ELSEVIER

Journal of Clinical Epidemiology 57 (2004) 1040–1048

Journal of Clinical Epidemiology

Common comorbidity scales were similar in their ability to predict health care costs and mortality

Anthony J. Perkins[a,b,*], Kurt Kroenke[b], Jürgen Unützer[c], Wayne Katon[d], John W. Williams, Jr.[e], Carol Hope[b,f], Christopher M. Callahan[a,b]

[a]*Indiana University Center for Aging Research, 1050 Wishard Boulevard, RG6, Indianapolis, IN 46202, USA*
[b]*Regenstrief Institute, Inc., Indianapolis, IN, USA*
[c]*Center for Health Services Research, UCLA Neuropsychiatric Institute, Los Angeles, CA, USA*
[d]*Department of Psychiatry, University of Washington, Seattle, WA, USA*
[e]*Center for Health Services Research in Primary Care, Durham Veterans Affairs Medical Center and Duke University Medical Center, Durham, NC, USA*
[f]*Purdue School of Pharmacy, West Lafayette, IN, USA*

Accepted 3 March 2004

Abstract

Objective: To compare the ability of commonly used measures of medical comorbidity (ambulatory care groups [ACGs], Charlson comorbidity index, chronic disease score, number of prescribed medications, and number of chronic diseases) to predict mortality and health care costs over 1 year.

Study Design and Setting: A prospective cohort study of community-dwelling older adults (n = 3,496) attending a large primary care practice.

Results: For predicting health care charges, the number of medications had the highest predictive validity (R^2 = 13.6%) after adjusting for demographics. ACGs (R^2 = 16.4%) and the number of medications (15.0%) had the highest predictive validity for predicting ambulatory visits. ACGs and the Charlson comorbidity index (area under the receiver operator characteristic [ROC] curve = 0.695–0.767) performed better than medication-based measures (area under the ROC curve = 0.662–0.679) for predicting mortality. There is relatively little difference, however, in the predictive validity across these scales.

Conclusion: In an outpatient setting, a simple count of medications may be the most efficient comorbidity measure for predicting utilization and health-care charges over the ensuing year. In contrast, diagnosis-based measures have greater predictive validity for 1-year mortality. Current comorbidity measures, however, have only poor to moderate predictive validity for costs or mortality over 1 year.

Keywords: Comorbidity; Utilization; Mortality; Chronic disease

1. Introduction

Researchers and others often seek to compare the burden of illness or outcomes of care across large populations. For example, to compare the health care costs or mortality due to a specific condition across two or more groups of patients, researchers often need to test and/or adjust for possible differences in comorbidity between groups. Most current measures of comorbidity rely on an accounting of the subjects' other diagnosed medical conditions and, in some commonly used approaches, an accounting of the severity of those conditions. Because of the large number of subjects and vagaries associated with medical records, it is often impractical to assess the burden of illness through manual review of individual written medical records. One solution to this problem has been to use diagnostic codes (e.g., the International Classification of Disease ICD codes) contained in administrative or clinical databases to describe the burden of illness among large cohorts.

Comorbidity refers to the total burden of illnesses across multiple potential conditions unrelated to the patient's principal or target diagnosis [1]. Severity of illness, by contrast, typically refers to a single target diagnosis. Different approaches have been used to define comorbidity, depending on the outcome measure, the clinical setting, and source of data. For inpatient settings, manual chart review and laboratory tests have been used to calculate comorbidity or severity scores [2–7], but community-based survey measures have often relied on self-reported symptoms, diagnoses,

* Corresponding author. Tel.: 317-630-7588; fax: 317-630-6611.
E-mail address: tperkins@regenstrief.org (A.J. Perkins).

0895-4356/04/$ – see front matter
doi: 10.1016/j.jclinepi.2004.03.002

quality of life, perceived health, or functional status to assess comorbidity and severity of illness [7,8]. Other strategies rely on administrative databases to calculate comorbidity using diagnostic codes [9–16] or prescriptions [16,17]. Measures can be simple (e.g., count of symptoms or diseases) or quite complex (e.g., using different weighting schemes). The choice and validity of the measure depends on the outcome of interest (e.g., mortality, utilization, costs, and treatment- specific variables), the efficiency in calculating the measure, and the availability and reliability of the data used to create the measure.

Although the predictive validity of individual comorbidity measures has been examined in detail [1–25], few published studies compare different measures within the same target population. Those studies that have compared predictive validity across measures tend to focus on comparisons in limited disease-specific populations (e.g., patients with cancer) [1,6,8,21–22]. Others have compared different variants of the same index [11–14] or compared scores for data derived from written medical records as opposed to administrative databases [10,20]. There are few published studies comparing the diagnosis-based with medication-based measures [18,24–26], and only three used the same population for these comparisons [18,25,26]. Furthermore, much current research applies the available comorbidity measures, or variations in the scoring of these measures, to patient populations in whom the measure has not been tested. For example, a research team might use a comorbidity measure developed to predict inpatient mortality to control for the impact of comorbidity on health care utilization among an outpatient population.

