanente, San Rafael, CA, USA

Abstract. As health plans assume financial risk for providing health care services, effectively managing the health of a population remains one of the toughest challenges. This article shows how risk assessment methods can be used to measure disease burden in the full population and to discriminate levels of future health care needs within specific disease cohorts. We also examine and compare the predictive power of claims-based models within a diabetic cohort.

Keywords: illness burden, disease profile, health management, diagnostic cost group, risk adjustment, disease management

1. Introduction

Effective management of health care for a population requires knowing the prevalence and distribution of its medical problems, which can vary dramatically by population. This is obvious at the very broad level of, say, comparing the illnesses of a pediatric population with those of Medicare beneficiaries, but is often true even across populations with very similar age/sex distributions; the illness burden of a given population is key in predicting its subsequent treatment costs.

Understanding the prevalence of diseases in a population and the associated needs for services facilitates rational planning. For example, when a Managed Care Organization (MCO) bids for a new block of business, it must estimate the cost and resources for providing for the medical problems of the new enrollees. Further, as it "staffs up" to accommodate the new population, it needs more refined estimates of the likely demand for specific specialist services, such as diabetic or cancer care.

Suppose, for example, that an established MCO that serves a population of 1,000,000 privately insured employees and their dependents considers submitting a bid to the State for the right to deliver care to a group of 100,000 Medicaid recipients. Even though the two populations may have very similar mean costs, a simple 10% expansion in costs and capacity is clearly too crude. Comparing the prevalence of medical problems between the new Medicaid group and existing population of insured may reveal important medical management issues that the State and the MCO should consider. For example, quadriplegia, paraplegia, and other neurological disorders are substantially more prevalent in Medicaid than in commercial populations; also, although mental retardation is vanishingly rare in commercially insured populations, many hundreds of the Medicaid recipients are likely to have such problems.

This article illustrates the usefulness of a population-based classification system for characterizing a population's disease burden as a first step in population-based health management. We illustrate the value of identifying each person's range of distinct medical problems encountered over a fixed time period, using the Diagnostic Cost Group (DCG) methodology for classification [1]. The methodology uses diagnoses recorded during inpatient and ambulatory care from multiple sites. It aggregates these diagnoses into approximately 100 condition categories (CCs) as its basic "building blocks" and supports rapid identification of groups of people who satisfy any of tens of thousands of distinct disease profiles. Such profiles can be as simple as "people with diabetes", or arbitrarily complex, such as, "people with acute complications of diabetes and congestive heart failure but without renal insufficiency".

Despite the extensive bibliography using the DCG methodology to predict costs for risk adjustment, this is the first publication to explore its uses in medical population profiling and for identifying people for case management. We use DCGs to profile and compare a privately insured and a Medicaid population and present their differences as to demographics, dis-

* Corresponding author.

ease profiles, and resource utilization. Similar analyses could be undertaken for comparing policy-relevant subgroups of commercially insured populations, such as employer groups, unions, and regional delivery systems.

In addition to disease prevalence, plans and payers need to consider the nature and cost of care that enrollees with specific problems receive. Are costs for people with a specific disease (such as diabetes mellitus) higher for certain groups because people in them are older and/or have more comorbidities? Or, are different pricing structures, or different utilization patterns, the explanation? We show how disease profiling can help explore these issues.

Finally, we quantify the capacity of DCG models to identify a subset of people with a given chronic condition whose future costs will be particularly high. We contrast DCG model performance with a traditional prior-cost model used by actuaries and others. Predicting future high cost cases is important for disease and catastrophic case management programs.

We choose diabetes in our case study because it is an extremely common and costly chronic condition, affecting nearly 16 million (5.9%) of the United States population [2–5]. In 1997, the total health expenditures incurred by people with diabetes (including costs not resulting from diabetes) approached $80 billion [2]. Because diabetes is common, expensive, and can require complex management, it is often the target of disease management strategies.

2. Methods

2.1. Study populations and data sources

The study populations were from two sources: a sample of 493,000 individuals covered by a state's Medicaid program in 1996–1997 (Medicaid); and 2 million individuals enrolled during 1996 and 1997 in various commercially insured, nationally-disbursed plans (Private). We retained all people with any year-1 and year-2 enrollment for these analyses, because we believe that full-population analyses (including partial year enrollees) are most useful and informative for population profiling and prediction. Partial year enrollees include those who die, a particularly expensive and important subgroup in any population, and recent entrants (who, in Medicaid, often enroll because of illness). Recent entrants provide useful, if incomplete, data for planning and prediction. For example, a man treated for several chronic ailments last year would be a good candidate for case management this year, whether he was previously enrolled for only one month or for the entire year. In this instance, including partial year enrollees helps health care managers identify population health risks, which is extremely valuable when managing relatively transient populations, such as Medicaid. For example, 45% of our Medicaid sample would have been excluded from the analysis if partial year enrollees were excluded.

In our Medicaid sample, 21% of enrollees were present because of poverty, while 45% were present because of medical reasons other than blindness/disability. The Private sample was from Medstat's MarketScan® Research Database, which contains inpatient and outpatient healthcare claims for individuals covered by the benefit plans of several large employers. More than 100 payers contribute data each year. 17% of enrollees in our commercially insured sample were covered by fully or partially capitated health plans with the balance in various fee-for-service arrangements. Since we did not have Medicare claims, we included only people who were younger than 65 at the end of year 1 in both samples.

