# Improved Comorbidity Adjustment for Predicting Mortality in Medicare Populations

*Sebastian Schneeweiss, Philip S. Wang, Jerry Avorn, and Robert J. Glynn*

**Objective.** To define and improve the performance of existing comorbidity scores in predicting mortality in Medicare enrollees.

**Data Sources.** Study participants were two Medicare populations who had complete drug coverage either through Medicaid or a statewide pharmacy assistance program: New Jersey Medicare enrollees ($N_{NJ}$ = 235,881) and Pennsylvania Medicare enrollees ($N_{PA}$ = 230,913).

**Study Design.** Frequently used comorbidity scores were computed for all subjects during the baseline year ( January 1, 1994, to December 31, 1994, and one year later in Pennsylvania). The study outcome was one-year mortality during the following year. Performance of scores was measured with the c-statistic derived from multivariate logistic regression models. Empirical weights were derived in the New Jersey population and the performance of scores with new weights was validated in the Pennsylvania population.

**Principal Findings.** A score based on ICD-9-diagnoses (Romano) performed 60 percent better than one based on patterns of medication use (Chronic Disease Score, or CDS-1) ($c$ = 0.771 vs. $c$ = 0.703). The performance of the Romano score was further improved slightly by inclusion of the number of different prescription drugs used during the past year. Modeling the 17 conditions included in the Romano score as separate binary indicators increased its performance by 8 percent ($c$ = 0.781). We derived elderly-specific weights for these scores in the New Jersey sample, including negative weights for the use of some drugs, for example, lipid lowering drugs. Applying these weights, the performance of Romano and CDS-1scores improved in an independent validation sample of Pennsylvania Medicare enrollees by 8.3 percent and 43 percent compared to the scores with the original weights. When we added an indicator of nursing home residency, age, and gender, the Romano score reached a performance of $c$ = 0.80.

**Conclusions.** We conclude that in epidemiologic studies of the elderly, a modified diagnosis-based score using empirically derived weights provides improved adjustment for comorbidity and enhances the validity of findings.

**Key Words.** Comorbidity adjustment, confounding (epidemiology), prediction, claims data, health services epidemiology, methods, elderly

As with any observational research, epidemiologic studies using data from administrative datasets have been criticized for incomplete control for confounding. Poorly measured patient characteristics, including comorbidities, can affect clinical decisions and may therefore bias results from etiologic studies using observational data (Walker 1996). However, it is widely acknowledged that in many situations, claims data may be the best source to study the outcomes of therapeutic and other clinical strategies in routine medical care of the elderly (Wang, Solomon et al. 2000; Ray et al. 2002). Improving comorbidity adjustment is therefore an important goal to increase the validity of findings from epidemiologic studies of older patients (Schneeweiss and Maclure 2000).

Comorbidity scores aggregate relevant comorbidities into a single variable. The attraction of such scores is that they are easy to apply and widely accepted tools. They can simplify data analysis, particularly when multiple hypotheses will be tested or when comorbidities are considered as time-varying confounders. In small studies or in subgroup analyses with limited sample size, the fact that comorbidity scores reduce the number of variables that need to be adjusted will improve statistical inference. However, in either of these circumstances scores must correctly model the underlying functional relation between comorbidity and outcome in a specific study population.

In an ideal world, one would have enough data and well-measured variables to completely adjust statistically for differences in risk. Such adjustments would be tailored to the outcome measure (e.g., death, readmission) and the population under study. In the real world, however, one may not have a large enough dataset, or all the data one would like. A second best alternative, then, is to use a comorbidity score that has

---

Supported by research grant no. R01-HS10881 from the U.S. Agency for Healthcare Research and Quality and no. R03-AG18395 from the National Institute on Aging to Dr. Schneeweiss, and no. RO1-AG18833 from the National Institute on Aging to Dr. Glynn.

