Consistency of Performance Ranking of Comorbidity Adjustment Scores in Canadian and U.S. Utilization Data

Sebastian Schneeweiss, MD, ScD, Philip S. Wang, MD, DrPH, Jerry Avorn, MD, Malcolm Maclure, ScD, Raia Levin, MSc, Robert J. Glynn, PhD, ScD

OBJECTIVE: The performance of standard comorbidity scores to control confounding is poorly defined in health care utilization data across elderly populations. We sought to evaluate and rank the performance of comorbidity scores across selected U.S. and Canadian elderly populations using health care utilization databases.

DESIGN: Cross-population validation study.

PARTICIPANTS: Study participants were residents age 65 years or older who had prescription drug coverage through state-funded programs selected from several large health care utilization databases available to the investigators: British Columbia, BC ($N = 141,161$), New Jersey, NJ ($N = 235,881$), and Pennsylvania, PA ($N = 230,913$).

MEASUREMENTS: We calculated 6 commonly used comorbidity scores for all subjects during the baseline year (1994 for NJ and PA, and 1995 for BC). These included scores based on diagnoses (Romano, Deyo, D'Hoore, Ghali) and prescription drugs (CDS-1, CDS-2). The study outcome was 1-year mortality. The performance of scores was measured by c-statistics derived from multivariate logistic regression that included age and gender.

MAIN RESULTS: Across these 4 large elderly populations, we found the same rank order of performance in predicting 1-year mortality after including age and gender in each model: Romano (c-statistic 0.754 to 0.771), Deyo (c-statistic 0.753 to 0.768), D'Hoore (c-statistic 0.745 to 0.760), Ghali (c-statistic 0.733 to 0.745), CDS-1 (c-statistic 0.689 to 0.738), CDS-2 (c-statistic 0.677 to 0.718), and age and gender alone (c-statistic 0.664 to 0.681). Performance was improved by an average of 6% by adding the number of different prescription drugs received during the past year.

CONCLUSIONS: Performance ranking of 6 frequently used comorbidity scores was consistent across selected elderly populations. We recommend that investigators use these performance data as one important factor when selecting a comorbidity score for epidemiologic analyses of health care utilization data.

KEY WORDS: comorbidity adjustment; confounding (epidemiology); prediction; claims data; methods.

J GEN INTERN MED 2004;19:444–450.

Comorbidity scores are commonly used tools to efficiently adjust for confounding in epidemiologic studies using health care utilization databases. Comorbidity scores aggregate all relevant comorbidities into a single variable. The attraction of such scores is that they are widely accepted tools that are easy to apply and can simplify data analysis, particularly when multiple hypotheses will be tested or when comorbidities are considered as time-varying confounders. There have been an increasing number of reports on the performance of comorbidity scores in adjusting confounding in observational studies.¹ These studies help researchers in selecting the best performing score for their study. However, it is not clear to what extent a score that demonstrated good performance in one population will perform well if applied in studies of other populations and databases.

The construct “comorbidity” reflects the aggregate effect of all clinical conditions a patient might have, excluding the disease of primary interest.² Because there is no gold standard, researchers have validated measures of comorbidity by how well they predict mortality or other outcomes. The predictive performance of comorbidity scores depends on several factors, including 1) the clinical conditions included in a score and their relative weights, 2) the distribution of comorbid conditions in the source population, 3) the endpoint of a study, such as 1-year mortality, and 4) the accuracy of the administrative data.³ The predictive performance of 2 scores can be validly compared when factors 2 to 4 are held constant. Several studies have explored the predictive validity of individual comorbidity measures in claims data,^{4–14} but most of these do not allow direct comparisons.¹ Even more important, there are no performance comparisons across population databases that test the generalizability of performance to similar but not identical populations.

We sought to compare the performance of 6 claims-based comorbidity scores in predicting 1-year mortality in subsets of elderly populations (65+) across selected Canadian and U.S. health care utilization databases.