2.2. Data analysis

Our data sources contained substantial information. The average number of diagnoses per outpatient claim was 1.5 in both the Private and Medicaid samples; the average number of diagnoses per person was 8 in the Private sample and 23 in Medicaid; the average number of claims per person in the two populations were 6 and 15, respectively.

Our outcome variable in the private sample was total covered expenses, excluding outpatient pharmacy (because pharmacy benefits varied widely across plans). We did not remove deductibles, copays, and coinsurance amounts. In Medicaid, we used the total Medicaid paid amount excluding pharmacy costs, for consistency with the private sample. Since Medicaid providers are prohibited from balance billing, the paid amount is equal to the Medicaid covered amount.

For people with partial year eligibility, the costs that we captured reflected only the medical expenses incurred when these people were covered. To make the cost variable comparable across individuals, we annualized and weighted cases as follows: a person's actual expenditure during the year was divided by the fraction of the year that he or she was present to produce the annualized amount. For example, a man who generated $6,000 during 6 months of plan enrollment had an "annualized" payment of $12,000. However, simply using annualized amounts in calculating means overstated the contribution of partial-year enrollees. Thus, we used the fraction of the year that each individual was eligible for coverage as a weight when examining annualized costs. The man in the above example, therefore, had a weight of 0.5. This process of annualizing and weighting observations resulted in unbiased estimates of average and total payments for groups with people who were eligible only part of the year.

2.3. Risk-adjustment method

The tool we used to describe illness burden in this paper is the DCG model as described in [1]. It uses all available diagnoses from physicians and hospitals as recorded in International Classification Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes in year 1, and classifies each diagnosis into one of 118 Condition Categories (CCs). Each CC contains ICD-9-CM codes that define similar clinical problems with similar expected costs. For example, there are 10 CCs for various types of Heart problems. CCs are not mutually exclusive; individuals can have multiple CCs. A person with no year-1 medical encounters has no CCs. Hierarchical

Conditions Categories (HCCs) are generated from CCs by imposing hierarchies, which identify only the most costly manifestation of each distinct disease. For example, a Metastatic Cancer (CC 5) patient may also have Skin Cancer (CC 11) and Benign Cancer (CC 12). However, when the hierarchies are applied, this individual is classified only in Metastatic Cancer. Either CCs or HCCs can be used to create population profiles, but only HCCs are used in predicting healthcare resource utilization.

For population-based analyses of illness burden, several approaches are available, such as the Adjusted Clinical Group (ACG) system [6] and Clinical Risk Groups (CRGs) [7]. Both ACGs and CRGs differ from DCGs in that they are "clinical groupers", classifying each person into exactly one of a predetermined number of clinical categories. In contrast, the DCG system constructs an entire clinical profile for each person. For example, a 50-year-old person with diabetes, a heart attack and asthma is classified in CCs 13 (Diabetes with Chronic Complication), 48 (Congestive Heart Failure), and 70 (Asthma). In the ACG system, the same three diseases cause the individual to be classified into ACG 4100 (2–3 other ADG combinations, age > 34), a category that also includes, for example, a 35-year-old person with Obesity and Myopia.

Starting by organizing all diagnoses in year 1 in a comprehensive disease profile, the DCG model predicts year 2 resource utilization by adopting an "accumulating" perspective; it assumes that a person with diseases A, B, and C will use at least as many resources as a person with problems A and B only. The simplest accumulating model is additive; the joint effect on expected cost of diseases A and B is the sum of the effects associated with each disease separately. Although reality is typically a bit more complicated than this, empirically we have found that additive models tend to fit the data well [1]. Caregivers and health plan managers can readily follow the additive approach of how each of a person's individual problems is combined in calculating his or her expected costs.

2.4. Creating population profiles

Our focus was on using the CCs to profile populations based on the range of medical problems that occur during a fixed period. We created the disease profile for a population by calculating the prevalence of CCs as rates per 10,000 persons. Since the CCs are not mutually exclusive, single individuals can appear in multiple categories. Because person–year was the analytic unit, we computed the mean total costs for the people in each CC. We did not break out costs by diagnosis; the cost associated with a CC was the total health care expenditure (for all their medical problems) for persons who had the condition that defined the CC. We further computed the relative cost of the people who had the medical problem that defined a CC by dividing their mean cost by the population mean cost. This indicates how much more expensive the care for people with this medical condition is than for the population average. We also described how much of the total cost incurred by the entire population was contributed by people with each specific CC.

2.5. Targeting care within a diabetic cohort

To illustrate the population-based paradigm in designing disease management programs, we first identified all people with diabetes in year 1 from a random sample of 300,000 people selected from the Private data and all 15,049 cases in the Medicaid sample.[1] In each diabetic cohort we calculated the prevalence of other CCs as rates per 10,000 to identify medical problems that were particularly common. We also calculated the excess prevalence of these problems as the ratio of their prevalence for people with diabetes divided by the prevalence for people in the same sample who did not have diabetes. We further tested whether the excess prevalence of each CC within each diabetic cohort was significant using Chi-squared statistics.

To illustrate the potential for disease management stratification in the private diabetic cohort, we used claims data to classify people in three ways: (1) focusing on Diabetes-Specific Severity (DSS), (2) using DCG expected cost categories (DCG), and (3) using categories defined by levels of year-1 total cost (Prior cost).