Address correspondence to Sebastian Schneeweiss, M.D., Sc.D., Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 221 Longwood Ave (BLI-341), Boston, MA 02115. Philip S. Wang is also with the Division of Pharmacoepidemiology and Pharmacoeconomics at Brigham and Women's Hospital as well as the Department of Epidemiology, Harvard School of Public Health, Boston. Jerry Avorn and Robert J. Glynn are with the Division of Pharmacoepidemiology and Pharmacoeconomics at Brigham and Women's. Additionally, Dr. Glynn is with the Department of Preventive Medicine at Brigham and Women's.

preassigned weights from a plausibly relevant population. Given the widespread use of scores such as the Charlson Index and Chronic Disease Score, this paper puts forward an alternative that is based on mortality in the Medicare population.

The construct "comorbidity" reflects the aggregate effect of all clinical conditions a patient might have, excluding the disease of primary interest (Greenfield and Nelson 1992; Last 1995). Because there is no gold standard, researchers have validated measures of comorbidity by how well they predict mortality. 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, for example, one-year mortality, and (4) the accuracy and completeness of the administrative data (Iezzoni 1997). The predictive performance of two scores can be validly compared when factors 2 to 4 are held constant. Several studies have explored the predictive validity of comorbidity measures in claims data (Roos et al. 1989; Romano, Roos, and Jollis 1993a, 1993b; Von Korff, Wagner, and Saunders 1992; Clark et al. 1995; Deyo, Cherkin, and Ciol 1992; D'Hoore, Sicotte, and Tilquin 1993; D'Hoore, Bouckaert, and Tilquin 1996; Ghali et al. 1996; Melfi et al. 1995; Poses et al. 1995). However, only a few studies sought to improve the performance of these scores (Clark et al. 1995; Zhang, Iwashyna, and Christakis 1999; Wang, Walker et al. 2000; Schneeweiss et al. 2001) and often only for specific subpopulations, for example, patients undergoing bypass surgery (Ghali et al. 1996).

From earlier studies it has become clear that specific weights must be applied to elderly populations to achieve acceptable comorbidity adjustment in epidemiologic research using claims data (Schneeweiss and Maclure 2000; Ghali et al. 1996). We sought to improve and validate the improved performance of an existing and frequently used ICD-9 diagnosis-based comorbidity score and a drug use-based score in predicting one-year mortality in U.S. Medicare patients aged 65 years or older.

## METHODS

### *Study Populations*

This study used a development sample from New Jersey and a validation sample from Pennsylvania. We defined a cohort of New Jersey Medicare enrollees aged 65 years or older who had complete drug coverage either through Medicaid (32 percent) or the Pharmacy Assistance for the Aged and

Disabled program (PAAD) in New Jersey (68 percent). Similar to Medicaid, PAAD has no deductible and no maximum benefit, but there is a nominal $2 copayment for each prescription. The PAAD program has more generous income eligibility criteria than those of Medicaid (less than $7,400 if single and less than $9,850 if married) and therefore includes patients above the poverty level. The eligibility criteria for PAAD were an annual income less than $16,200 if single and less than $19,850 if married. The combined Medicaid-PAAD population is therefore more representative for Medicare beneficiaries. The baseline year started on January 1, 1994, and the follow-up year started on January 1, 1995. All patients had a pharmacy claim during the four months prior to the baseline year and survived the baseline year. Patients eligible for the development sample were $N_{NJ} = 235{,}881$.

We similarly defined a validation sample from Pennsylvania Medicare enrollees aged 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 less than $13,000 if single and $16,200 if married. The baseline year started on January 1, 1995, and the follow-up year started on January 1, 1996. All patients had a pharmacy claim during the four months prior to the baseline year and survived the baseline year. The validation study comprised $N_{PA} = 230{,}913$ subjects.

### *Scores*

Since there is a large number of scores and their combinations that could theoretically be evaluated, we restricted this study to the best performing scores in predicting mortality using ICD-9 diagnosis and pharmacy records that were identified in an earlier, Canadian claims data study (Schneeweiss et al. 2001).