METHODS

Study Populations

The data presented below are derived from 4 elderly populations in Canada and the United States that were available to the investigators:

Received from the Division of Pharmacoepidemiology and Pharmacoeconomics (SS, PSW, JA, RL, RJG), Brigham and Women's Hospital and Harvard Medical School; Department of Epidemiology (SS, MM), Harvard School of Public Health, Boston, Mass; University of Victoria (MM), Victoria, British Columbia, Canada; Division of Preventive Medicine (RJG), Brigham and Women's Hospital and Harvard Medical School; and Department of Biostatistics (RJG), Harvard School of Public Health, Boston, Mass.

Address correspondence and requests for reprints to Dr. Schneeweiss: Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont Street (Suite 3030), Boston, MA 02120 (e-mail: schneeweiss@post.harvard.edu).

1) All British Columbia residents age 65 years or older on March 31, 1996, who had at least 1 health care encounter paid for by the Ministry of Health (for a prescription medication, medical service, or hospitalization) during the 4 months prior to the baseline year (April 1, 1995 to March 31, 1996). British Columbia covers prescription medications and medical services costs for all residents 65 years or older through a tax-funded province-wide system, with a dispensing fee of \$7.60 per prescription. As part of a larger policy study,¹⁵ we obtained linkable claims data for all patients who had filled at least one prescription for an angiotensin-converting enzyme (ACE) inhibitor or calcium channel blocker during the period January 1, 1995 to December 31, 1997. Patients who died or were admitted to long-term care during the baseline year were excluded. The cohort of eligible patients ($N = 141,161$) was followed for 1 year after baseline from April 1, 1996 to March 31, 1997. We assessed comorbidity during the baseline year using several approaches and then measured mortality during the follow-up year. Results from British Columbia were presented previously¹⁶ and were used in this study for an international comparison.

2) We identically defined a cohort of Medicare enrollees age 65 years or older who had complete drug coverage either through Medicaid (32%) or the Pharmacy Assistance for the Aged and Disabled program (PAAD) (68%) in New Jersey. Similar to Medicaid, PAAD has no deductible and no maximum benefit, but there is a nominal \$2 copayment for each prescription. PAAD has more generous income eligibility criteria than those of Medicaid and therefore includes patients above the poverty level. The program's eligibility criteria were an annual income <\$15,700 if single and <\$19,250 if married. The baseline year started on January 1, 1994; the follow-up year started on January 1, 1995. All patients had a pharmacy claim during the 4 months prior to the baseline year and had received an ACE inhibitor or calcium channel blocker at least once during the observation period. A total of 108,247 patients were eligible for this study.

3) We identically defined a Medicare/PAAD population in New Jersey without the restriction to patients with ACE inhibitor or calcium channel blocker ($N = 235,881$).

4) We similarly defined a validation sample from Pennsylvania Medicare enrollees age 65 years or older who had complete drug coverage through the Pharmacy Assistance Contract for the Elderly (PACE). The eligibility criteria for PACE were an annual income <\$13,000 if single and <\$16,200 if married. There was a \$6 copayment for each prescription. The baseline year started on January 1, 1995. The follow-up year started on January 1, 1996. All patients had a pharmacy claim during the 4 months prior to the baseline year without the restriction to patients with ACE inhibitor or calcium channel blocker and survived the baseline year ($N = 230,913$).

All personal identifiers were stripped from the data prior to analysis and transformed to untraceable coded study numbers to protect participants' privacy.

Scores

Original research on the metric properties of comorbidity indices for claims data was identified by a literature search using Medline and HealthStar databases, bibliographies, and expert consultations. We identified 6 distinct indices of comorbidity for use in administrative databases.^{4,7-10,12} Four of the 6 scores use diagnostic information from ICD-9 codes and are based on the Charlson Index, which was originally designed for clinical data.¹⁷ Two scores are based on outpatient drug utilization data. The composition and development of these scores are summarized elsewhere.^{1,16} We did not include a list of 30 conditions published by Elixhauser et al.¹⁸ Such lists may perform slightly better, but they lose the advantage of simplicity of comorbidity scores that provide a single number per patient. They will therefore be of limited utility in studies that model comorbidity as a time-varying covariate and in studies of moderate or small sample size.

