Volume 11 • Number 5 • 2008
VALUE IN HEALTH

The Estimation Power of Alternative Comorbidity Indices

Onur Baser, PhD,[1]* Liisa Palmer, PhD,[2] Judith Stephenson, SM[3]

[1]STATinMED Research and University of Michigan, Ann Arbor, MI, USA; [2]Thomson Healthcare, Washington, DC, USA; [3]HealthCore, Inc. Wilmington, DE, USA

ABSTRACT

Objective: Health-care expenditures are strongly influenced by overall illness burden. Appropriate risk adjustment is required for correct policy analysis. We compared three risk adjustment methods: the Charlson comorbidity index (CCI), the chronic disease score (CDS), and the Agency for Healthcare Research and Quality's comorbidity index (AHRQCI) in terms of their estimation power in analyzing health-care expenditures.

Method: Data from the Thomson MarketScan® Research Databases (Thomson Healthcare, Ann Arbor, MI) were used to estimate total health-care expenditures of migraine patients treated by a triptan. Seven distinct multivariate models were evaluated for model fit (CCI only, CDS only, AHRQCI only, CCI + CDS, CCI + AHRQCI, CDS + AHRQCI, and CCI + CDS + AHRQCI). The estimation power of these indices (alone and in combination) was evaluated using Bayesian and Akaike information criteria, log-likelihood scores, and pseudo R^2 values.

Results: Confirming results from previous studies, when comorbidity indices were considered individually the results were inconclusive. Statistically the best performance was observed in the model that included all three of the comorbidity measures (CCI + CDS + AHRQCI); however, the practical differences in the estimated values were small.

Conclusion: Low correlation between these comorbidity indices shows that it is possible to have potential risk factors that are not captured in the single comorbidity index. Each comorbidity measure considers different risks, and the collinearity of the three measures is not strong enough to preclude using them simultaneously in the same model.

Keywords: comorbidity, health-care costs, regression analysis, risk adjustment.

Introduction

A comorbidity is a condition other than the diagnosis of interest that may influence the treatment outcome. Increasingly, summary measures or comorbidity indices are being applied in health services research as proxy measures of overall health status. In fact, it has become standard in health services research to include a summary measure of an individual's comorbid conditions along with other clinical and demographic characteristics within the multivariate framework. Being able to capture effectively an individual's general health status is particularly important when analyzing health-care claims data.

Three common comorbidity measures are the Charlson comorbidity index (CCI), the chronic disease score (CDS), and the Agency for Healthcare Research and Quality's comorbidity index (AHRQCI). CCI contains 19 categories of comorbidity, defined primarily using ICD-9-CM diagnosis codes (a few procedure codes are also employed). Each category has an associated weight based on the adjusted risk of 1-year mortality. The overall comorbidity score reflects the cumulative increased likelihood of 1-year mortality—the higher the score, the more severe the burden of comorbidity. CDS is a risk-adjusted metric based on age, sex, and history of dispensed drugs. AHRQCI contains a comprehensive set of 30 comorbidity measures, defined using ICD-9-CM diagnosis codes associated with increased hospital length of stay, charges, and mortality.

The predictive validity of individual comorbidity measures has been examined in detail. Some of these studies compared different variations for the same index [1–8], some compared the individual indices in disease-specific populations [9–16], some evaluated validity using different administrative claims data [17,18], some compared diagnosis-based index scores with drug-based scores [19–21], and others analyzed the predictive power in estimating mortality [22–25]. Few studies, however, compare the use of different measures within the same population when estimating health-care expenditures [26–28]. Schneeweiss and Maclure [26] examined the performance of comorbidity scores for use in epidemiologic research, Perkins et al. [27] evaluated the ability of commonly used comorbidity measures to predict mortality and health-care costs over a 1-year period, and Farley et al. [28] compared several comorbidity indices and count

Address correspondence to: Onur Baser, STATinMED Research and University of Michigan 110 Depot Street, Ann Arbor, MI 48104, USA. E-mail: obaser@statinmed.com
10.1111/j.1524-4733.2008.00343.x

*Formerly an employee of Thomson Healthcare.

  1098-3015/08/946 946–955

measures to predict health-care expenditures. Each of these studies, however, focused on individual comorbidity indices rather than evaluating the effect of using multiple indices within a single model.

Claims-based comorbidity scores depend both on the type of clinical conditions included in the development of the score and the assigned relative weight, thus each comorbidity score can proxy the severity of disease from a different perspective (the distribution of comorbid conditions in the source population, the study end point, and the accuracy of the administrative data were assumed to be constant among the different index measures when we compared them). Therefore, combining the index scores in a given model can be more informative than choosing just one. It has been suggested that using a single comorbidity index might be methodologically unsound [1,4,5,29]. Investigators might consider potential risk factors that are not captured in the single comorbidity index, but may be relevant to a patient's particular condition or procedure. For example, one disadvantage of using medical-based indices such as CDS is that individuals who do not receive or fill a prescription will not be detected [30]. Nevertheless, this individual would be detected with diagnosis-based scores such as CCI or AHRQCI. Conversely, diagnosis-based indices assume that different diagnoses or differing levels of severity within a single diagnosis are similar with respect to outcome of interest [31]. A patient with hypertension and diabetes, for example, could be considered sicker than a patient with "only" a stroke. While an important advantage of administrative data sets is that they offer large, representative samples from the population of interest, applying a single comorbidity index may negate this benefit.