Our purpose in the present study was to compare the predictive validity of commonly used measures of medical comorbidity among a large cohort of vulnerable older adults cared for in a single primary care practice that is served by a state-of-the-art electronic medical record system. Because we are interested in measures that can be used with electronic medical records or administrative databases, we limited the comparison to those measures that can be obtained by using data from such sources. We were particularly interested in determining if more complex measures predict health care charges and mortality better than simpler measures. We also examined the impact of modifications in the scoring on the performance of these scales.

2. Overview of common comorbidity measures

Ambulatory care groups (ACGs) were designed to provide a case-mix adjustment measure for the ambulatory care setting [15–16]. ACGs are based on the premise that a measure of a population's illness burden can help explain variation in health care resource consumption. Using primary and secondary diagnoses (ICD-9 codes) from ambulatory visits over a defined time period, patients are first classified into one of 34 ambulatory diagnostic groups (ADGs) by clustering similar conditions based on their expected impact on health services resource consumption. Patients are then classified into one of 51 distinct ACGs, based on age, sex, and constellation of ADGs. Weiner et al. [16] showed that ACGs explained more than 50% of ambulatory resource use retrospectively and approximately 20% prospectively. One disadvantage of this scoring system is that some ACG categories will have very small numbers. Analysts may have to combine categories to avoid computational difficulties. In addition, the large number of categories makes it more difficult to make comparisons of comorbidity across groups.

Another measure that uses ICD-9 codes is the Charlson comorbidity index [2–3]. This index was first developed in an inpatient setting, using medical record review to predict risk of mortality. The index assigns nonzero weights to 19 conditions (all other conditions are given a score of 0), based on their risk of mortality. The weights can take on values of 1, 2, 3, or 6 and the weights are then summed for each patient. There have been many variations adapting the index for use with administrative databases by mapping ICD-9 codes to the 19 comorbid conditions defined by the original scale [9,11–14]. Here we use the adaptation of Deyo et al. [9]. The Charlson index has been shown to predict mortality [9–14,19–20,25] and is associated with length of stay and hospital charges [9,19,25]. Notably, when used among cohorts assembled from ambulatory settings, those subjects without a hospitalization in the prior year would have a score of zero. Some investigators have resolved this problem by allowing outpatient diagnosis to contribute to the score.

The final measure using ICD-9 codes is a simple count of comorbid conditions. The outcome is calculated by simply summing the comorbid conditions (usually limited to chronic conditions) over a defined time period. The number of comorbid conditions has been shown to predict mortality [2]. One face-validity concern with this method is that it assumes that different diagnoses (e.g., stroke, cancer) or differing levels of severity within diagnoses are similar with respect to the outcome of interest [2], and that, for example, a patient with hypertension and diabetes has twice the comorbidity as a patient with only a stroke.

A comorbidity measure that is based on prescribed medications is the chronic disease score (CDS). The CDS [17,18] increases with the number of different chronic diseases as inferred from the subject's medication profile. Medications used in the management of acute diseases (e.g., antibiotics) or common symptoms (e.g., nasal congestion) are not included in the scoring. Individual medications are mapped to medication classes, which are then mapped to different chronic diseases. Weights are assigned to each CDS class, based on expert judgment. The original CDS was calculated by summing the weights of each unique CDS class for each patient [17]. A revised version of the CDS uses empirical weights to calculate the CDS [18]. Both scores were shown to predict mortality and health care resource utilization after adjusting for demographics and previous resource utilization. The main disadvantage of the CDS measure is the

need for continual updates as new medications are introduced and the limitation in defining comorbidity as only chronic diseases treated with prescription medications.

Notably, these measures are often applied to source data, populations, or outcomes outside the scope of the original developmental work for the index. For example, the Charlson comorbidity index might be used to measure utilization in an outpatient population of older adult patients with depression by using administrative data. Although this application may have face validity, there is limited literature to assess the true predictive validity of the adaptation.

3. Methods

Community-dwelling patients at least 60 years of age with a scheduled primary care appointment between July 15, 1999, and August 31, 2001, were eligible for the study. During this time period, 4,386 patients had a scheduled visit and 3,496 (80%) completed a screen. Of the 880 people who were not seen, 499 (56%) did not show up for any of their scheduled appointments during the time period. Additionally, 111 (12%) were missed by the recruiter, 105 (12%) refused, and 175 (20%) were ineligible. All data for the study were obtained from the Regenstrief Medical Record System (RMRS) [27]. The RMRS is a comprehensive electronic medical record system composed of different modules. The pharmacy module contains information on dispensing and charges of all medications by the inpatient and outpatient pharmacy. The RMRS also maintains a number of other databases including diagnoses, vital signs, results of laboratory and diagnostic tests, full-text discharge summaries, preventive health maneuvers, and detailed information on all inpatient and outpatient charges. It also contains death certificate information from the Indiana State Board of Health for all registered patients who die in, or outside of, Indiana.