Although the specific algorithms used by Disease Management (DM) companies for diabetes are generally proprietary and likely depend on the particular data sources available, they are typically based on the diabetes care and management studies published by the American Diabetes Association (ADA) [2,3,8–12]. These studies identify diagnostic and risk factors among people with diabetes, and provide quality-of-care indicators for clinicians for care management. For example, the ADA standards of medical care for patients with diabetes mellitus emphasize the higher risk of diabetes patients with diabetic ketoacidosis and hyperosmolar hyperglycemic nonketotic syndrome, hypertension, cardiovascular disease, atherosclerotic vascular disease, diabetic neuropathy, and problems involving the feet [8]. However, no attention is drawn to the presence of septicemia. The diabetes-specific severity-based (DSS) model used in this paper was based on a review of previous studies and focuses on factors that clinicians found important in classifying diabetic patients (table 1). Since we had only administrative databases, such useful information, as "living alone" or "an HbA1C blood test level exceeding 11%", was not available in these (or most) administrative data sets, and was therefore not used in our analysis. This DSS, administrative-data-based methodology classified people into three mutually exclusive groups: stable, at risk, and high risk. As shown in table 1, information used to classify people includes the number of hospitalizations or ER visits, and the appearance of ICD-9-CM codes indicating vascular disorder, stroke, or amputation.

In the second method, we used DCG-benchmarked model predictions to classify people with diabetes into four mutually exclusive expected cost categories: \$1,000–\$4,999, \$5,000–\$9,999, \$10,000–\$24,999, and \$25,000+. Note that no one

[1] That is, we selected all people with a diagnosis in any of CCs 13 Diabetes with Chronic Complications, 14 Diabetes with Acute Complications/Nonproliferative, and 15 Diabetes with No or Unspecified Complications.

Table 1
Disease severity-based classification for diabetic cohort.[a]

Stable	At risk	High risk
Dx of diabetes	>1 hospitalization for non-diabetic medical conditions, or >1 ER visit for non-diabetic medical conditions, or active vascular disorders, CVA, TIA, PVD, or >74 years old	>1 hospitalization for diabetes related medical conditions, or >1 ER visit for diabetes related medical conditions, or hospitalization for Vascular disorders, CVA, TIA, PVD, or 1 or more episodes of hospitalization for DKA or HHNK, or amputation

[a] Diabetes = ICD-9-CM codes 250.xx, 251.0. CVA = cerebrovascular accident (ICD-9-CM codes 434.xx, 436, 437, 437.0, 437.1, 437.3–437.6, 437.8, 437.9). TIA = transient ischemic attack (ICD-9-CM codes 435.xx). PVD = peripheral vascular disease (ICD-9-CM codes 443.xx). DKA = diabetic ketoacidosis (ICD-9-CM code 250.1). HHNK = hyperglycemic hyperosmolar nonketotic (ICD-9-CM code 250.2). Amputation = ICD-9-CM codes 895, 896, 897.

with diabetes is predicted to cost less than $1,000. In general, sorting of the population in terms of next year's expected costs does not require any dollar figures; the DCG predicted relative risk scores (calculated from the DCG benchmark population) suffice.

Prior cost is another "population-based" measure that has been used in risk assessment to set capitation rates [13,14] and to identify future high-cost users [15]. Variations of prior-cost models are widely used by actuaries and underwriters [16]. Our third way of classifying the diabetic cohort sorted people by levels of year-1 or "prior" costs. To make a clear comparison with the DCG methodology, we divided the population into four groups, based upon increasing levels of prior cost, with the same number of people as the DCG four categories of increasing predicted cost described above. We labeled these groups Moderate, Medium, High and Very High (Prior Cost).

Finally, we compared each method's ability to predict year-2 costs in this diabetic cohort using R^2 values. We first calculated R^2s viewing each of the categorical partitions (DSS, DCG, and prior cost) as a prediction model; we also used each of the two continuous predictors (DCG score and prior cost) to divide the diabetic cohort into 50 2-percentile groups and calculate the R^2 for each partition.

3. Results

3.1. Comparisons across two populations

Table 2 contrasts the distribution of annual health care costs in the two data sets. On average, individuals covered by the Medicaid program cost 16% more in year 2 ($1,965 vs. $1,700) and these costs vary less (coefficients of variation, that is CVs, are 280 and 438) than in the Private sample. The Medicaid sample has twice as many minors (age < 18) as the Private data (55 vs. 26%), and fewer people over age 44 (14 vs. 35%), leading to a 13-year difference in mean age (21 vs. 34). Despite the younger age profile, 81% of the Medicaid enrollees have positive expenditures, as compared to only 74% in the Private sample; the hospitalization rate is also much higher in Medicaid (11 vs. 5%).

Both prevalence and distribution of illness differ dramatically across populations. Table 3 illustrates this pattern for a selection of 47 Condition Categories (CCs) chosen from the 118 used in the DCG system. The selected conditions comprise nine clusters of disease categories, most of which are high-cost and chronic. The first two columns of numbers display prevalence rates per 10,000 persons; the next two columns show the relative annual cost in year 2 of persons falling in each condition group (calibrated so that for each sample the average cost weight is one); the last two columns indicate the percentage of year-2 costs for each whole population that is contributed by people with each specified medical problem.