The goal of this analysis was to compare the performance of CCI, CDS, and AHRQCI, and to examine the efficacy of their combined use in a single model in terms of estimating health-care expenditures. Of particular interest was determining if the concurrent use of these indices in a single model will predict health-care expenditures better than using only a single index.

We address these issues by evaluating the health-care expenditures of patients with a diagnosis of migraine and a filled prescription for a triptan as a case study. Approximately 120 million Americans suffer from moderate to severe attacks of migraines and the aggregate cost of providing health care for persons with migraine is substantial [32]. Triptans are family of tryptamine drugs currently marketed for the treatment of migraine and cluster headaches. Alternative treatment choices for migraines differ not only in efficacy and side effects, but also in cost [33]. To estimate risk adjusted health-care expenditures of migraine patients with triptan treatment would be the first step to analyze the average treatment effect of triptan use with respect to other treatment choices. Because previous studies found that comorbid conditions for migraine patients influence health-related expenditures, it is important to control for these factors, appropriately [34].

Methods

Data Sources

Data used for the analysis were derived from Thomson's MarketScan® Commercial Claims and Encounters Database and MarketScan Medicare Supplemental and COB Database (Thomson Healthcare, Ann Arbor, MI) from the period of January 1, 2001 to December 31, 2004. In 2004, these databases represented the health services of approximately 20 million employees, dependents, and retirees in the United States, with primary or Medicare supplemental coverage through privately insured fee-for-service, point-of-service, or capitated health plans.

Sample Selection Criteria

An analytic sample was extracted from the MarketScan databases. Individuals who satisfied all of the following criteria were included:

1. at least 12 years of age or older;
2. at least one outpatient diagnosis of migraine (ICD-9-CM = 346.xx) during the patient identification period (January 1, 2002 through December 31, 2003);
3. at least one filled prescription for any of the following triptans: sumatriptan (Imitrex®, Imigran®), rizatriptan (Maxalt®), naratriptan (Amerge®, Naramig®), zolmitriptan (Zomig®), eletriptan (Relpax®), almotriptan (Axert®, Almogran®), and frovatriptan (Frova®, Migard®);
4. at least 6 months of continuous enrollment before the first triptan prescription;
5. at least 12 months of continuous enrollment after the first triptan prescription;
6. eligible for medical and drug benefits during the 18-month study period.

Key Variables and Definitions

A patient-level analytic file containing all variables was constructed from the enrollment and claims data in the MarketScan databases. In addition to demographic variables such as age and sex, the analytic file included patient location based on US Census categories (northeast, north central, south, and west) and urban/rural residency. The file also contained information about each patient's health insurance and his or her provider's specialty associated with health-care visits of interest.

The dependent variable—total health expenditures—was calculated as the sum of all inpatient, outpatient, and outpatient pharmaceutical expenditures for all medical services. This included all

services paid for by insurance, as well as out-of-pocket copayments and deductibles. The analytic file contains patients with fee-for-service health plans and those with partially or fully capitated plans. Data on cost were not available, however, for the capitated plans. Therefore, the value of patients' service utilization under capitated plan was imputed using average payments from the MarketScan fee for service by region.

Using standard methodology, which is briefly described in the following sections, CCI, CDS, and AHRQCI scores were calculated as measures of overall health status for each patient in the analytic file. The presence or absence of the comorbidities was evaluated during the preperiod (6 months before index) for migraine populations with triptan use.

CCI. The most commonly used index in health outcomes studies is CCI, which assigns a weight ranging from 1 to 6 according to disease severity for 19 conditions [34]. CCI contents and weighting scheme are based on Cox proportional hazards modeling [35]. The weights for each condition are summed and a score is assigned to each patient. The original index was developed in an inpatient setting, using medical review to predict the risk of mortality. The index has since adopted several weights, some of which allow outpatient diagnoses to contribute to the score [3,29]. Regardless of the version, CCI has practically insignificant effects in predicting health-care utility and indices [36,37].

CDS. CDS is an aggregate comorbidity measure based on current medication use. Von Korff et al. [20] created CDS to serve as an indicator of an individual's morbidity and overall health status. The index was established under the following guidelines: 1) the score should increase with the number of chronic diseases under treatment; 2) the score should increase with the complexity of the treatment regimen; 3) progressive and life-threatening diseases should result in higher scores; and 4) treatment regimens should target diseases and not just symptoms. The first three guidelines are indicative of disease severity while the last ensures some symptomatic medications are excluded from the score.

AHRQCI. AHRQCI is based on a comprehensive set of 30 ICD-9-CM comorbidity flags [38,39]. Diagnosis codes that represent potential complications are excluded as indicators of comorbid illness. AHRQCI uses diagnosis-related groups (DRGs) to differentiate secondary diagnoses from comorbidities. For example, if heart failure appeared on a record, it will be counted as a comorbidity only if the diagnosis does not fall into any of the cardiac DRGs. The final comorbidity scores can be calculated as the sum of comorbid conditions present. Current coding for the index is available from the Agency for Healthcare Research and Quality [38]. The algorithm for the most recently version of the AHRQCI is presented in the appendix.