#### *Romano's Adaptation (Score) of the Charlson Index for Use with Claims Data*

The Charlson Index is a weighted sum of presence or absence of each of 19 conditions; each condition is assigned a weight from 1 to 6, with higher weights indicating greater severity (Charlson et al. 1987). A person's Charlson Index is the sum of these weights. The Romano implementation of the Charlson Index has been shown to perform best in adjusting for comorbidity in claims data, closely followed by the Deyo score (Deyo, Cherkin, and Ciol 1992; Schneeweiss et al. 2001). Each condition was identified by the corresponding sets of five-digit ICD-9-CM diagnoses as delineated in their original publications (Romano, Roos, and Jollis 1993a, 1993b). The Romano score was calculated using ICD-9-CM codes derived from all hospital

discharges, which can contain up to nine diagnoses in New Jersey and Pennsylvania, and all diagnoses associated with ambulatory physician services during the baseline year. Earlier work demonstrated that including ambulatory diagnoses in addition to hospital discharge diagnoses can moderately improve the prediction of mortality from $c = 0.757$ to 0.771 (16 percent) (Schneeweiss et al. 2001).

We further identified 30 conditions by using ICD-9-CM codes described by Elixhauser et al. (1998). The Elixhauser system does not combine these conditions into a weighted score; instead the conditions are included as binary indicators in a multivariate regression model to adjust for confounding.

### *Chronic Disease Score*

The Chronic Disease Score is a comorbidity measure obtained from a weighted sum of the use of medications from 30 different classes (Von Korff, Wagner, and Saunders 1992). An integer weight between 1 and 5 is given to each of the selected medication classes, which are then summed to an overall score. We adapted the original coding to calculate the Chronic Disease Score (CDS-1). Drugs that have become available since 1992 were assigned to an appropriate category based on the condition for which the medication is prescribed. For example, only cimetidine was originally specified as an indicator for ulcer disease, and we expanded this list to include any H2 antagonist or proton pump inhibitor. For drugs that were available at the time the score was developed, but have since had their indications expanded to include one of the scored chronic diseases, the disease categories were not changed with regard to that drug, for example, methotrexate for cancer but now more frequently used for rheumatoid arthritis. We calculated the CDS-1 for each patient based on all prescriptions filled during the baseline year.

We further used the number of distinct prescription drugs (distinct chemical entities) dispensed during the baseline year as a crude comorbidity measure because it moderately increased the predictive performance in earlier research from 0.771 to 0.783 (12 percent) (Schneeweiss et al. 2001). Medications that were equal in the first eight digits of the American Hospital Formulary Services code (1996) were considered to be the same substance.

### *Study Endpoint*

The primary study endpoint was all-cause mortality during the follow-up year. We used both Medicare and Medicaid eligibility files to identify deaths (Yuan et al. 2000) and had complete, one-year follow-up information on all patients.

*Data Quality*

The accuracy of pharmacy claims data within Medicaid, PAAD, and PACE has been found to be very good (Lessler and Harris 1984). The strength and limitations of diagnostic coding in Medicaid and Medicare are frequently examined and discussed (Iezzoni et al. 1992; Fisher et al. 1992; Roos, Sharp, and Cohen 1991; Avorn 1991; Bright, Avorn, and Everitt 1989). Misclassification of ICD-9-CM diagnoses in New Jersey and Pennsylvania are likely to be similar to that observed by other studies using Medicare claims data (Fisher et al. 1992; Fowles et al. 1995; Romano and Mark 1994; Glynn et al. 1999a).

*Data Analysis*

One-year mortality was modeled by multivariate logistic regression models that included age and gender, and each of the two scores separately. Comorbidity scores were modeled as continuous variables because earlier studies found no improvement in predictive performance when scores were grouped into three categories (Schneeweiss et al. 2001). Second-degree polynomials were considered and removed from the final models if they did not significantly improve the model fit (likelihood ratio test $p < 0.01$). C-statistics ( = area under the receiver-operating-characteristic [ROC] curve) were calculated as measures of discrimination (Ash and Shwartz 1997). The c-statistic ranges from 0 to 1, with 1 indicating perfect prediction and 0.5 indicating chance prediction. For reference, the Framingham Heart Study predicts 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 (Wilson et al. 1998). It has been suggested that c-statistics between 0.7 and 0.8 can be considered as acceptable, and between 0.8 and 0.9 as excellent (Hosmer and Lemeshow 2000). Higher values are rarely observed in population-based research, and are considered outstanding. The correlation between higher c-statistics and better control for confounding has been demonstrated (Schneeweiss et al. 2001). Asymptotic 95 percent confidence limits have been reported for c-statistics (Liebetrau 1983).