Diagnosis-based Scores. The Charlson Index is a list of 19 conditions; each has a weight assigned from 1 to 6. The Charlson Index score is the sum of weights for all conditions that a patient has. For the Deyo and Romano implementations of the Charlson Index, the corresponding sets of 5-digit ICD-9-CM diagnoses as delineated in their original publications were used.^{5,9} The Romano implementation is often called the Dartmouth-Manitoba score. These 2 scores differ only modestly in the ICD-9-CM codes mapping the conditions of the Charlson Index.⁵ For the D'Hoore implementation of the Charlson comorbidity index, the first 3 digits of the ICD-9 as described in D'Hoore et al. were used.¹⁰ The Ghali adaptation of the Charlson Index was calculated with the reduced set of diagnoses and new weights that improved prediction of in-hospital mortality in bypass surgery patients.¹²

The 4 scores were calculated using ICD-9-CM codes derived from all hospital discharges, which can contain up to 16 diagnoses in British Columbia and 9 diagnoses in New Jersey and Pennsylvania, and all diagnoses associated with ambulatory physician services during the baseline year. Although all 4 diagnosis-based scores were developed for hospital discharge diagnoses, studies have found that including ambulatory diagnoses in addition to hospital discharge diagnoses improves the prediction of mortality.^{16,19,20}

Prescription Medication-based Scores. The Chronic Disease Score (CDS) uses outpatient pharmacy dispensing data to assign patients to chronic disease groups. An integer weight is given to each comorbidity category represented by selected medication classes, which are summed to an overall score.⁷ For calculation of the CDS-1, we followed the original coding.⁷ Drugs that became available since 1992 were assigned to an appropriate category based on the condition for which the medication is prescribed. CDS-2

was developed later and includes more drug classes.⁸ It is based on the fairly crude 4-digit American Hospital Formulary Services classification of drugs.²¹ We used the weights that were published for predicting primary care visits.⁸

Combining several sources of morbidity information, including hospital discharge information, ambulatory or nursing home diagnoses, and pharmacy claims, can reduce residual confounding bias.^{16,19-20} In earlier studies, we found that combining a diagnosis-based comorbidity score with the number of distinct prescription medications used during the past year was found to be a particularly practical and effective improvement.^{16,19} We therefore used the number of distinct prescription drugs (distinct chemical entities) dispensed during the baseline year as an additional crude comorbidity measure.

Study Endpoint

The primary study endpoint was all-cause mortality during the follow-up year. We used British Columbia vital statistics information to identify deaths in Canada, and both Medicare and Medicaid eligibility files in the U.S. populations.²² We used complete 1-year follow-up information on all patients.

Data Quality

The accuracy of pharmacy claims data within Medicaid, PAAD, and PACE is considered very good.²³ The strength and limitations of diagnostic coding in Medicaid and Medicare have been frequently examined.²⁴⁻²⁶ In British Columbia, pharmacy dispensing data including medication, strength, and the number of units are entered into a computer network by the pharmacist when a prescription is filled, and underreporting and misclassification appear to be minimal.²⁷ Although previous reports indicate reasonable levels of accuracy and completeness of diagnostic coding,²⁸ misclassification of ICD-9-CM diagnoses in British Columbia, New Jersey, and Pennsylvania is probably similar to that seen in research using other administrative databases.^{25,29-32}