Analysis

Total expenditures were estimated as a function of each unique comorbidity index and as a function of the combination of the comorbidity indices, adjusting for an individual's demographic and utilization characteristics. These baseline variables included age, sex, geographic region of residence, and insurance plan type. Variables noting the year of first triptan prescription and provider specialty were also included. Seven unique models, each based on the same baseline variables, were evaluated using the following combinations of comorbidity indices:

- Model 1: baseline variables + CCI;
- Model 2: baseline variables + CDS;
- Model 3: baseline variables + AHRQCI;
- Model 4: baseline variables + CCI + CDS;
- Model 5: baseline variables + CCI + AHRQCI;
- Model 6: baseline variables + CDS + AHRQCI;
- Model 7: baseline variables + CCI + CDS + AHRQCI.

If there is a strong collinearity between the measures, it would be redundant to include all of them into the model because this may increase the variance of estimated coefficients.

Variance inflation factor (VIF) was used to determine the presence of multicollinearity among the comorbidity indices. VIF expresses the degree to which collinearity among the predictors degrades the precision of an estimate. According to an informal rule of thumb, there is evidence of multicollinearity if the largest VIF is greater than 10 and the mean of all VIFs is considerably larger than 1 [40].

Log-linear and generalized linear models (GLMs) are two commonly used methods of analyzing health-care expenditure data. Manning and Mullahy [41] describe the criteria necessary for choosing between the two. The Park test, which is applied for family selection in a GLM [41], indicated that a GLM model with gamma family was most appropriate for this analysis.

Bayesian and Akaike information criteria (BIC and AIC, respectively), log-likelihood scores, and pseudo R^2 values were calculated for all models to determine the best fit statistically. Pseudo R^2 values are the percentage reduction of log-likelihood values from the fully restricted model (model with no covariates) to the models in question (Model 1 to Model 7).

To compare the prediction performance of the models, we split the sample randomly into two equal parts—the training subsample and the test subsample. Each of the seven models was fitted on the training

Table 1 Sample statistics (N = 47,743)

Variable	Mean	SD
Total cost	7,495	12,742
Index year 03	0.3072	0.4613
Index year 04	0.0825	0.2752
Age	40.2700	12.3600
Female	0.8358	0.3705
North central	0.2619	0.4400
South	0.4252	0.4944
West	0.2065	0.4048
Urban	0.7788	0.4151
Capitated health plan	0.2464	0.4309
OB/GYN	0.0083	0.0907
Pediatrician	0.0140	0.1174
Pain specialist	0.0004	0.0210
Psychiatrist	0.0183	0.1342
Other physician	0.3360	0.4723
CCI	0.8730	2.0213
AHRQCI	0.4497	0.7903
CDS	2.2622	2.5166

CCI, Charlson comorbidity index; AHRQCI, Agency for Healthcare Research and Quality's comorbidity index; CDS, chronic disease score.

subsample and used to form predictions for all individuals in the test subsample. Each individual's predicted expenditures were compared with that individual's actual expenditures. The average squared prediction error (ASPE) was then computed for each model.

$$\text{ASPE} = \sum (\text{E}(\text{Y}_\text{k}) - \text{Y}_\text{k})^2,$$

where E[.] is the expected value of the term in parentheses, and the index k runs through the m individuals in the test subsample. The ASPE of different models can be compared directly, with better models producing smaller ASPE [42,43].

Statistical analysis is done using STATA software, version 9 (STATA Corp., College Station, TX).

Results

The analytic sample consisted of 47,743 migraine patients who used a triptan between January 1, 2002 and December 31, 2003. Demographic characteristics for the sample are provided in Table 1. Sample patients were on average 40 years of age and disproportionately female (83%). Nearly half resided in the South and almost three-quarters lived in urban areas. Twenty-four percent of the sample was insured by a capitated plan.

Mean scores for CCI, CDS, and AHRQCI were 0.87, 2.27, and 0.45, respectively. Scores ranged from 0 to 28 for CCI, 0 to 17 for CDS, and 0 to 9 for AHRQCI.

Table 2 presents the results of the multicollinearity analysis. This analysis indicated no evidence of multicollinearity in any of the models according to the criteria mentioned earlier. Furthermore, when considering the conditional correlations associated with Model 7, only 10% of the variation in CCI is explained by CDS and AHRQCI; while 17% of the variation in AHRQCI is explained by CCI and CDS; and 24% of the variation in CDS is explained by CCI and AHRQCI. These results indicate that each index measures things that the other two do not. Consequently, including all three indices in the same model would control for unmeasurable variation which may otherwise occur by excluding them.

Table 3 presents the coefficients, *P*-values, marginal effects, and estimated cost values for each of the multivariate models. Adjusted costs range from $6300 to $6670, depending on the model. The marginal effect of a one-point increase in CCI ranges from $646 to $1075, depending on the model. The corresponding incremental effects are $556 to $877 and $1500 to $2724 for CDS and AHRQCI indices, respectively.

In order to choose the best model in terms of predictability, BIC, AIC, log-likelihood scores, and pseudo R^2 values were calculated for each model. Table 4 presents the performance of each comorbidity measure independently and in combination.