Most diseases are more prevalent in Medicaid than in the Private population, although diseases of older people (most heart disease, cancers, and other neoplasms) are less common in the much younger Medicaid cohort. Excess prevalence in the Medicaid database is extraordinarily high for Septicemia (Blood Poisoning)/Shock (CC 2), Mental Retardation (CCs 36 through 39), Quadriplegia (CC 40), Paraplegia (CC 41),

Table 2
Costs, demographics, and utilization experience in two populations.[a]

	Medicaid	Private
N	493,238	1,975,759
Years	1996–1997	1996–1997
Mean months of eligibility		
year 1	10.3	11.1
year 2	10.2	11.2
Year-1 cost ($)		
mean	$1,671	$1,613
standard deviation	$6,708	$7,254
coefficients of variation (×100)	401	450
median	$252	$228
Year-2 cost ($)		
mean	$1,965	$1,700
standard deviation	$5,501	$7,440
coefficients of variation (×100)	280	438
median	$326	$239
% female	57.7	52.0
Age (mean)	20.8	34.0
% age 0–17	54.5	25.8
% age 18–44	31.8	39.2
% age 45–64	13.7	35.0
% with non-zero year-2 costs	80.6	73.7
% ever hospitalized in year 2	11.2	4.6

[a] For persons with at least one month of eligibility in each of years 1 and 2.

Table 3
Prevalence and year-2 costs for people by selected year-1 medical conditions in two populations.[a]

CC[b]	Condition category	Year-1 prevalence rates per 10,000		Year-2 relative cost[c]		% of year-2 total cost	
		Medicaid	Private	Medicaid	Private	Medicaid	Private
–	*All*	10,000	10,000	1.0	1.0	100	100
	ACC001: Infectious and Parasitic						
1	HIV/AIDS	10	4	3.6	5.4	0.3	0.2
2	Septicemia (Blood Poisoning)/Shock	39	11	5.1	11.5	2.0	1.3
3	Central Nervous System Infections	16	8	3.9	5.0	0.6	0.4
4	Other Infectious Disease	1,931	898	1.2	1.4	23.2	12.3
	ACC002: Malignant Neoplasm						
5	Metastatic Cancer	15	19	9.3	12.5	1.4	2.4
6	High Cost Cancer	23	18	6.6	12.4	1.5	2.2
7	Moderate Cost Cancer	34	45	5.5	7.0	1.9	3.2
8	Lower Cost Cancers/Tumors	58	173	4.6	3.7	2.7	6.5
	ACC003: Benign/In Situ/Uncertain Neoplasm						
9	Carcinoma In Situ	13	20	3.9	4.4	0.5	0.9
10	Uncertain Neoplasm	77	191	3.3	2.8	2.5	5.4
11	Skin Cancer, except Melanoma	10	56	3.6	2.5	0.4	1.4
12	Benign Neoplasm	344	996	2.2	1.8	7.5	18.1
	ACC004: Diabetes						
13	Diabetes with Chronic Complications	40	30	6.3	8.7	2.5	2.6
14	Diabetes with Acute Complications/Nonproliferative	52	31	4.5	6.0	2.3	1.9
15	Diabetes with No or Unspecified Complications	295	265	3.2	3.5	9.4	9.4
	ACC010: Mental Retardation						
36	Profound Mental Retardation	4	0[d]	12.2	2.3	0.5	0.0
37	Severe Mental Retardation	9	0[d]	10.3	5.6	0.9	0.0
38	Moderate Mental Retardation	11	0[d]	10.6	2.5	1.2	0.0
39	Mild/Unspecified Mental Retardation	59	1	5.3	3.0	3.1	0.0
	ACC011: Neurological						
40	Quadriplegia	8	2	9.7	9.7	0.8	0.2
41	Paraplegia	10	1	7.6	7.6	0.7	0.1
42	High Cost Neurological	87	48	5.1	5.2	4.5	2.5
43	Moderate Cost Neurological	318	99	3.2	3.5	10.0	3.5
44	Low Cost Neurological	309	289	2.6	2.5	7.9	7.3
	ACC013: Heart						
48	Congestive Heart Failure	110	59	5.0	7.2	5.5	4.2
49	Heart Arrhythmia	38	47	4.1	4.2	1.6	2.0
50	Acute Myocardial Infarction	16	12	4.3	6.2	0.7	0.7
51	Other Acute Ischemic Heart Disease	59	52	4.2	5.2	2.5	2.7
52	Chronic Ischemic Heart Disease	184	200	3.7	4.3	6.7	8.6
53	Valvular and Rheumatic Heart Disease	86	85	3.6	3.9	3.1	3.3
54	Hypertensive Heart Disease	42	47	3.5	3.9	1.5	1.8
55	Other Heart Diagnoses	83	59	4.4	4.8	3.6	2.8
56	Heart Rhythm and Conduction Disorders	107	115	3.7	3.3	3.9	3.8
57	Hypertension (High Blood Pressure)	550	738	2.6	2.3	14.6	17.3
	ACC016: Lung						
64	Chronic Obstructive Pulmonary Disease	350	120	3.1	3.8	10.9	4.6
65	Higher Cost Pneumonia	19	5	6.8	10.3	1.3	0.5
66	Moderate Cost Pneumonia	15	8	5.0	5.8	0.7	0.4
67	Lower Cost Pneumonia	371	145	2.4	3.1	9.0	4.5
68	Pulmonary Fibrosis and Other Chronic Lung Disorders	32	21	3.9	4.2	1.2	0.9
69	Pleural Effusion/Pneumothorax	31	19	5.9	8.5	1.8	1.6
70	Asthma	577	259	1.7	1.8	9.7	4.7
71	Other Lung Disease	1,933	858	1.5	1.9	28.7	16.3
	ACC019: Urinary System						
76	Dialysis Status	2	1	11.0	27.7	0.2	0.2
77	Kidney Transplant Status	2	3	5.6	9.6	0.1	0.3
78	Renal Failure	23	15	7.5	18.1	1.8	2.8
79	Nephritis	15	8	6.4	8.5	0.9	0.7
80	Other Urinary System	911	583	2.1	2.2	18.9	13.1