We obtained all diagnostic ICD-9 codes for all ambulatory visits and inpatient stays for the year prior to the index visit. Programs written in SAS (version 8.02; SAS Institute, Cary, NC, USA) were used to calculate ACG categories using outpatient data only and the Charlson comorbidity index using inpatient data only, according to methods described by the developers. In addition, we calculated modified Charlson and ACG scores by using both inpatient and outpatient data. To calculate a simple count of diagnoses, we summed the number of 10 common chronic conditions (arthritis, coronary artery disease, cancer, congestive heart failure, chronic obstructive pulmonary disease, diabetes, hypertension, liver disease, renal disease, and stroke). We chose these conditions both because of their prevalence and because they are conditions typically included on community-based surveys collecting data on self-reported diagnoses.

In addition to obtaining diagnostic codes from the electronic medical record, we also obtained all prescription medications from the pharmacy database for the prior year. Medical devices, durable medical equipment, and over-the-counter medications were excluded. To address newer medications not included in the original CDS, we updated and revised the CDS categories applying a strategy used by the original developers. A pharmacist classified each medication into a pharmacy class and subclass. The pharmacy subclasses were then mapped to a chronic disease class. We had six different physician–scientists representing general internal medicine, geriatrics, and psychiatry independently apply weights to each of the chronic conditions, based on the likely impact on health services use over 1 year. Raters could apply a weight of 0–3,with 3 indicating a condition that would result in the most health care utilization. Two different weighting systems were designed, one to predict health care costs (CDS–HC) and the other attempting to predict mortality (CDS–M). In this manner, a chronic disease could be assigned a different weight in terms of its expected impact on utilization than its expected impact on mortality. The final Chronic Disease Score is calculated by summing the weights for each unique chronic disease class represented in the medication profile of each patient.

We also examined the predictive validity of several simpler variants of the CDS. The overall goal of these variations was to determine to what extent the predictive validity improved with the more complex weighting schemes and categorizations. The first variations are the pharmacy subclass scores (PSS–HC for health care costs and PSS–M for mortality), which are calculated by summing the weight for each unique pharmacy subclass rather than lumping these within a chronic disease. Additionally we specified the unweighted count of chronic diseases classes, an unweighted count of pharmacy subclasses, and a simple count of the prescription medications as measures of comorbidity. Table 1 provides a representation of how the score varies across such categories.

All outpatient and inpatient visits and associated charges for the year following the index visit were obtained from the electronic medical records system. Total charges were defined as the sum of all charges from outpatient and inpatient visits. Utilization was also obtained for the year prior to the index date to be used in our sensitivity analyses. The number of inpatient visits was recoded as any inpatient stay versus no inpatient stay. The number of ambulatory visits was defined as the sum of all primary care visits, medical specialty visits, surgical specialty visits, and emergency room visits. In addition, death during the year following the index visit and date of death was assessed using the latest version of death certificate information from the Indiana State Board of Health for all registered patients who died (in or outside of Indiana).

We transformed the cost data by adding 1 to all values and then taking the natural logarithm. To assess how the different chronic disease measures varied across demographic measures, *t*-tests and correlations were used. Spearman rank correlations were used to assess agreement across the different measures of comorbidity. We used linear regression and analysis of covariance to assess the variation explained by each comorbidity measure on ambulatory visits

Table 1
Variations in calculation of comorbidity scores based on medications for two hypothetical patients

Patient	Drug	Pharmacologic class	Pharmacologic subclass[a]	Chronic disease[a]
Patient 1				
	atenolol	cardiovascular agent	beta adrenergic blocking agent	hypertension
	benazepril	cardiovascular agent	angiotensin-converting enzyme inhibitor	congestive heart failure
	losartan	cardiovascular agent	angiotensin II receptor antagonist	congestive heart failure
	Score HC = 3		Score HC = 5	Score HC = 3
Patient 2				
	isosorbide	cardiovascular agent	coronary vasodilator	coronary artery disease
	donepezil		cholinesterase inhibitor	Alzheimer disease
	celecoxib	anti-inflammatory	COX II inhibitor	osteoarthritis
	nitroglycerin	cardiovascular agent	coronary vasodilator	coronary artery disease
	Score HC = 4		Score HC = 6	Score HC = 6

[a] Weighted score.

and total cost. Logistic regression was used to assess how well each comorbidity measure predicted any inpatient stay or death.

4. Results

Table 2 presents demographic, utilization, and comorbidity measures for the 3,496 patients. The average age was 68.9 years; 2,418 (69%) patients were women and 1,959 (56%) were African American. Approximately 75% of the patients had Medicare and 959 (27%) had Medicaid. Physicians identified 709 (20%) patients as smokers and 1,581 (45%) patients were obese. This group of patients averaged eight ambulatory visits per year and total health care charges of about $6,600 in the year following the index visit; 785 (22%) had an inpatient stay. Patients averaged 7 medications, 6.4 unique pharmacy subclasses, and 5.2 chronic disease classes.