When considered individually, results were inconclusive. This is consistent with previous literature. In terms of ASPE, CDS is the most powerful. According to the pseudo R^2, AHRQCI results in greater percentage reduction in likelihood value, while BIC indicates that CCI is the best of the three. When evaluating the combination of comorbidity indices, Model 4 (CCI + AHRQCI) has the worst performance when compared to Model 6 (CDS + AHRQCI) and Model 5 (CCI + CDS). Nevertheless, when comparing Model 6 and Model 5, the ASPE criteria clearly indicate that Model 6 is favored, while pseudo R^2 values indicate that Model 5 is favored. When the data presented in Table 4 is evaluated in its entirety, the model with the best fit is the most expanded model—Model 7—which includes all three comorbidity measures (CCI + CDS + AHRQCI). Smaller values, indicating better model fit, were observed for the AIC, BIC, and log-likelihood scores for Model 7. Hence, minimum ASPE is achieved

Table 2 Analysis of multicollinearity

Model	VIF	Mean VIF	Conditional correlation
Model 4			
CCI	1.11		
AHRQCI	1.12	1.38	0.29
Model 5			
CCI	1.06		
CDS	1.10	1.37	0.22
Model 6			
AHRQCI	1.18		0.37
CDS	1.20	1.39	
Model 7			
CCI	1.12		0.10
AHRQCI	1.24		0.17
CDS	1.22	1.38	0.24

VIF, variance inflation factor; CCI, Charlson comorbidity index; AHRQCI, Agency for Healthcare Research and Quality's comorbidity index; CDS, chronic disease score.

Table 3 Model parameters

Model	Coefficient	*P*-value	Marginal effect	Estimated cost
Model 1				
CCI	0.1613	<0.0001	$1076	$6670
Model 2				
AHRQCI	0.4115	<0.0001	$2724	$6620
Model 3				
CDS	0.1324	<0.0001	$878	$6629
Model 4				
CCI	0.1142	<0.0001	$738	
AHRQCI	0.3258	<0.0001	$2104	$6457
Model 5				
CCI	0.1270	<0.0001	$815	
CDS	0.1114	<0.0001	$715	$6417
Model 6				
AHRQCI	0.3035	<0.0001	$1957	
CDS	0.0962	<0.0001	$620	$6449
Model 7				
CCI	0.1024	<0.0001	$647	
AHRQCI	0.2377	<0.0001	$1502	
CDS	0.0881	<0.0001	$557	$6320

CCI, Charlson comorbidity index; AHRQCI, Agency for Healthcare Research and Quality's comorbidity index; CDS, chronic disease score.

and, compared with other models, highest percentage reduction in log-likelihood value is observed by Model 7.

Discussion

As a forecasting problem, if there is a variable in a model that has a *P*-value of 0.0001 in the prediction equation, every metric (AIC, BIC, etc.) will indicate to keep that variable. The resulting model will perform much better in out-of-sample prediction as the ASPE values demonstrate in Table 4. With a sufficiently large sample, in this case 47,743 observations, it is possible to have significant relationships between the comorbid measures and the outcome variable because of the sample size while there is no underlying correlation that exists. Thus, we performed sensitivity analysis of the results for randomly chosen subsamples. The required sample sizes for each model, ranged from 1520 to 1734, and was chosen based on a 5% significance level, 90% power, and a specified effect size (derived from changes in pseudo R^2 values). Conclusions for these models are not different than those presented for the original study models.

It is possible to have potential risk factors that are not captured in the single comorbidity index. For example, AHRQCI covers depression, psychosis, and drug and alcohol dependence, while neither the CDS nor the CCI includes these conditions. Moreover, AHRQCI weights each condition equally where the CDS and CCI do not. For example, in oncology where CCI weights are 1 or 2 depending on the cancer type and CDS weight all cancer drugs as a 3. These differences are confirmed by the low correlations among the three commonly measured index scores. Ninety percent of CCI variation cannot be explained by the AHRQI and CDS while 83% of AHRQI variation cannot be explained by CCI and CDS. Similarly, 76% of CDS variation is not explained by the CCI and AHRQI.

Low correlation has two important statistical consequences. Health-care expenditures are strongly influenced by overall illness burden and in administrative data sets; there is a limited set of variables from which to choose to capture the burden of illness. Because low correlation suggests that these indices measure different aspects of overall illness burden, using a single index might leave a considerable portion to illness

Table 4 Comorbidity score performance

Model	AIC	BIC	Log-likelihood	Pseudo R^2	ASPE
Model 1 (CCI)	19.59916	–466,960	–467,845	0.0124	189769
Model 2 (AHRQCI)	19.59648	–467,088	–467,781	0.0125	188356
Model 3 (CDS)	19.61134	–466,379	–468,136	0.01178	187489
Model 4 (CCI & AHRQCI)	19.54664	–465,458	–466,591	0.0150	182944
Model 5 (CCI & CDS)	19.53419	–465,052	–466,293	0.0156	181369
Model 6 (AHRQCI & CDS)	19.54403	–464,583	–466,528	0.0152	180969
Model 7 (CC, AHRQCI & CDS)	19.50363	–461,502	–465,563	0.0170	147089

AIC, Akaike information criteria; BIC, Bayesian information criteria; ASPE, average squared prediction error; CCI, Charlson comorbidity index; AHRQCI, Agency for Healthcare Research and Quality's comorbidity index; CDS, chronic disease score.