Data Analysis

One-year mortality was modeled by multivariate logistic regression models that included age, gender, and each of the comorbidity scores. The scores and number of distinct prescription drugs used in the previous year were modeled as continuous variables; earlier studies showed no improvement in predictive performance compared to grouping scores in 3 categories.¹⁶ Second-degree polynomials were considered for age and the scores, and were removed from the final models if they did not significantly improve the model fit (likelihood ratio test $P < .01$). The area under the receiver operating characteristic (ROC) curve or c-statistic was calculated, as it is a widely accepted measure of discrimination.³³ The c-statistic ranges from 0

to 1, with 1 indicating a perfect prediction and 0.5 indicating a chance prediction. For example, the Framingham Heart Study equations predict the incidence of coronary heart disease based on age, blood pressure, smoking, diabetes, and LDL and HDL levels with a c-statistic of 0.77.³⁴ c-statistics between 0.7 and 0.8 are generally considered as acceptable and between 0.8 and 0.9 as excellent.³⁵ Higher values are rarely observed in studies of the general population and are described as outstanding. It is assumed that in most cases, a higher c-statistic is a marker for better control for confounding.¹⁶ Asymptotic 95% confidence limits were calculated for c-statistics.³⁶

We calculated the relative improvement in percent between c-statistics of 2 scores beyond the predictive power of age and gender alone as $\{[(c_1 - c_b) - (c_2 - c_b)] / (c_1 - c_b)\} * 100$, with c_b being the c-statistic of the corresponding baseline age-gender adjusted logistic regression model(s).³⁶ Because the comparison of absolute performance depends on many factors that vary among populations and studies, we chose to compare the relative ranking of scores across the 4 study populations.

RESULTS

Population

At the beginning of the baseline year, the BC population was on average 75.4 years old (standard deviation ± 6.7), and 58% were female. The U.S. populations were comparable in age but had higher proportions of female enrollees (78%, 77%, and 80.7%; Table 1). The distributions of comorbidity indices during the baseline period are shown in Table 2. There were 17,690 deaths in the total NJ population (7.5%), and 20,684 deaths in PA (7.7%). In the cardiovascular subpopulations, 5,569 deaths occurred during the follow-up year in BC (3.9%), and 7,236 in NJ (6.7%).

Performance

The Romano adaptation of the Charlson Index performed best among the 6 scores at predicting 1-year mortality consistently across all 4 populations, with c-statistics ranging from $c = 0.754$ in the cardiovascular disease population of NJ to 0.771 in the total BC population (Table 3). The performance of Deyo's version was not statistically significantly different from the Romano implementation. The 3-digit ICD-based D'Hoore score and Ghali's adaptation consistently performed less well, and both CDS-1 and CDS-2 performed even less well (Table 3). In all 4 study populations the same rank order of predictive performance (c-statistic) was found: Romano $\geq$ Deyo $>$ D'Hoore $>$ Ghali $>$ CDS-1 $>$ CDS-2. On average, scores predicted mortality better in the restricted BC cohort than in the identically defined restricted Medicare cohort, mostly due to better prediction by the baseline age-gender model. However, the c-statistics of the 2 drug-based scores (CDS-1 and CDS-2) were 56% and 65% higher in the BC population compared to a similarly defined NJ population adjusted for the

Table 1. Characteristics of Four Elderly Populations in British Columbia, New Jersey, and Pennsylvania During the Baseline Year

	British Columbia (Cardio-vasc.*)	New Jersey/ PAAD (Cardio-vasc.*)	New Jersey/ PAAD (Total)	Pennsylvania/ PACE (Total)
N	141,161	108,247	235,881	230,913
Age, y ($\pm$ SD)	75.4 ($\pm$ 6.7)	76.7 ($\pm$ 7.2)	77.6 ($\pm$ 7.5)	77.6 ($\pm$ 7.0)
Female, %	58.0	78.0	77.2	80.7
Average number of all hospitalizations	0.36 ($\pm$ 0.9)	0.49 ($\pm$ 1.0)	0.39 ($\pm$ 0.9)	0.48 ($\pm$ 1.0)
Average number of distinct prescription drugs [†] ($\pm$ SD)	7.4 ($\pm$ 5.1)	9.4 ($\pm$ 6.0)	7.7 ($\pm$ 5.9)	6.5 ($\pm$ 4.6)
Average number of distinct ICD diagnoses [‡] ($\pm$ SD)	3.5 ($\pm$ 2.7)	9.3 ($\pm$ 9.0)	8.1 ($\pm$ 8.8)	14.6 ($\pm$ 10.8) [§]
Average number of physician visits/services ($\pm$ SD)	8.8 ($\pm$ 8.6)	9.9 ($\pm$ 8.4)	8.5 ($\pm$ 8.4)	14.0 ($\pm$ 12.0) [§]

* Restricted to patients receiving ACE inhibitors or calcium channel blockers during the observation period.