Figure 1 shows the distribution of each comorbidity score. The CDS–HC has the widest variation, with a range of 0 to 24 and the mode occurring at 7. The other medication-based measures had similar distributions. In this sample, the Charlson comorbidity index approaches a dichotomous measure of whether the patient had a hospitalization, with 2,889 (83%) of patients having a Charlson comorbidity index of 0. The Charlson comorbidity index calculated using both inpatient and outpatient data has fewer scores of 0, but the majority of scores are 2 or less (data not shown). The number of chronic diseases has a similar range (0–7) and distribution to the Charlson comorbidity index when that index is modified to include both inpatient and outpatient data. The ACGs have a wider range of values than the other diagnostic measures, with the mode occurring in the group represented by 2–3 ADG combinations.

African Americans had significantly lower medication comorbidity scores and a lower Charlson comorbidity index than whites; however, they had significantly more diagnosed chronic diseases. Women had significantly higher scores than men for all comorbidity measures except the number of chronic diseases and ACGs. Patients with Medicare or Medicaid and smokers had significantly higher comorbidity scores for all measures. Although low-income patients had a significantly lower Charlson comorbidity index and fewer chronic diseases, they tended to have higher comorbidity scores when the medication-derived measures were used. Increasing levels of obesity (in terms of the body mass index, BMI) were associated with increasing levels of comorbidity for all measures except the Charlson comorbidity index and ACGs. Increasing age was associated with decreased comorbidity for all measures except ACGs (not significant) and the number of chronic diseases (associated with increasing comorbidity). All comorbidity measures were significantly correlated with the multiple measures of utilization during the prior year.

Spearman rank correlations for all comorbidity measures are given in Table 3. The medication-derived comorbidity measures were highly correlated, as expected. The medication-derived measures were moderately correlated with each of the diagnostic measures with the correlations ranging from 0.41 to 0.53. The diagnostic measures were moderately correlated with one another, but not as highly correlated as the medication measures. The original and modified ACG measures were highly correlated ($r = 0.95$), but the original and modified Charlson measures were only moderately correlated ($r = 0.57$).

Regression models showed that increasing comorbidity on each measure predicted both log of total charges and number of outpatient visits. All measures remained significant even after adjusting for demographic factors and utilization in the year prior to the index visit. The percent variation explained in the logarithm of total charges and the number of ambulatory visits by each comorbidity measure are given in Table 4.

For log of total charges in the unadjusted models, the number of medications performed best, closely followed by the number of pharmacy subclasses and ACGs (inpatient and outpatient data), with each model predicting approximately 13% of the variation in charges. The variation explained by PSS and CDS ranged from 10% to 12%; the least variance was explained by the Charlson comorbidity index and the

Table 2
Study population demographics, utilization, and comorbidity

	Mean or % (SD) (N = 3,496)
Age, years	68.9 (7.0)
African American, %	56.0
Female, %	69.2
Current smoker, %	20.3
Medicare, %	74.5
Medicaid, %	27.4
County insurance plan for medically indigent, %	53.1
BMI	
≤19, %	6.1
20 24, %	20.6
25 29, %	28.1
≥30, %	45.2
Utilization, prior year	
Any hospital stay, %	22.4
Charges, $	6,626 (18,237)
Ambulatory visits, no.	8.0 (6.6)
ER visits, no.	1.1 (1.8)
Primary care visits, no.	3.9 (2.9)
Died, %	2.8
Comorbidity measures	
CDS HC score	6.9 (4.6)
PSS HC score	7.5 (5.4)
CDS M score	5.4 (3.8)
PSS M score	5.7 (4.3)
Unique CDS classes, no.	5.2 (3.3)
Unique pharmacy subclasses, no.	6.4 (4.4)
Prescription medications, no.	7.0 (5.1)
Chronic diseases, no.	2.2 (1.3)
Charlson comorbidity index	0.4 (1.1)
Modified Charlson index, both inpatient and outpatient	1.1 (1.5)
ACGs	
Acute minor, %	5.0
Acute major, %	6.1
Chronic, %	12.0
2 3 ADG combinations, %	29.8
4 5 ADG combinations, 0 1 major ADGS, %	19.9
4 5 ADG combinations, ≥2 major ADGS, %	4.5
6 9 ADG combinations, 0 1 major ADGS, %	10.0
6 9 ADG combinations, ≥2 major ADGS, %	9.3
≥10 ADG combinations, 0 3 major ADGS, %	2.8
≥10 ADG combinations, ≥4 major ADGs, %	0.7
Modified ACGs, both inpatient and outpatient	
Acute minor, %	4.9
Acute major, %	5.7
Chronic, %	12.0
2 3 ADG combinations, %	27.4
4 5 ADG combinations, 0 1 major ADGS, %	18.0
4 5 ADG combinations, ≥2 major ADGS, %	3.9
6 9 ADG combinations, 0 1 major ADGS, %	9.4
6 9 ADG combinations, ≥2 major ADGS, %	11.0
≥10 ADG combinations, 0 3 major ADGS, %	4.8
≥10 ADG combinations, ≥4 major ADGs, %	2.9