Table 5 Model parameters using indicators individually

	Model 1		Model 2		Model 3		Model 4			Model 5			Model 6			Model 7		
	ME	*P*	ME	*P*	ME	*P*	ME	*P*	VIF	ME	*P*	VIF	ME	*P*	VIF	ME	*P*	VIF
ccicon1	5,526	0.00								4,523	0.00	1.04	4,557	0.00	1.04	4,156	0.00	1.04
ccicon2	2,849	0.00								1,870	0.00	1.17	2,173	0.00	1.07	1,673	0.00	1.18
ccicon3	2,576	0.00								2,168	0.00	1.13	2,115	0.00	1.04	1,894	0.00	1.13
ccicon4	1,447	0.00								687	0.01	1.03	957	0.00	1.02	487	0.07	1.03
ccicon5	1,630	0.00								1,471	0.00	1.07	1,443	0.00	1.06	1,389	0.00	1.07
ccicon6	2,596	0.00								1,967	0.00	1.34	1,914	0.00	1.10	1,595	0.00	1.36
ccicon7	2,467	0.00								1,341	0.00	1.25	1,643	0.00	1.03	970	0.00	1.26
ccicon8	309	0.00								277	0.00	1.02	275	0.00	1.01	246	0.00	1.02
ccicon9	286	0.00								217	0.00	1.02	242	0.00	1.01	196	0.00	1.02
ccicon10	1,410	0.00								1,078	0.00	1.09	1,157	0.00	1.07	990	0.00	1.09
ccicon11	1,470	0.00								867	0.00	1.06	1,075	0.00	1.02	730	0.00	1.06
ccicon12	1,291	0.00								856	0.00	1.71	793	0.00	1.37	655	0.00	1.76
ccicon13	2,764	0.00								2,529	0.00	1.02	2,217	0.00	1.02	2,212	0.00	1.02
ccicon14	2,350	0.00								1,200	0.00	1.33	2,229	0.00	1.02	1,267	0.00	1.33
ccicon15	1,322	0.00								1,222	0.00	1.37	1,200	0.00	1.10	1,202	0.00	1.38
ccicon16	1,202	0.00								851	0.00	1.12	1,139	0.00	1.01	831	0.00	1.12
ccicon17	978	0.00								902	0.00	1.22	950	0.00	1.06	902	0.00	1.22
ccicon18	N/A	N/A								N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ccicon19	1,238	0.00								958	0.00	1.96	1121	0.00	1.01	914	0.00	1.96
cds_ht																		
cds_ri																		
cds_as			670	0.00			491	0.00	1.07				480	0.00	1.10	427	0.00	1.08
cds_ra			133	0.90			−295	0.78	1.00				−416	0.69	1.00	−433	0.68	1.00
cds_c.			1,553	0.00			874	0.00	1.09				880	0.00	1.03	786	0.00	1.10
cds_pk			809	0.00			664	0.00	1.01				635	0.00	1.01	462	0.02	1.01
cds_ht			924	0.00			622	0.00	1.15				635	0.00	1.10	537	0.00	1.16
cds_db			1,459	0.00			905	0.00	1.72				523	0.00	1.36	632	0.00	1.76
cds_ep			1,982	0.00			1,437	0.00	1.13				1,719	0.00	1.06	1,345	0.00	1.13
cds_ar			323	0.18			82	0.73	1.01				−246	0.29	1.02	−134	0.56	1.02
cds_ac			−837	0.35			−669	0.45	1.00				−410	0.65	1.00	−507	0.56	1.00
cds_ul			2,435	0.00			1,910	0.00	1.02				2,053	0.00	1.02	1,774	0.00	1.02
cds_gc			4,552	0.20			5,607	0.12	1.00				6,467	0.08	1.00	6,836	0.07	1.00
cds_gt			2,637	0.00			2,700	0.00	1.01				2,017	0.00	1.01	2,211	0.00	1.01
cds_ch			1,577	0.00			1,401	0.00	1.11				1,293	0.00	1.12	1,239	0.00	1.12
cds_mg			568	0.06			606	0.04	1.01				887	0.00	1.01	875	0.00	1.01
cds_tb			17,358	0.00			11,779	0.00	1.00				4,923	0.00	1.00	3,948	0.01	1.01
AHRQCI_01					8,370	0.00	4,930	0.00	1.02	2,381	0.01	1.11				922	0.23	1.11
AHRQCI_02					1,937	0.00	1,529	0.00	1.02	1,603	0.00	1.02				1,259	0.00	1.02
AHRQCI_03					2,332	0.00	1,813	0.00	1.02	1,700	0.00	1.03				1,268	0.00	1.03