[a] Mean year-2 annualized costs are $1,965 in the Medicaid data, and $1,700 in the Private data.
[b] CC = number of the condition category in the DCG/HCC classification system.
[c] Year-2 relative cost is calculated as the group mean divided by the sample mean.
[d] 3 CCs in mental retardation with "0" prevalence actually represents between 0.12 and 0.27 in each CC, all prevalences of less than 1/2 per 10,000.

Moderate Cost Neurological (CC 43), and High Cost Pneumonia (CC 66). Also, people with Mental Retardation are much more expensive among the Medicaid enrollees than in the Private sample, possibly reflecting the fact that the more severe the problem, the more likely a person is to leave private insurance and become Medicaid eligible. In contrast, people in the five Neurological CCs are similarly expensive in the two populations and, for people in most other CCs, Medicaid payments are lower. Costs for Septicemia (Blood Poisoning)/Shock (CC 2), High Cost Cancer (CC 6), Dialysis Status (CC 76), and Renal Failure (CC 78) are strikingly low in Medicaid. Relative costs for Medicaid enrollees with these conditions are only half of those in the Private sample. Both disease prevalence and relative costs affect the percentage of total costs contributed by each CC. High prevalence CCs, such as Other Infectious Disease, Hypertension, and Other Lung Disease, all have large shares of the total costs although their relative costs are low. Moderate prevalence medical conditions with high costs (such as Diabetes, Congestive Heart Failure, Chronic Ischemic Heart Disease, and Chronic Obstructive Pulmonary Disease) also impose large healthcare burdens. The relative contributions to total cost by people in individual CCs are very different for the two populations. Three disease categories, Mental Retardation, Neurological, and Lung, account for much bigger percentages in the Medicaid sample as compared with the Private population. Quantification of such differences should help in allocating resources in Medicaid.

3.2. Comparisons for a diabetic cohort

Understanding disease burden is also important for care and disease management programs that target specific cohorts. For example, table 4 compares the prevalence of medical problems encountered in two diabetic cohorts selected from the Private and the Medicaid samples. The selected CCs include: all comorbidities commonly associated with diabetes (nutritional, heart, cerebrovascular, and vascular conditions); conditions with at least 10% prevalence among people with diabetes; and conditions with much elevated prevalence among people with diabetes (at least six times greater prevalence in the Private diabetic cohort). All the excess prevalence rates in the diabetic cohorts as shown in the last two columns are statistically significant ($p < 0.01$). Over 30% of people with diabetes also have health problems that are classified in CCs such as Major Symptoms, Hypertension, Screening, Minor Symptoms, and Other Musculoskeletal/Connective Tissue. Elevated rates of High Cost Eye, Cardiovascular and Other Vascular Diseases, and Renal Failure CCs among people with diabetes are not surprising; these are well recognized diabetic complications. However, while health care practitioners are aware that a diabetic cohort is at higher risk for complications such as infections, these data quantify the extent to which they experience more severe sequelae, such as Septicemia/Shock, High Cost Pneumonia, and Bone/Joint Infections/Necrosis. Whether these are unrelated coexisting conditions or complications of diabetes, the presence of these serious comorbidities signals individuals likely to consume more resources. To physicians, diabetes case managers and health plans, this table sends the message that people with diabetes are not only at risk for the commonly recognized vascular complications of diabetes but also have more and more severe complications in multiple organ systems. When caring for and managing a diabetic cohort, comorbidities matter immensely.

This table also shows that the lower expenditures for people with diabetes observed in Medicaid (in table 3) are definitely not explained by a lower comorbidity burden; in fact, the prevalence of almost all other diseases is far higher for Medicaid diabetics than for those in the private sample.

Table 5 summarizes our comparison among the disease-specific severity (DSS), Diagnostic Cost Group (DCG), and Prior Cost based classifications, as predictors of future costs in the private diabetic cohort. The DSS high-risk category identifies 0.8% of the 9,154 diabetes patients, whose costs account for 3.9% of the total costs incurred by all diabetes patients. This group averages 5.6 times higher costs than the $6,184 average of all people with diabetes.