Abbreviations: ACG, ambulatory care group; ADG, ambulatory diagnostic group; BMI, body mass index; CDS, chronic disease score; HC, health care; M, mortality; PSS, pharmacy subclass score.

number of chronic diseases (5.5%). After adjusting for age, race, and gender, the rank ordering of the scores remained the same. After adjusting for demographics and prior utilization, the number of medications, the number of pharmacy subclasses, and both pharmacy subclass score models performed best, explaining more than 16% of the variation. Using inpatient and outpatient data to compute modified Charlson and ACGs scores resulted in better-fitting models.

For number of ambulatory visits in the unadjusted models, ACGs and the number of medications performed best, with each model predicting approximately 15% of the variation in charges. The worst-performing measures were the CDS–M (8.0%), the Charlson comorbidity index (4.7%), and the number of chronic diseases (3.7%). After adjusting for age, race, and gender, the rank ordering remained similar. After adjusting for demographics and prior utilization, the number of prescription medications, the number of pharmacy subclasses, and ACGs were the best models, explaining approximately 19% of the variation. Addition of outpatient data to calculate the modified Charlson score resulted in better-fitting models, however, the addition of inpatient data to calculate the ACG score resulted in worse fitting models.

Logistic regression analysis showed that increasing comorbidity is significantly associated with increased odds of having an inpatient admission during the following year, as well as increased odds of death. All measures remained significant after adjusting for demographic factors and prior utilization. The area under the receiver operator characteristic (ROC) curve from the logistic regression models is given in Table 5.

In predicting inpatient admission, the modified ACGs model has the highest ROC value, at 0.678, followed by ACGs (0.663), the number of medications (0.659), the number of pharmacy subclasses (0.658), and the PSS–M (0.656). The Charlson comorbidity index has the lowest value, at 0.608. After adjusting for age, race, and gender, the rank ordering of the models remained similar. After adjusting for demographics and prior utilization, the pharmacy subclass scores were the best-predicting models, followed by the CDS–Mortality, the number of medications, and the number of pharmacy subclasses. Using inpatient and outpatient data to compute modified Charlson and ACGs scores resulted in better-fitting models.

In predicting mortality, the diagnostic measures performed much better than the medication-based measures. The best model was the ACG score calculated using inpatient and outpatient data (area under the ROC curve = 0.739), followed by the Charlson comorbidity index calculated using inpatient and outpatient data (0.731). After adjusting for age, race, and gender, the rank ordering of the models was similar. After adjusting for demographics and prior utilization, the diagnostic measures were the best models (area under the ROC curve 0.763–0.792); among these, the Charlson comorbidity index calculated using both inpatient and outpatient data performed best.

We recalculated the measures in Tables 4 and 5, categorizing the continuous measures of comorbidity (all measures

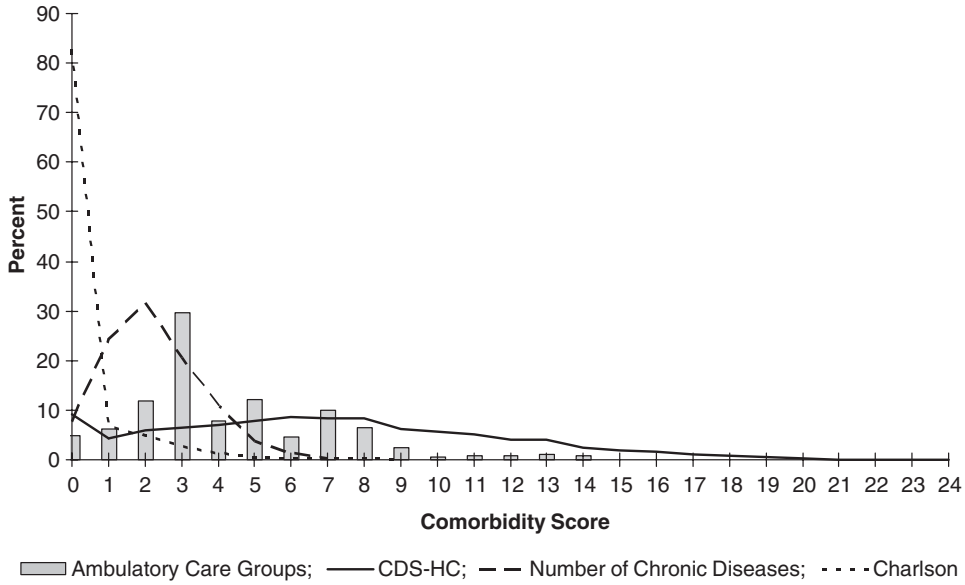

Fig. 1. Distribution of scores on comorbidity measures. Bars: ambulatory care groups. Solid line: Chronic Disease Score. Dashed line: number of chronic diseases. Dotted line: Charlson comorbidity index.

except ACGs) (data not shown). Categorizing the variables only slightly improved the fit of each model for these comorbidity measures and did not change the relative performance ratings. The last reclassification involved putting in classification variables representing each of the 10 chronic diseases instead of using the number of diseases. This reclassification improved the model fit. For prediction of any inpatient admission and death, this model became the second-best model for predicting inpatient admission and the best for predicting death.