Table 5 continued

	Model 1		Model 2		Model 3		Model 4			Model 5			Model 6			Model 7		
	ME	*P*	ME	*P*	ME	*P*	ME	*P*	VIF	ME	*P*	VIF	ME	*P*	VIF	ME	*P*	VIF
AHRQCI_04					5,657	0.01	2,678	0.12	1.01	2,808	0.10	1.01				1,379	0.37	1.01
AHRQCI_05					4,712	0.00	3,669	0.00	1.01	−444	0.41	1.11				−361	0.49	1.11
AHRQCI_06					1,507	0.00	760	0.00	1.14	927	0.00	1.09				387	0.00	1.16
AHRQCI_07					9,141	0.00	8,963	0.00	1.01	6,472	0.00	1.02				6,411	0.00	1.02
AHRQCI_08					5,763	0.00	3,081	0.00	1.04	4,803	0.00	1.02				2,526	0.00	1.05
AHRQCI_09					3,349	0.00	2,320	0.00	1.05	1,040	0.00	1.33				647	0.00	1.35
AHRQCI_10					2,693	0.00	625	0.03	1.55	132	0.64	1.58				−693	0.02	1.96
AHRQCI_11					7,826	0.00	3,383	0.00	1.19	3,710	0.00	1.11				1,541	0.02	1.26
AHRQCI_12					1,624	0.00	1,503	0.00	1.02	1,482	0.00	1.02				1,413	0.00	1.02
AHRQCI_13					6,806	0.00	6,239	0.00	1.02	3,234	0.01	1.07				3,070	0.01	1.07
AHRQCI_14					6,010	0.00	5,205	0.00	1.01	2,460	0.00	1.12				2,205	0.00	1.12
AHRQCI_15					5,361	0.01	5,981	0.00	1.00	3,027	0.08	1.00				3,240	0.06	1.01
AHRQCI_16					24,603	0.00	20,065	0.00	1.00	3,827	0.12	1.96				2,677	0.23	1.96
AHRQCI_17					14,877	0.00	13,049	0.00	1.01	6,377	0.00	1.31				5,505	0.00	1.31
AHRQCI_18					29,468	0.00	25,095	0.00	1.01	2,975	0.01	1.17				1,981	0.07	1.18
AHRQCI_19					5,137	0.00	4,001	0.00	1.06	496	0.10	1.29				−88	0.76	1.34
AHRQCI_20					5,286	0.00	3,538	0.00	1.06	2,906	0.00	1.25				2,060	0.00	1.28
AHRQCI_21					6,856	0.00	4,251	0.00	1.01	6,049	0.00	1.01				3,756	0.00	1.01
AHRQCI_22					4,023	0.00	4,101	0.00	1.01	3,537	0.00	1.01				3,619	0.00	1.01
AHRQCI_23					5,009	0.00	2,988	0.00	1.01	2,778	0.00	1.01				2,240	0.00	1.01
AHRQCI_24					2,393	0.00	1,395	0.00	1.03	1,353	0.00	1.03				679	0.01	1.04
AHRQCI_25					2,392	0.00	2,168	0.01	1.01	1,479	0.05	1.01				1,521	0.04	1.01
AHRQCI_26					2,160	0.00	1,830	0.00	1.04	1,267	0.00	1.05				1,029	0.00	1.05
AHRQCI_27					1,448	0.05	1,197	0.08	1.07	1,273	0.07	1.07				1,188	0.08	1.07
AHRQCI_28					3,681	0.00	2,774	0.00	1.08	3,475	0.00	1.08				2,630	0.00	1.08
AHRQCI_29					4,504	0.00	2,855	0.00	1.09	4,091	0.00	1.05				2,689	0.00	1.09
AHRQCI_30					2,720	0.00	2,008	0.00	1.04	2,429	0.00	1.04				1,878	0.00	1.04
Mean VIF									1.15			1.22			1.17			1.21
Predicted cost	6,550.32		6,598.08		6,570.52		6,377.61			6,362.42			6,299.81			6,205.73		
Log-likelihood	−467,273.65		−467,620.46		−467,420.57		−465,997.91			−465,884.06			−465,411.91			−464,693.55		
AIC	19.58		19.59		19.58		19.52			19.52			19.50			19.47		
BIC	−467,920.10		−467,258.90		−467,496.90		−470,180.00			−470,375.20			−472,594.30			−471,481.20		
Pseudo R^2	0.0132		0.0127		0.0131		0.0161			0.0163			0.0174			0.0189		
ASPE	188,521		188,750		188,052		160,452			159,854			145,456			137,545		

VIF, variance inflation factor; AHRQCI, Agency for Healthcare Research and Quality's comorbidity index; AIC, Akaike information criteria; BIC, Bayesian information criteria; ASPE, averaged square prediction method.

burden to the unmeasurable part of the model. The resulting estimating coefficients may have omitted variable bias. Omitted variable bias exists when a relevant covariate is excluded from the model and has serious statistical consequences. In particular, estimated coefficients and the resulting conclusions for every covariate in the model may not be "trustworthy."

The second consequence of low correlation is related to multicollinearity, which refers to any linear relationship among covariates in a regression model. Investigators are often reluctant to use more than one comorbidity index in a single model because of possible multicollinearity. In the presence of multicollinearity, the estimate of one variable's impact on the outcome while controlling for others tends to be less precise than if model covariates were uncorrelated with one another [44]. When the goal of the model is a prediction as opposed to the estimation of the effect of a specific variable, one should not be concerned about multicollinearity between variables in the model. That said, even if we are interested in the effect of a specific variable, such as treatment effects, there is no harm in using these three indices simultaneously in a single model because there is no evidence of multicollinearity. If we exclude these indices for the sake of avoiding multicollinearity, we run the risk of omitted variable bias—a much bigger problem than multicollinearity.