In addition to the absolute improvement in predictive power, we calculated the difference between two c-statistics, $c_1$ and $c_2$, in percent 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). A difference in $c$ between two comorbidity scores of 0.703 and 0.771 when the baseline $c$ is 0.658 corresponds to a 60 percent increase in

$c$: $[(0.771 - 0.658) - (0.703 - 0.658)]/(0.771 - 0.658) = 0.60$. Percent differences are presented in addition to the absolute c-values.

### *Derivation of Elderly Specific Weights*

We decomposed the diagnosis-based (Romano) and medication-based (CDS-1) scores into their individual component conditions, and modeled one-year mortality as a function of a linear vector of equally weighted binary indicators of all conditions in the total Medicare population. Based on the independent one-year mortality risk of each individual condition we assigned each condition a new weight. We increased the weights by 1 point with each 0.3 increase in the ln (OR). A weight of 1 therefore refers to an exp (0.45-0.15) = 35 percent increase in risk of dying during the follow-up year (see appendix).

## RESULTS

### *Population*

At the beginning of the baseline year, the New Jersey population was on average 78 years old (standard deviation: $\pm$ 7.5) and 77 percent were female (Table 1). The Pennsylvania population was comparable in age and gender, but had more physician visits and diagnoses recorded per subject (Table 1). The distributions of comorbidity indices during the baseline period are shown in Table 2. A total of 17,690 deaths occurred during the follow-up year in New Jersey (7.5 percent) and 20,684 deaths (7.7 percent) in Pennsylvania.

### *Improving Performance*

The Romano adaptation of the Charlson index (c = 0.771) performed 60 percent better than the CDS-1 score ($c = 0.703$) in the New Jersey Medicare/ PAAD population (Table 3). Including quadratic terms for age or the scores did not improve the predictive power (<2 percent).

Modeling the 17 conditions that are included in the Romano score as separate binary indicators ("$\text{Romano}_{17}$") increased the performance by 8.1 percent in Medicare claims data (Table 3). $\text{Romano}_{17}$ and the 30 diagnosis indicators suggested by Elixhauser performed equally well ($c = 0.781$). A marginal improvement could be achieved by adding indicators for acute myocardial infarction and stroke to the Elixhauser set of conditions (+1.6 percent).

Table 1: Characteristics of Two Medicare Populations in New Jersey and Pennsylvania during the Baseline Year

| | *New Jersey/ PAAD** | *Pennsylvania/ PACE*** |
|---|---|---|
| *N* | 235,881 | 230,913 |
| Age | 77.6 (±7.5) | 77.6 (±7.0) |
| Female | 77.2 % | 80.7 % |
| Average number of all hospitalizations | 0.39 (±0.9) | 0.48 (±1.0) |
| Proportion of patients with any hospitalization | 23.1% | 26.7% |
| Average number of distinct prescription drugs † | 7.7 (±5.9) | 6.5 (±4.6) |
| Proportion of patients with any prescription drug | 92.1% | 96.1% |
| Average number of distinct ICD diagnoses †† | 8.1 (±8.8) | 14.6 (±10.8) |
| Proportion of patients with any ICD diagnosis | 83.3% | 96.9% |
| Average number of physician visits | 8.5 (±8.4) | 14.0 (±12.0) |
| Proportion of patients with any physician visit | 82.5% | 96.4% |
| Nursing home residents | 13.5 % | 10.5 % |
| Proportion of patients dying in follow-up year | 7.5% | 7.7% |

*PAAD = Pharmacy Assistance for the Aged and Disabled program.