[†] Prescription medications that have different chemical structures but may be of the same therapeutic group.

[‡] ICD-9 diagnoses at hospital discharge and ambulatory care visits that differ in their first 3 digits.

[§] The Pennsylvania claims databases allowed for more diagnostic and service coding for ambulatory care.

predictive power of age and gender alone. Combining the number of different prescription drugs used in the past year with the Romano score improved predictive power consistently by an average of 6% in all 4 study populations. Modeling scores as tertiles or including quadratic terms did not improve prediction.

DISCUSSION

Among 4 elderly North American populations we found a consistent performance ranking of comorbidity scores in predicting 1-year mortality using claims data. This finding is reassuring for researchers who want to select the best

Table 2. Distributions of Six Comorbidity Scores During the Baseline Year in Elderly in British Columbia, New Jersey, and Pennsylvania

	Mean	Standard Deviation	Percentage with Score = 0	Median	75th Percentile	Max
British Columbia (Cardio-vasc.*)						
CDS-1	4.4	2.3	0.0	4.0	6	19
CDS-2	3.1	1.3	0.0	2.8	3.6	12
D'Hoore	1.2	1.8	56.1	0.0	2	18
Ghali	0.4	1.2	86.7	0.0	0	11
Deyo	0.5	1.1	73.8	0.0	1	12
Romano	0.5	1.1	73.3	0.0	1	14
New Jersey/PAAD (Cardio-vasc.*)						
CDS-1	5.2	2.1	0.0	5	6	20
CDS-2	3.5	1.5	0.0	3	4	11
D'Hoore	1.6	1.9	41.8	1	2	17
Ghali	1.1	2.0	67.7	0	1	11
Deyo	1.4	1.9	45.4	1	2	15
Romano	1.5	1.9	43.5	1	2	16
New Jersey/PAAD (Total)						
CDS-1	4.1	3.0	16.1	4	6	22
CDS-2	2.9	1.7	11.5	3	4	11
D'Hoore	1.3	1.8	50.0	1	2	17
Ghali	0.9	1.8	74.0	0	1	11
Deyo	1.2	1.8	52.2	0	2	16
Romano	1.3	1.8	50.3	0	2	16
Pennsylvania/PACE (Total)						
CDS-1	3.8	2.5	12.4	4	5	19
CDS-2	2.9	1.5	7.1	3	4	11
D'Hoore	1.8	2.0	32.8	1	3	17
Ghali	1.3	2.1	61.9	0	2	11
Deyo	1.6	1.9	34.1	1	2	19
Romano	1.8	2.0	32.0	1	3	20

* Restricted to patients receiving ACE inhibitors or calcium channel blockers during the observation period.

Table 3. Prediction of One-Year Mortality by Six Comorbidity Scores Measured One Year Earlier in the Four Study Populations