Although the predictive validity of individual comorbidity measures has been examined in detail, this study is the first to evaluate the effects of using multiple indices within a single model. We used previously validated algorithms to create index measures, so that we can compare our single-measure models with previously published worked. As in previous studies, model results were inconclusive.

In general, summing the indicators to a single value is problematic, because this assumes that the effect in the original population is the same as in the study population. For example, it is unlikely that the effects of comorbid disease on the risk of 1-year mortality found in the original population used to develop the Charlson index (medical service patients seen at Cornell Medical Center, New York, NY) are the same as the effect on total expenditures (as used in our study of patients with migraines and triptan treatment). A similar argument can be made with CDS and AHRQCI. Because the current study has a relatively large sample size, we tested to see if the predictive power would be different with individual indicators rather than the summation scores—basically seven models with indicator functions (Table 5). The marginal effects of individual indicators vary widely. The predicted total costs ranged between \$6205 to \$6598 and were not significantly different from the predicted costs presented in Table 3 where summations of indices are used for risk adjustment. In sum, the results of these analyses do not change the conclusion of this study. In particular, there is no evidence of collinearity between the three types of indices and using these three unique indices in a single model statistically increases the predictive power of our estimations.

While the changes in AIC, BIC, ASPE, or pseudo R^2 values determine the ranking of our models in terms of predictive power, the actual difference in the predicted total cost determines what we might call practical significance. The differences can be statistically significant without being especially large. A statistically significant value being practically or clinically insignificant often occurs when working with large samples. Therefore, a discussion of the practical significance along with statistical significance of the estimates is appropriate. Results from this health expenditure analysis suggest that the difference is significant statistically, not practically.

The results of this study should be also be viewed in light of limitations that are inherent in retrospective claims data analysis. Correct categorization of insurance database information depends on correct codings by clinicians and other medical staff. The accuracy of diagnostic coding cannot be evaluated in a claims-based study. Patients may also receive treatment that is not submitted to their health plan for reimbursement and thus not included in claims data. Data on comorbidities were limited to conditions coded on medical claims within the time frame studied. Finally, the sensitivity analysis applied in this study does not utilize the full potential of the database. Ideally, one would perform variable selection procedure and the use of predictive modeling based on a data mining process using SAS Enterprise Miner (SAS Institute Inc., Cary, NC), and then investigate the results via model comparison.

While more work is warranted to evaluate these findings can be supported in other population's theses results are, nevertheless, important to pharmacoeconomical researchers for several reasons. First, we analyzed the use of three common comorbidity indices (CCI, CDS, and AHRQCI) as methods of risk adjustment in a sample of triptan-using migraine patients and found that each measures different risks. Second, we demonstrated that the collinearity of the three measures was not strong enough to prevent including them simultaneously in the same model. Finally, we showed that summing the indicators into a single value ignores the variations in the effect of each indicator on healthcare expenditures but the estimated risk adjusted expenditures are not different from those estimated using summation scores.

Supplementary material for this article can be found at: http://www.ispor.org/publications/value/ViHsupplementary.asp