While it is often not possible to directly compare methods that partition a population differently, in this instance, it is easy to see that the four DCG-based categories stratify the population better than the DSS categories do. Specifically, the top DCG category has both three times more people (209 vs. 69) and substantially higher average costs ($45,300 vs. $34,300) than the top DSS category, while the bottom two DCG categories combined have both slightly more people (86.3 vs. 84.9%) and minimally lower cost ($4,370 vs. $4,640) than the least risky (so-called "stable") DSS category. Also, the DSS method is intrinsically categorical and, given that the data lack additional clinical detail, cannot be used to further split the "stable" group, while the DCG method is able to use these same, imperfect data to identify 2 low cost subgroups, each with a sizeable number of people, whose next year's costs differ by a factor of more than two ($6,724 vs. $3,307).

Table 5 shows that prior cost is similar to the DCG model in its enhanced ability to identify subsets of people with diabetes who are relatively healthy or particularly expensive the following year. Both models are able to distinguish subgroups whose future costs ranged from about twice the average cost of the general population to more than 20 times that average. R^2s for predicting next year's cost in this private diabetic cohort are 4.7% for the DSS partition, 12.2% for the DCG partition, and 10.5% for the prior-cost partition. We further use the two continuous predictions (DCG score or prior-year costs) to divide the diabetic cohort into 50 2-percentile groups. Again, the R^2 is a bit higher for the DCG partition than the prior-cost partition (13.7 vs. 12%). People in the lowest 2-percentile group identified by the DCG scores are 40% ($1,141 vs. $1,861) less expensive than those identified by prior-year costs. In contrast, people in the highest DCG-predicted 2-percentile group are 14% ($49,148 vs. $42,997) more expensive than those in the prior-cost predicted group.

Table 4
Prevalence of selected conditions with presence of diabetes (rates per 10,000).[a]

CC[b]	Condition category[c]	Prevalence in the diabetic cohort		Excess within diabetic cohort[d]	
		Private	Medicaid	Private[e]	Medicaid[e]
		$n = 9,154$	$n = 15,049$	–	–
2	Septicemia (Blood Poisoning)/Shock	76	160	8.9	4.6
4	Other Infectious Disease	1,221	2,059	1.5	1.1
12	Benign Neoplasm	1,794	1,332	1.7	4.3
16	Protein-Calorie Malnutrition	17	64	4.8	3.7
17	Moderate Cost Endo/Metab/Fluid-Electlyte	377	1,192	5.3	4.8
18	Other Endocrine, Metabolic, Nutritional	2,809	3,532	3.1	6.5
19	Liver Disease	50	126	6.2	7.2
23	Low Cost Gastrointestinal	1,863	3,463	2.1	2.0
24	Bone/Joint Infections/Necrosis	118	136	8.3	12.2
26	Other Musculoskeletal/Connective Tissue	3,898	4,869	2.1	3.4
48	Congestive Heart Failure	504	1,253	9.1	17.0
49	Heart Arrhythmia	234	292	4.8	9.6
50	Acute Myocardial Infarction	105	185	8.7	17.4
51	Other Acute Ischemic Heart Disease	404	653	7.8	16.0
52	Chronic Ischemic Heart Disease	1,296	1,829	6.6	13.8
53	Valvular and Rheumatic Heart Disease	304	459	3.6	6.2
54	Hypertensive Heart Disease	320	433	7.2	14.4
55	Other Heart Diagnoses	338	674	5.9	10.4
56	Heart Rhythm and Conduction Disorders	438	561	3.8	6.1
57	Hypertension (High Blood Pressure)	3,381	4,638	4.6	11.0
58	High Cost Cerebrovascular Disease	51	136	3.8	6.1
59	Low Cost Cerebrovascular Disease	410	664	7.3	12.3
60	High Cost Vascular Disease	375	646	7.7	14.6
61	Thromboembolic Vascular Disease	174	227	5.8	10.8
62	Atherosclerosis	190	475	8.3	15.0
63	Other Circulatory Disease	367	683	3.2	7.8
65	High Cost Pneumonia	32	78	6.8	4.5
71	Other Lung Disease	1,501	3,447	1.8	1.8
72	High Cost Eye	737	820	6.0	5.9
73	Low Cost Eye	1,231	2,429	2.4	1.5
75	Low Cost Ear, Nose, and Throat	2,993	4,760	1.1	1.1
76	Dialysis Status	5	31	5.5	32.5
77	Kidney Transplant Status	33	24	10.9	13.1
78	Renal Failure	184	296	16.2	20.1
79	Nephritis	54	175	7.6	18.3
80	Other Urinary System	1,315	2,524	2.2	2.9
83	Low Cost Genital	1,844	2,039	1.6	2.2
91	Chronic Ulcer of Skin	268	399	16.6	17.6
92	Other Dermatological	2,384	2,451	1.8	1.9
97	Other Injuries and Poisonings	2,481	3,156	1.4	1.3
99	Major Symptoms	3,161	4,946	2.5	2.3
100	Minor Symptoms, Signs, Findings	3,697	4,822	2.1	2.2
117	Screening/Observation/Special Exams	3,551	5,943	1.4	1.2