Score*	Model	British Columbia (Cardio-vasc. [†])			New Jersey/PAAD (Cardio-vasc. [†])			New Jersey/PAAD (Total)			Pennsylvania/PACE (Total)		
		c	95% Confidence Limits		c	95% Confidence Limits		c	95% Confidence Limits		c	95% Confidence Limits	
CDS-1	Age + gender	0.681	0.674	0.695	0.664	0.658	0.671	0.658	0.654	0.662	0.669	0.665	0.673
	Age + gender + CDS-1	0.738	0.731	0.744	0.689	0.682	0.695	0.703	0.699	0.707	0.695	0.691	0.699
CDS-2	Age + gender + CDS-2	0.718	0.711	0.725	0.677	0.670	0.683	0.690	0.686	0.694	0.683	0.679	0.687
	Age + gender + D'Hoore	0.745	0.739	0.752	0.745	0.739	0.751	0.760	0.757	0.764	0.747	0.743	0.750
Ghali	Age + gender + Ghali	0.745	0.738	0.752	0.738	0.732	0.744	0.739	0.735	0.743	0.733	0.729	0.737
	Age + gender + Deyo	0.768	0.762	0.775	0.753	0.748	0.759	0.768	0.764	0.772	0.757	0.753	0.761
Romano	Age + gender + Romano	0.771	0.764	0.777	0.754	0.748	0.760	0.771	0.767	0.775	0.757	0.754	0.761
	Age + gender + number of drugs used in past year	0.745	0.738	0.752	0.701	0.695	0.707	0.713	0.709	0.717	0.701	0.697	0.705
Romano + # of Rx	Age + gender + Romano + number of drugs used in past year	0.783	0.778	0.789	0.756	0.750	0.761	0.775	0.771	0.779	0.760	0.756	0.764

* Scores are modeled as continuous variables.

† Restricted to patients receiving ACE inhibitors or calcium channel blockers during the observation period.

‡ Prescription medications that have different chemical structures but may be of the same therapeutic group.

performing standard score to control for comorbidity. Based on these data it can be assumed with greater certainty that the performance ranking can be generalized to other large general populations age 65 years or older, or large subgroups thereof, such as cardiovascular patients.

Comorbidity scores are useful because they are easy to apply and save time and resources, a major issue when analyzing massive health care databases and testing multiple hypotheses. They can also increase the efficiency of statistical inference, which may become an issue in claims data when analyzing small population subgroups or when comorbidities are modeled as time-varying covariates in longitudinal studies. However, adjusting for a score should not be regarded as successfully controlling for all confounding caused by comorbidity,³⁷ because even scores with improved performance impose a functional relation between comorbidities and outcome. The narrower a subpopulation is defined, the more likely a score developed for a more general population will perform insufficiently. Scores are useful for preliminary analyses to indicate the direction and magnitude of confounding, which can guide decisions about further adjustment. It remains unclear how much more confounding can be controlled by using traditional multivariate modeling techniques to control comorbidity.

We found that diagnosis-based scores consistently performed better than medication-based scores at predicting future mortality. This is consistent with earlier findings that sicker patients are less likely to be treated for comorbid conditions.³⁸ In particular, medications with some preventive effects, such as oral antidiabetic agents³⁹ or lipid-lowering drugs,⁴⁰ are less frequently prescribed in very sick patients, causing them to appear artificially “healthier” in medication-based scores. Another limitation of medication-based scores is that their list of drugs must be updated regularly as new drugs and drug classes become available.

In earlier studies, we had found that the number of distinct medications received during the previous year was a better predictor of mortality than the CDS-1.¹⁶ This finding is consistently true among our 4 large study populations. It is not surprising that the Ghali score in our sample of elderly beneficiaries did not perform as well as in the original publication because its weights were developed in patients undergoing bypass surgery.^{1,12}

Numeric differences in the performance between some scores were small because the c-statistic (or area under the ROC curve) has a limited sensitivity to detect additional improvements in prediction once a certain level is reached.⁴¹ Nevertheless, if researchers had the option to select one score, we would recommend choosing the numerically best performing even if the absolute difference in c is small. In practice, often not all desirable data sources (5-digit ICD-9-CM diagnostic data plus pharmacy dispensing data) are available. Based on our ranking, researchers can pick the next best alternative and discuss the increased potential for residual confounding compared to the better performing scores.