References

1 Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. J Clin Epidemiol 1993;46:1075–9.
2 Romano PS, Roos LL, Jollis JG. Further evidence concerning the use of a clinical comorbidity index with ICD9-CM administrative data. J Clin Epidemiol 1993;46:1085–90.
3 Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992;45:613–19.
4 D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. J Clin Epidemiol 1996;49:1429–33.
5 Ghali WA, Hall RE, Rosen AK, et al. Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. J Clin Epidemiol 1996;49:273–8.
6 Zhang JX, Iwashyna TJ, Christakis NA. The performance of different lookback periods and sources of information for Charlson comorbidity adjustment in Medicare claims. Med Care 1999;37:1128–39.
7 Needham DM, Scales DC, Laupacis A, Pronovost PJ. A systematic review of the Charlson comorbidity index using Canadian administrative databases: a perspective on risk adjustment in critical care research. J Crit Care 2005;20:12–19.
8 Cleves MA, Sanchez N, Draheim M. Evaluation of two competing methods for calculating Charlson's comorbidity index when analyzing short-term mortality using administrative data. J Clin Epidemiol 1997; 50:903–8.
9 Miskulin DC, Martin AA, Brown R, et al. Medical Directors Dialysis Clinic Inc. Predicting 1 year mortality in an outpatient haemodialysis population: a comparison of comorbidity instruments. Nephrol Dial Transplant 2004;19:413–20.
10 Di Bari M, Virgillo A, Matteuzzi D, et al. Predictive validity of measures of comorbidity in older community dwellers: the Insufficienza Cardiacanegli Anziani Residenti a Dicomano Study. J Am Geriatr Soc 2006; 54:210–16.
11 Fried L, Bernardini J, Piraino B. Comparison of the Charlson Comorbidity Index and the Davies score as a predictor of outcomes in PD patients. Perit Dial Int 2003;23:568–73.
12 Dominick KL, Dudley TK, Coffman CJ, Bosworth HB. Comparison of three comorbidity measures for predicting health service use in patients with osteoarthritis. Arthritis Rheum 2005;53:666–72.
13 Soares M, Salluh JI, Ferreira CG, et al. Impact of two different comorbidity measures on the 6-month mortality of critically ill cancer patients. Intensive Care Med 2005;31:408–15.
14 McGregor JC, Kim PW, Perencevich EN, et al. Utility of the Chronic Disease Score and Charlson Comorbidity Index as comorbidity measures for use in epidemiologic studies of antibiotic-resistant organisms. Am J Epidemiol 2005;161:483–93.
15 Gabbe BJ, Magtengaard K, Hannaford AP, Cameron PA. Is the Charlson Comorbidity Index useful for predicting trauma outcomes? Acad Emerg Med 2005; 12:318–21.
16 Yan Y, Birman-Deych E, Radford MJ, et al. Comorbidity indices to predict mortality from Medicare data: results from the national registry of atrial fibrillation. Med Care 2005;43:1073–7.
17 Schneeweiss S, Wang PS, Avorn J, et al. Consistency of performance ranking of comorbidity adjustment scores in Canadian and U.S. utilization data. J Gen Intern Med 2004;19:444–50.
18 Wang PS, Walker A, Tsuang M, et al. Strategies for improving comorbidity measures based on Medicare and Medicaid claims data. J Clin Epidemiol 2000; 53:571–8.
19 Clark DO, Von Korff M, Saunders K, et al. A chronic disease score with empirically derived weights. Med Care 1995;33:783–95.
20 Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992;45:197–203.
21 Wahls TL, Barnett MJ, Rosenthal GE. Predicting resource utilization in a Veterans Health Administration primary care population. Comparison of methods based on diagnoses and medications. Med Care 2004;42:123–8.
22 Holman CD, Preen DB, Baynham NJ, et al. A multipurpose comorbidity scoring system performed better than the Charlson index. J Clin Epidemiol 2005; 58:1006–14.
23 Martins M, Blais R. Evaluation of comorbidity indices for inpatient mortality prediction models. J Clin Epidemiol 2006;59:665–9.
24 Rius C, Perez G, Martinez JM, et al. An adaptation of Charlson comorbidity index predicted subsequent mortality in a health survey. J Clin Epidemiol 2004; 57:403–8.
25 Chaudhry S, Jin L, Meltzer D. Use of a self-report-generated Charlson Comorbidity Index for predicting mortality. Med Care 2005;43:607–15.
26 Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. Int J Epidemiol 2000;29:891–8.
27 Perkins AJ, Kroenke K, Unutzer J, et al. Common comorbidity scales were similar in their ability to predict healthcare costs and mortality. J Clin Epidemiol 2004;57:1040–8.
28 Farley JF, Harley CR, Devine JW. A comparison of comorbidity measurements to predict healthcare expenditures. Am J Manag Care 2006;12:110–19.
29 de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. A critical review of available methods. J Clin Epidemiol 2003;56:221–9.
30 Johnson RE, Hornbrook MC, Nichols GA. Replicating the chronic disease score (CDS) from automated pharmacy data. J Clin Epidemiol 1994;47: 1191–9.
31 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40:373–83.

32 Warshaw LJ, Burton WN. Cutting the costs of migraine: role of the employee health unit. JOEM 1998;40:943–53.
33 Stang PE, Osterhaus JT, Celentano DD. Migraine. Patterns of healthcare use. Neurology 1994;34:344–50.
34 De Lissovoy G, Lazarus SS. The economic cost of migraine. Present state of knowledge. Neurology 1994;44(6 Suppl. 4):S56–62.
35 Charlson ME, Szatrowski TP, Peterson J, et al. Validation of a combined comorbidity index. J Clin Epidemiol 1994;47:1245–51.
36 Petersen LA, Pietz K, Woodard LD, Byrne M. Comparison of the predictive validity of diagnosis-based risk adjusters for clinical outcomes. Med Care 2005;43:61–7.
37 Stukenborg GJ, Wagner DP, Connors AF Jr. Comparison of the performance of two comorbidity measures, with and without information from prior hospitalizations. Med Care 2001;39:727–39.
38 Healthcare cost and Utilization Project (HCUP) US Tools and Software Page. Comorbidity Software, Version 3.0. Available from: http://www.hcupus.EI.gov/toolssoftware/comorbidity/comorbidity.jsp [Accessed August 24, 2006].
39 Elixhauser A, Steiner C, Harris DR. Comorbidity measures for use with administrative data. Med Care 1998;36:8–27.
40 Chatterjee S, Hadi AS, Price B. Regression Analysis by Example (3rd ed.). New York: John Wiley and Sons, 2000.
41 Manning WG, Mullahy H. Estimating log models. To transform or not to transform? J Health Econ 2001; 20:461–94.
42 Duan N, Manning WG, Willard G, et al. A comparison of alternative models for the demand for medical care. J Bus Econ Stud 1983;1:115–26.
43 Duan N. Smearing estimate: a nonparametric retransformation method. J Am Stat Assoc 1983;78:605–10.
44 Wooldridge JM. Introductory Econometrics: A Modern Approach. Cincinnati, OH: South-Western, 2001.