5. Discussion

We compared the predictive validity of several comorbidity measures that can be calculated using electronic medical records or administrative databases on four different outcomes. In predicting total charges and the number of ambulatory visits over 1 year, the comorbidity indices based on medications or outpatient diagnoses performed better than the Charlson index (which was developed to predict inpatient mortality) or the total number of chronic conditions. For predicting hospitalization and mortality over 1 year, the ACGs performed the best, but there was little variation across all of the measures. A simple count of medications performed as well as the more complex weighting schemes in predicting each of the four primary outcomes.

In examining these findings, we are aware of the potential for circular reasoning. If we assume that prior utilization is the best predictor of future utilization, each of these measures can be viewed as a direct measure of prior utilization. The

Table 3
Spearman correlations of the CDS weighted measures with other measures of comorbidity

	Medication-based measures							Diagnosis-based measures			
	CDS HC	PSS HC	CDS M	PSS M	CDS classes, no.	Pharmacy classes, no.	Medications, no.	Charlson index	Modified Charlson index[a]	Chronic Diseases, no.	ACGs
CDS HC	1.000										
PSS HC	0.958	1.000									
CDS M	0.960	0.907	1.000								
PSS M	0.928	0.959	0.953	1.000							
CDS classes, no.	0.962	0.936	0.905	0.890	1.000						
Pharmacy classes, no.	0.914	0.948	0.851	0.893	0.961	1.000					
Medications, no.	0.907	0.943	0.845	0.889	0.951	0.992	1.000				
Charlson index	0.315	0.339	0.325	0.352	0.322	0.363	0.370	1.000			
Modified Charlson index[a]	0.412	0.438	0.422	0.455	0.395	0.426	0.432	0.569	1.00		
Chronic diseases, no.	0.506	0.485	0.534	0.517	0.479	0.459	0.456	0.325	0.514	1.000	
ACGs	0.427	0.453	0.397	0.425	0.461	0.502	0.512	0.428	0.485	0.351	1.000
Modified ACGs[a]	0.433	0.459	0.407	0.436	0.466	0.509	0.520	0.561	0.533	0.364	0.946

Abbreviations: ACG, ambulatory care group; CDS, chronic disease score; HC, health care; M, mortality; PSS, pharmacy subclass score.
[a] Based on both inpatient and outpatient data.

Table 4
Percent variation explained by each comorbidity score for log (total charges) and ambulatory visits

	Log(total charges)		Ambulatory visits	
	R^2	Adjusted[a] R^2	R^2	Adjusted[a] R^2
CDS HC	10.5	10.6	9.8	10.3
PSS HC	12.4	12.5	11.8	12.2
CDS M	9.7	9.8	8.0	8.8
PSS M	11.7	11.9	10.0	10.7
CDS classes, no.	10.9	11.0	11.6	12.0
Pharmacy subclasses, no.	13.2	13.3	14.3	14.6
Medications, no.	13.4	13.6	14.7	15.0
Hospitalization in prior year	5.6	5.9	4.5	5.8
Charlson comorbidity index	5.5	5.7	4.7	5.9
Modified Charlson[b]	8.0	8.2	6.9	8.0
Chronic diseases, no.	5.5	5.8	3.7	5.0
ACGs, collapsed	12.0	12.1	15.8	16.4
Modified ACGs[b]	13.0	13.2	14.1	14.9

Abbreviations: ACG, ambulatory care group; CDS, chronic disease score; HC, health care; M, mortality; PSS, pharmacy subclass score.

[a] Adjusted for age, race, and gender.

[b] Based on both inpatient and outpatient data.

more times patients visit a physician, the more likely they are to accumulate diagnoses and medications. A dichotomous indicator of hospitalization in the past year predicts hospitalization or mortality in the coming year about as well as more complex measures. In concluding that the measures are remarkably similar in performance despite major differences in data input and scoring, we could also conclude they are similar in their limited capacity to explain future variation in health care use or mortality.

These data also suggest that indices designed to measure inpatient mortality, for example, should not be expected to be excellent measures of outpatient utilization. The data presented here are not intended to support the use of a measure in a population or for an outcome beyond the intended purpose. Note, however, that these comorbidity measures are nonetheless used in this fashion in the extant literature. Our data show how the predictive validity of the instrument changes as a function of the type of data inputs used to calculate scores. In many cases, the predictive validity improves substantially with adjustment for age, gender, race, and prior health services utilization. The predictive validity of both the Charlson comorbidity index and the ACGs for hospitalization or 1-year mortality improved when the data source included both inpatient and outpatient diagnostic codes. Although all of the measures except ACGs were treated as continuous variables in the analyses, recoding the continuous measures to be categorical gave similar results.