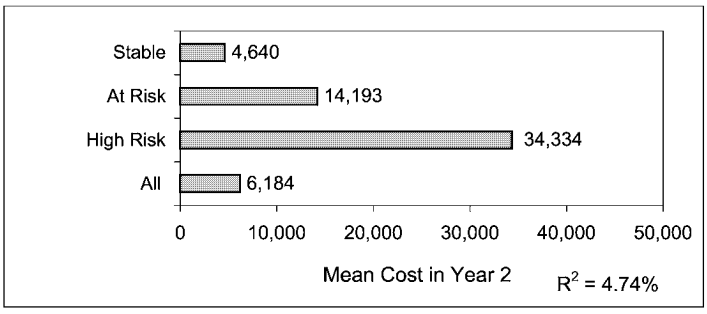
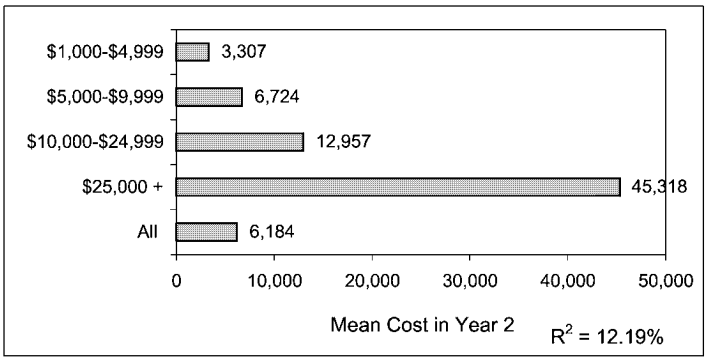
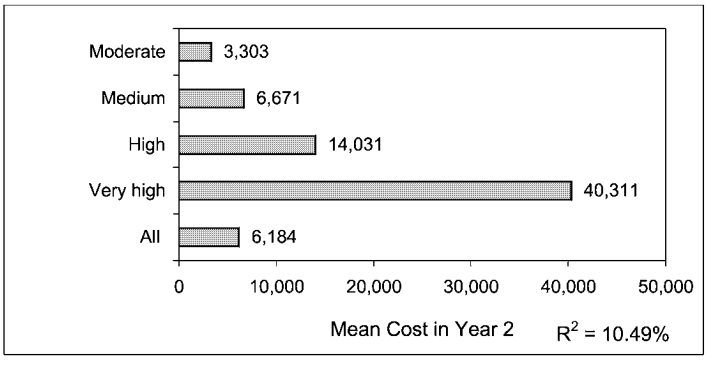
[a] A random sample of 300,000 from Private data (Medstat 1996–1997 data) and the full Medicaid sample ($n = 493,238$).
[b] CC = number of the condition category in the DCG/HCC classification system.
[c] Based on year-1 medical conditions.
[d] Ratio of diabetic/non-diabetic prevalence rates.
[e] All numbers are significant at $p = 0.01$ level.

4. Discussion

Clearly, populations differ in disease prevalence, which affects their health care needs. Because each population has its own "medical signature", understanding the prevalence of disease and the distribution of medical problems is important in population-based health management.

Unlike commercial health insurance plans, Medicaid is a government program that provides heath insurance to the poor and disabled. Not surprisingly the disease profiles of our Medicaid and Private under-age-65 samples differ markedly: Medicaid beneficiaries have more recorded medical problems, but lower resource utilization in condition-defined cohorts. This is typical across most conditions, but most prominent for conditions with high costs. For many relatively rare and serious conditions, disease prevalence is more than double in Medicaid than in the Private sample. Conversely, the relative health care expenditures for Medicaid enrollees with specific medical problems are less than a half or a third of those in the privately insured population.

Table 5
DSS vs. DCG vs. prior cost based prospective classifications of a diabetic cohort*: distribution and next year's costs.

	% of People	% of Costs	Relative Cost
DSS (Diabetes-specific severity-based classification)			
	84.9	64.0	0.8
	14.4	32.1	2.3
	0.8	3.9	5.6
	100.0	100.0	1.0
DCG (Expected cost categories)			
	59.5	32.1	0.5
	26.8	29.0	1.1
	11.4	23.4	2.1
	2.3	15.5	7.3
	100.0	100.0	1.0
Prior Cost (Expected cost categories)			
	59.5	32.0	0.5
	26.8	29.0	1.1
	11.4	25.1	2.3
	2.3	13.9	6.5
	100.0	100.0	1.0

* 9,154 individuals with diabetes are identified from a random sample of 300,000 from Private data (Medstat 96-97 data)

The high prevalence of disease among the younger Medicaid enrollees reflects the fact that many enroll when they become disabled and need medical care. In contrast, most people in the Private sample are active employees, their spouses and children, leading to a rather typical under-age-65 U.S. population. The disease profiles show the distinct illness burdens in the populations and should be of great value in care management.

Why are Medicaid's costs within CCs so much lower? We looked to see if lower comorbidity burden could explain this, since younger people are more likely to have medical problems "one at a time", and additional comorbidities add so much to total care costs within a cohort of people defined by the presence of a single disease. Although the Medicaid population is substantially younger, the DCG framework enabled us to see that disease prevalence for almost all CCs is not only higher in Medicaid than in the private sample in its full population, but also within its diabetic cohort. Thus, the lower expenditures for people with diabetes in Medicaid occur in spite of a higher burden of comorbid disease. Other potential explanations for lower spending in Medicaid are (1) lower prices per unit of service and (2) lower volume or less intense mix

of services (underutilization). Medicaid does generally spend less than other payers for the same services, and it is possible to examine the effect of differential pricing in these data. However, utilization comparisons are trickier, since Medicaid beneficiaries are more likely than the commercially insured to receive care through other programs, such as Medicare or the Veteran's Health Administration (VHA) system. In our data, we neither know who is eligible for such care, nor do we capture the care itself or its cost. The impact of such utilization could be considerable, since about one sixth of all Medicaid enrollees are also entitled to Medicare benefits.