A c of 0.77 means that for all possible pairs of individuals who died with individuals who lived, 77% of the time the model correctly attributed a higher risk of death to the person who died than to the person who lived. It has been shown how predictive validity can be translated into confounder adjustment.¹⁶ From several examples⁴²⁻⁴⁴ and text books,³⁵ it appears that a c above 0.7 is a model with acceptable performance and above 0.8 with excellent predictive performance; however, the meaningfulness of any level of performance depends on what is otherwise achievable using other methods. Large investments, such as additional data and analyses, yield only small numeric gains in c above 0.75. Whether those gains are worth their price depends on the benefits of a "truer" analysis and the costs of error, issues that are unique to each problem studied.

The data for our study are now several years old; however, we are not aware of important changes in diagnostic coding for Medicare patients or changes to the structure and interpretation of relevant data fields. In the unlikely event that there were significant changes, they would affect the absolute performance in predicting mortality (c -statistic) but should not change the ranking in any systematic way.

Our study was limited to predicting mortality, the ultimate outcome of care. However, the ranking of scores may be different when predicting other health care outcomes, such as annual expenditures or number of services. The study was further limited to health care utilization databases from 3 states/provinces that were available to the authors and originally requested for other purposes. Although this selection may not be representative for both countries, it provides sufficient variability in characteristics of patients, health care systems, and recording practice for claims. The Pennsylvania database had more extensive coding of diagnoses related to ambulatory visits and services and the British Columbia data had more discharge diagnoses available. This may have affected the absolute level of performance, which is the reason why a relative ranking is the most valid approach to compare the performance of scores across populations.

We recommend that investigators use these performance data as one important factor when selecting a comorbidity score for epidemiologic analyses of health care utilization data. The Charlson-based Romano and Deyo scores using published weights in combination with a simple count of the different prescription drugs received during the past year appear to be well performing comorbidity measures in epidemiologic studies.

Supported by grants R01-HS10881 from the U.S. Agency for Healthcare Research and Quality, and R03-AG19463, R03-AG18395, and R01-AG18833 from the National Institute on Aging.

REFERENCES

- Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol.* 2000;29:891-8.
- Greenfield S, Nelson EC. Recent developments and future issues in the use of health status assessment measures in clinical setting. *Med Care.* 1992;30:MS23-MS41.
- Iezzoni LI. *Risk Adjustment for Measuring Healthcare Outcomes.* 2nd ed. Chicago, Ill: Health Administration Press; 1997.
- Roos LL, Sharp SM, Cohen MM, Wajda A. Risk adjustment in claims-based research: the search for efficient approaches. *J Clin Epidemiol.* 1989;42:1193-1206.
- 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.
- 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.
- von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. *J Clin Epidemiol.* 1992;45:197-203.
- 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.
- 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.
- D'Hoore W, Sicotte C, Tilquin C. Risk adjustment in outcome assessment: the Charlson Comorbidity Index. *Methods Inf Med.* 1993;32:382-7.
- D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson index with administrative data bases. *J Clin Epidemiol.* 1996;49:1429-33.
- 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.
- Melfi C, Holleman E, Arthur D, Katz B. Selecting a patient characteristics index for the prediction of medical outcomes using administrative claims data. *J Clin Epidemiol.* 1995;48:917-26.
- Poses RM, Smith WR, McClish DK, Anthony M. Controlling for confounding by indication for treatment. Are administrative data equivalent to clinical data? *Med Care.* 1995;33:AS36-AS46.
- Schneeweiss S, Walker AM, Glynn RJ, Maclure M, Dormuth C, Soumerai SB. Outcomes of reference drug pricing for angiotensin-converting enzyme inhibitors. *N Engl J Med.* 2002;346:822-9.
- Schneeweiss S, Seeger J, Maclure M, Wang P, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol.* 2001;154:854-64.
- 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.
- Elixhauser A, Steiner C, Harris R, Coffey RM. Comorbidity measures for use with administrative data. *Med Care.* 1998;36:8-27.
- Schneeweiss S, Wang PS, Avorn J, Glynn RJ. Improved comorbidity adjustment for predicting mortality in Medicare populations. *Health Serv Res.* 2003;38:1103-20.
- 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.
- American Society of Health-System Pharmacists: AHFS Drug Information 96. Bethesda, Md; 1996.
- Wang PS, Walker A, Tsuang M, Orav EJ, Levin R, Avorn J. Strategies for improving comorbidity measures based on Medicare and Medicaid claims data. *J Clin Epidemiol.* 2000;53:571-8.
- Yuan Z, Cooper GS, Einstadter D, Cebul RD, Rimm AA. The association between hospital type and mortality and length of stay. *Med Care.* 2000;38:231-45.
- Lessler JT, Harris BSH. Medicaid data as a source for postmarketing surveillance information, final report. Research Triangle Park, NC: Research Triangle Institute; 1984.
- Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T.