All comorbidity measures significantly predict increasing total charges, increasing number of ambulatory visits, and the increasing probability of an inpatient admission or death, even after adjusting for demographic factors and prior utilization. The observed values of R^2 for total cost and number of outpatient visits adjusting for age, race, and gender were similar to or slightly higher than reported values from models adjusting for age and gender in previous studies [18,25]. The areas under the ROC curve for predicting any inpatient admission were similar to the values obtained in an earlier study [25]. The areas under the ROC curve for the medication-based measures were lower than had been reported [25]; the areas under the ROC curve for the Charlson comorbidity index were similar to those from previous studies [11,13,25].

Previous studies examining the predictive validity of different comorbidity measures have focused on comparing measures (usually chart-review measures) in limited disease-specific populations [1,6,8,21–22] or on comparing different variations of the Charlson comorbidity index [11–14]. Only three published studies [18,25,26], all using administrative

Table 5
Area under the ROC curve by each comorbidity score for inpatient admission and mortality

	Any inpatient stay		Mortality	
	Area under the ROC curve	Area under the ROC curve, adjusted[a]	Area under the ROC curve	Area under the ROC curve, adjusted[a]
CDS HC	0.638	0.648	0.612	0.662
PSS HC	0.651	0.661	0.619	0.672
CDS M	0.643	0.651	0.625	0.666
PSS M	0.656	0.665	0.626	0.678
CDS classes, no.	0.637	0.649	0.621	0.670
Pharmacy subclasses, no.	0.658	0.669	0.634	0.679
Medications, no.	0.659	0.670	0.636	0.677
Previous hospitalization	0.619	0.643	0.642	0.692
Charlson comorbidity index	0.608	0.636	0.656	0.719
Modified Charlson index[b]	0.648	0.664	0.731	0.755
Chronic diseases, no.	0.633	0.643	0.659	0.695
ACGs, collapsed	0.663	0.672	0.713	0.744
Modified ACGs[b]	0.678	0.685	0.739	0.767

Abbreviations: ACG, ambulatory care group; CDS, chronic disease score; HC, health care; M, mortality; PSS, pharmacy subclass score; ROC, receiver operator characteristic.

[a] Adjusted for age, race, and gender.

[b] Based on both inpatient and outpatient data.

databases, have compared medication-based and diagnosis-based measures. Clark et al. [18] compared ACGs with the original CDS and revised CDS. They showed that the revised CDS and ACGs performed similarly with respect to predicting charges and primary care visits. Fishman et al. [26] compared ACGs, hierarchical coexisting conditions (HCCs), and RxRisk (a medication-based measure) in a population of adults and children. They showed that HCCs performed best ($R^2 = 15.4\%$), followed by ACGs (10.1%) and RxRisk (8.7%). Schneeweiss et al. [25] studied similar outcome and comorbidity measures in a sample of predominantly white older adults with cardiovascular disease in Canada. Similar to our results, they found that the number of prescription medications was the best predictor of charges and physician visits and that the Charlson comorbidity index predicted mortality better than the medication-based measures.

Although reports have shown that indices based on actual chart-review measures perform better than measures calculated through administrative databases [10,20,28], the data for the current study are taken from a state-of-the-art electronic medical record used in the day-to-day care of all patients in the targeted health care system. Thus, the dataset for these analyses includes diagnoses used in day-to-day care rather than those simply used for billing purposes. Although such a database may still have limitations reflecting both clinical practice and capture of information, we do not believe these limitations would bias against one or more of the comparison comorbidity measures. Notably, the number of chronic conditions tabulated in these analyses emanates from physician diagnoses, rather than from the self-report used in many studies. This could potentially alter the reported predictive validity of this measure.

6. Conclusion

In an outpatient setting, a simple count of prescribed medications may be the most efficient comorbidity measure for predicting utilization and health care charges over the ensuing year. The count of medications also compares well to other available measures in predicting hospitalization and mortality over the coming year. Diagnosis-based measures had the greatest predictive validity for 1-year mortality. Despite concerns about face validity, the less complex indices such as a simple count of chronic conditions or medications performed at least as well as the more complex measures. Given the modest variation in outcomes explained by common comorbidity measures, however, there is obviously room for improvement in their predictive validity for total costs, utilization, or mortality. Also, all current measures tend to rely heavily on physician actions and assessments (notably diagnosis or prescription) or patients' help-seeking behavior, which itself would be expected to generate physician action. Ideally, one would like to identify a comorbidity index that is less reliant on variability in health care access, help-seeking, and financial incentives or disincentives across providers, patients, and health care systems. Candidate variables might be those that incorporate population-based measures of physiologic function or observed physical or cognitive performance. In choosing measures of comorbidity for specific analyses, researchers must remain aware of the intent of the original developers and the effect of varying the source of data, the patient population, or the outcome measure on the reported validity of the index.