**PACE = Pennsylvania Assistance Contract for the Elderly.

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

††ICD-9 diagnoses that differ in their first three digits from hospital as well as ambulatory care diagnoses during the baseline year.

Table 2: Distributions of Two Comorbidity Scores during the Baseline Year in Seniors in New Jersey/PAAD and Pennsylvania/ PACE Medicare Populations

| | *Mean* | *Standard Deviation* | *Percentage with Score = 0* | *Median* | *75th Percentile* | *Max* |
|---|---|---|---|---|---|---|
| New Jersey/PAAD: | | | | | | |
| CDS-1* | 4.1 | 3.0 | 16.1 | 4 | 6 | 22 |
| Romano** | 1.3 | 1.8 | 50.3 | 0 | 2 | 16 |
| Pennsylvania/ PACE: | | | | | | |
| CDS-1 | 3.8 | 2.5 | 12.4 | 4 | 5 | 19 |
| Romano | 1.8 | 2.0 | 32.0 | 1 | 3 | 20 |

*Chronic Disease Score (Von Korff, Wagner, and Saunders 1992).

**Romano's adaptation of the Charlson Index for use with claims data (Romano, Roos, and Jollis 1993a, 1993b).

The combination of the ICD-based Romano score with the number of distinct prescription medications used during the baseline year improved the performance by on average 5.8 percent compared to the Romano score alone in the New Jersey Medicare population.

Table 3: Comorbidity Scores and Their Performance in a New Jersey Medicare/ PAAD Population in Predicting One-Year Mortality

| | | | *New Jersey /PAAD* | |
|---|---|---|---|---|
| *Score* | *Model* | *Degrees of Freedom* | *c* | *95% Confidence Interval* |
| | Age + gender | 2 | 0.658 | (0.654–0.662) |
| CDS-1 | Age + gender + CDS-1 | 3 | 0.703 | (0.699–0.707) |
| Romano (1 score) | Age + gender + Romano | 3 | 0.771 | (0.767–0.775) |
| Romano + CDS-1 | Age + gender + Romano + CDS-1 | 4 | 0.777 | (0.773–0.780) |
| Romano + # of RX* | Age + gender + Romano + numb. of Rx | 4 | 0.775 | (0.771–0.779) |
| Romano (17 cat.) † | Age + gender + $\text{Romano}_{17}$ | 19 | 0.781 | (0.777–0.784) |
| Elixhauser (30 cat.) | Age + gender + $\text{Elixhauser}_{30}$ | 32 | 0.781 | (0.778–0.785) |
| Elixhauser plus †† | Age + gender + $\text{Elixhauser}_{30}$ + AMI + stroke | 34 | 0.783 | (0.779–0.786) |

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

†All Romano coded conditions were included as 17 binary categories in the model: 1 = condition present) or 0 = condition absent.

††Thirty Elixhauser indicators based on ICD-9-CM diagnoses plus two indicators for acute myocardial infarction (AMI) and stroke according to Romano's ICD-9-CM coding.

We then derived new weights for the Romano and the CDS-1 scores based on their component's independent association with dying during the following year. For the Romano score, the new weights were always in the same direction (larger than 0) as the original Charlson weights. Except for two conditions (dementia and AIDS) the new weights departed by only one point from the original weight, if at all (Table 4). Drug therapy for five conditions of the CDS-1 were associated with a reduced one-year mortality (gold salts for rheumatoid arthritis; cromolyn in asthma or rhinitis therapy; anti-acne therapy; lipid lowering therapy; ergot derivates for treating migraine) compared to receiving no prescription for these conditions (Table 5). Of these, only antilipidemics were used by a substantial proportion (10 percent) of the New Jersey Medicare/ PAAD population and were associated with a 40 percent reduction in one-year mortality.