The total medical costs for people with similar illnesses also differ significantly across populations. Medical problems that are relatively rare and inexpensive in one population may be much more consequential in another. For example, Mental Retardation and Pregnancy are all much more common in Medicaid. Also, the cost of a medical problem in a delivery system depends heavily on the nature of the insurance benefit. For example, people with End Stage Renal Disease (ESRD) or HIV will incur substantial extra costs in a system with a drug benefit. Therefore, care management and resource allocation should be adapted to the populations served.

The DCGs not only predict future costs as accurately as prior spending, they also provide clinical descriptions at the group or individual level. Using this clinical information, the DCGs can easily create disease burden profiles to help health care managers identify population risks and allocate resources while managing population health. On the other hand, the population disease profiles are extremely useful in managing high cost diseases with multiple comorbidities, such as diabetes. The CCs provide valuable information for disease managers to better understand these populations, and to create individual clinical profiles that identify people at particularly high risk. Because the DCGs provide credible estimates of future costs *for individuals*, they can be powerful management tools. However, it is important for managers to understand that only for moderately large groups (n's of at least several hundred) are actual costs likely to be similar to predicted ones. The CCs identify individuals with "high risk" and potentially manageable combinations of comorbidities; when actually assigning cases to managers, more detailed medical record data would be consulted. Compared with a disease management program based solely on the severity of the principal medical problem, the DCG model recognizes the range of distinct medical problems for each individual, which may matter as much as, or more than, disease-specific severity. This matches the experience of caregivers who must manage their patients from a "whole person" perspective. With its straightforward, accumulative methodology, the DCG methodology can be used to assess and describe populations. It supports a program that manages populations rather than "diseases".

Acknowledgement

We would like to thank John Z. Ayanian, Lisa I. Iezzoni, Marilyn Schlein Kramer, Diane Laurent, and Larry Gertzog for helpful discussion and comments on this paper.

References

[1] A.S. Ash, R.P. Ellis, G.C. Pope, J.Z. Ayanian, D.W. Bates, H. Burstin, L.I. Iezzoni, E. McKay and Y. W., Using diagnoses to describe populations and predict costs, Health Care Financing Review 21(3) (2000) 7–28.

[2] American Diabetes Association, Economic consequences of diabetes mellitus in the U.S. in 1997, Diabetes Care 21 (1998) 296–309.

[3] L.A. Lavery, H.R. Ashry, W. van Houtum et al., Variation in the incidence and proportion of diabetes-related amputations in minorities, Diabetes Care 19 (1996) 48–52.

[4] T.A. Hodgson and A.J. Cohen, Medical care expenditures for diabetes, its chronic complications, and its comorbidities, Preventive Medicine 29 (1999) 173–186.

[5] Diabetes: An American epidemic, Newsweek 136 (2000, September 4) 40–47.

[6] J.P. Weiner, A. Dobson, S.L. Maxwell and K. Coleman, Risk-adjusted Medicare capitation rates using ambulatory and inpatient diagnoses, Health Care Financing Review 17(3) (1996) 77–100.

[7] R.F. Averill, N.I. Goldfield, J. Eisenhandler, J.H. Muldoon, J.S. Hughes, J.M. Neff, J.C. Gay, L.W. Gregg, D.E. Gannon, B.V. Shafir, F.A. Bagadia and B.A. Steinbeck, Development and evaluation of clinical risk groups (CRGs), US Department of Commerce, Cooperative Agreement No. 70NANB5H1013 (2000).

[8] American Diabetes Association, Standards of medical care for patients with diabetes mellitus, Diabetes Care 23 (2000) S32–S42.

[9] American Diabetes Association, Type 2 diabetes in children and adolescence, Diabetes Care 22 (2000) 381–389.

[10] K.A. Javor, J.G. Kotsanos, R.C. McDonald, A.D. Baron, J.G. Kesterson and W.M. Tierney, Diabetic ketoacidosis charges relative to medical charges of adult patients with type I diabetes, Diabetes Care 20 (1997) 349–354.

[11] A.D. Mooradian, S. McLaughlin, C.C. Boyer and J. Winter, Diabetes care for older adults, Diabetes Spectrum 12 (1999) 70–77.

[12] www.diabetes.org

[13] J.P. Newhouse, E.M. Sloss, W.G. Manning and E.B. Keeler, Risk adjustment for a children's capitation rate, Health Care Financing Review 15(1) (1993) 39–54.

[14] S.M.C. Payne, R.D. Cebul, M.E. Singer, J. Krishnaswamy and K. Gharrity, Comparison of risk-adjustment systems for the Medicaid-eligible disabled population, Medical Care 38 (2000) 422–432.

[15] R.T. Meenan, M.C. O'Keefee-Rosetti, M.C. Hornbrook, D.J. Bachman, M.J. Goodman, P.A. Fishman and A.V. Hurtado, The sensitivity and specificity of forecasting high-cost users of medical care, Medical Care 37 (1999) 815–823.

[16] G.N. Herrle, Rating and underwriting, in: *The Managed Health Care Handbook*, 2nd edn., ed. P.R. Kongstvedt (Aspen Publishers, Gaithersburg, 1999).