References

[1] Rochon PA, Katz JN, Morrow LA, McGlinchey-Berroth R, Ahlquist MM, Sarkarati M, Minaker KL. Comorbid illness is associated with survival and length of hospital stay in patients with chronic disability: a prospective comparison of three comorbidity indices. Med Care 1996;34:1093–101.

[2] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chron Dis 1987;40:373–83.

[3] Charlson ME, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. J Clin Epidemiol 1994;47:1245–51.

[4] Linn BS, Linn MW, Gurel L. Cumulative Illness Rating Scale. J Am Geriatr Soc 1968;16:622–6.

[5] Parkerson GR Jr., Broadhead WE, Tse CKJ. The Duke Severity of Illness Checklist (DUSOI) for measurement of severity and comorbidity. J Clin Epidemiol 1993;46:379–93.

[6] Poses RM, McClish DK, Smith WR, Bekes C, Scott WE. Prediction of survival of critically ill patients by admission comorbidity. J Clin Epidemiol 1996;49:743–7.

[7] Rozzini R, Frisoni GB, Ferrucci L, Barbisoni P, Sabatini T, Ranieri P, Guralnik JM, Trabucchi M. Geriatric Index of Comorbidity: validation and comparison with other measures of comorbidity. Age Ageing 2002;31:277–85.

[8] Parkerson GR Jr., Connis RT, Broadhead WE, Patrick DL, Taylor TR, Tse CK. Disease-specific versus generic measurement of health-related quality of life in insulin-dependent diabetic patients. Med Care 1993;31:629–39.

[9] Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992;45:613–9.

[10] Van Doorn C, Bogardus ST, Williams CS, Concato J, Towle VR, Inouye SK. Risk adjustment for older hospitalized persons: a comparison of two methods of data collection for the Charlson index. J Clin Epidemiol 2001;54:694–701.

[11] D'Hoore W, Sicotte C, Tilquin C. Risk adjustment in outcome assessment: the Charlson comorbidity index. Methods Inf Med 1993;32:382–387.

[12] D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative databases. J Clin Epidemiol 1996;49:1429–33.

[13] Ghali WA, Hall RE, Rosen AK, Ash AS, Moskowitz MA. Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. J Clin Epidemiol 1996;49:273–8.

[14] Romano PS, Roos LL, Jollis JG. Further evidence concerning the use of a clinical comorbidity index with ICD-9-CM administrative data. J Clin Epidemiol 1993;46:1085–90.

[15] Starfield B, Weiner J, Mumford L, Steinwachs D. Ambulatory care groups: a categorization of diagnoses for research and management. Health Serv Res 1991;26:53–74.

[16] Weiner JP, Starfield BH, Steinwachs DM, Mumford LM. Development and application of a population-oriented measure of ambulatory care case-mix. Med Care 1991;29:452–72.

[17] Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992;45:197–203.

[18] Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. Med Care 1995;33:783–95.

[19] Beddhu S, Bruns FJ, Saul M, Seddon P, Zeidel ML. A simple comorbidity scale predicts clinical outcomes and costs in dialysis patients. Am J Med 2000;108:609–13.

[20] Kieszak SM, Flanders WD, Kosinski AS, Shipp CC, Karp H. A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. J Clin Epidemiol 1999;52: 137–142.

[21] Geissler HJ, Holzl P, Marohl S, Kuhn-Regnier F, Mehlhorn U, Sudkamp M, de Vivie ER. Risk stratification in heart surgery: comparison of six score systems. Eur J Cardiothorac Surg 2000;17:400–6.

[22] Poses RM, McClish DK, Smith WR, Huber EC, Clemo FL, Schmitt BP, Alexander D, Racht EM, Colenda CC 3rd. Results of report cards for patients with congestive heart failure depend on the method used to adjust for severity. Ann Intern Med 2000;133:10–20.

[23] Hall SF, Rochon PA, Streiner DL, Paszat LF, Groome PA, Rohland SL. Measuring comorbidity in patients with head and neck cancer. Laryngoscope 2002;112:1988–96.

[24] Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. Int J Epidemiol 2000;29:891–8.

[25] 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: 854–864.

[26] Fishman PA, Goodman MJ, Hornbrook MC, Meenan RT, Bachman DJ, O'Keefe Rosetti MC. Risk adjustment using automated ambulatory pharmacy data: the RxRisk Model. Med Care 2003;41:84–99.

[27] McDonald CJ, Tierney WM, Martin DK, Overhage JM. The Regenstrief Medical Record System: 20 years experience in hospitals, clinics, and neighborhood health centers. MD Comput 1992;9:206–17.

[28] Malenka DJ, McLerran D, Roos N, Fisher ES, Wennberg JE. Using administrative data to describe casemix: a comparison with the medical record. J Clin Epidemiol 1994;47:1027–32.