- Comorbidities, complications, and coding bias: does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA*. 1992;267:2197-203.
26. Fisher ES, Whaley FS, Krushat M, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *Am J Public Health*. 1992;82:243-8.
27. Anderson GM, Kerluke KJ, Pulcins IR, Hertzman C, Barer ML. Trends and determinants of prescription drug expenditures in the elderly: data from the British Columbia Pharmacare Program. *Inquiry*. 1993;30:199-207.
28. Roos LL, Sharp SM, Cohen MM. Comparing clinical information with claims data: some similarities and differences. *J Clin Epidemiol*. 1991;44:881-8.
29. Williams JI, Young W. Inventory of studies on the accuracy of Canadian health administrative databases. Technical Report, Institute for Clinical Evaluative Sciences (ICES); December 1996.
30. Fowles JB, Lawthers AG, Weiner JP, et al. Agreement between physician's office records and Medicare part B claims data. *Health Care Financ Rev*. 1995;16:189-99.
31. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. *Med Care*. 1994;32:81-90.
32. Glynn RJ, Monane M, Gurwitz JH, Choodnovskiy I, Avorn J. Agreement between drug treatment and a discharge diagnosis of diabetes. *Am J Epidemiol*. 1999;149:541-9.
33. Ash AS, Shwartz M. Evaluating the performance of risk-adjustment methods: dichotomous outcomes. In: Iezzoni LI, ed. *Risk Adjustment for Measuring Healthcare Outcomes*. 2nd ed. Chicago, Ill: Health Administration Press; 1997.
34. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837-47.
35. Hosmer DW, Lemeshow S. *Applied Logistic Regression*, 2nd ed. New York, NY: Wiley; 2000.
36. Liebetrau AM. *Measures of Association. Quantitative Application in the Social Sciences*. Vol. 32. Beverly Hills, Calif: Sage Publications; 1983.
37. Maclure M, Schneeweiss S. Causation of bias: the episcopo. *Epidemiology*. 2001;12:114-22.
38. Redelmeier DA, Tan SH, Booth GL. The treatment of unrelated disorders in patients with chronic medical diseases. *N Engl J Med*. 1998;338:1516-20.
39. Glynn RJ, Monane M, Gurwitz JH, Choodnovskiy I, Avorn J. Aging, comorbidity, and reduced rates of drug treatment for diabetes mellitus. *J Clin Epidemiol*. 1999;52:781-90.
40. Glynn RJ, Knight EL, Levon R, Avorn J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology*. 2001;12:682-9.
41. Moons KGM, Stijnen T, Michel BC, et al. Application of treatment thresholds to diagnostic test evaluation: an alternative to the comparison of areas under receiver operating characteristics curves. *Med Decis Making*. 1997;17:447-54.
42. Hannan EL, Kilburn H, Lindsey ML, Lewis R. Clinical versus administrative data bases for CABG surgery. Does it matter? *Med Care*. 1992;30:892-907.
43. Grover S, Coupal L, Hu X-P. Identifying adults at increased risk of coronary disease: how well do the current cholesterol guidelines work? *JAMA*. 1995;274:801-6.
44. Wilson PWF, D'Agostino RB Sr., Levy L, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837-47.