The performance of the Romano and CDS-1 scores using the empirical weights for elderly improved in the development sample by 6.6 percent and 35 percent compared to the scores with the original weights (Table 5). When the Romano score was combined with the number of distinct prescriptions and an indicator for nursing home residency in addition to age and gender, the c-statistic increased slightly from 0.779 to 0.803.

Table 4: Conditions According to the Romano Adaptation of the Charlson Index for Use with Claims Data (Romano, Roos, and Jollis 1993a,1993b) with original Charlson Index Weights (Charlson et al. 1987) and Weights Derived from New Jersey Medicare Data ($N = 235,881$) for the Same Conditions

| *Conditions* | *Prevalence in New Jersey in %* | *Original Charlson Weights* | *New Jersey Odds- ratio Estimates** | *95% Confidence Interval* | *Assigned NJ Medicare Weights*** |
|---|---|---|---|---|---|
| Myocardial infarct | 4.3 | 1 | 1.23 | (1.15–1.31) | 1 |
| Congestive heart failure | 15.1 | 1 | 2.09 | (2.01–2.17) | 2 |
| Peripheral vascular disease | 13.3 | 1 | 1.55 | (1.49–1.61) | 1 |
| Cerebrovascular disease | 11.4 | 1 | 1.42 | (1.36–1.48) | 1 |
| Dementia | 6.2 | 1 | 2.16 | (2.06–2.27) | 3 |
| Chronic pulmonary disease | 12.4 | 1 | 1.66 | (1.59–1.73) | 2 |
| Connective tissue disease | 2.2 | 1 | 1.09 | (0.98–1.21) | 0 |
| Ulcer disease | 3.4 | 1 | 1.03 | (0.96–1.11) | 0 |
| Mild liver disease | 0.3 | 1 | 1.73 | (1.41–2.12) | 2 |
| Diabetes | 12.0 | 1 | 1.37 | (1.31–1.44) | 1 |
| Hemiplegia | 2.1 | 2 | 1.44 | (1.33–1.56) | 1 |
| Moderate or severe renal disease | 1.6 | 2 | 2.54 | (2.34–2.76) | 3 |
| Diabetes with end organ damage | 5.1 | 2 | 1.57 | (1.48–1.67) | 2 |
| Any tumor | 5.8 | 2 | 1.85 | (1.75–1.95) | 2 |
| Leukemia | † | 2 | | | |
| Lymphoma | † | 2 | | | |
| Moderate or severe liver disease | 0.1 | 3 | 3.24 | (2.47–4.26) | 4 |
| Metastatic solid tumor | 1.7 | 6 | 5.94 | (5.50–6.40) | 6 |
| AIDS | 0.1 | 6 | 3.26 | (2.13–4.98) | 4 |

*Odds ratio of dying during the follow up comparing patients with the condition versus subjects without.

**A 35% increase in risk of dying is reflected in a one-point increase in weights (see appendix for scoring rule).

†Leukemia and lymphoma are included in the "any tumor" category in Romano's adaptation of the Charlson Index for use with claims data (Romano, Roos, and Jollis 1993a,1993b).

### *Validation of Improved Performance*

In a comparably defined validation sample of Pennsylvania Medicare/PACE enrollees, the performance of the Romano score improved by 8.3 percent, and for the CDS-1 by 43 percent, when the new weights for elderly were applied compared to the scores with the original weights (Table 6). A model combining the Romano score with the number of prescription drugs performed best ($c = 0.767$).

## DISCUSSION

Evaluating the risk of mortality in large groups of older patients will become increasingly important with the aging of populations throughout the industrialized world. We tested the properties of several comorbidity adjustment scores and calibrated weights to better fit elderly Medicare populations, which substantially improved the capacity to predict mortality and control confounding in epidemiologic studies.

In two Medicare populations, an ICD-9-based score performed better than a medication-based score at predicting one-year mortality of a large typical population of patients over 65. Some of the inferiority of the medication-based score in predicting short-term outcomes is apparently attributable to the use of preventive treatments or treatment for benign conditions in healthier patients. For example, elderly women who are generally healthy and aware of health risks are likely to take lipid-lowering drugs and hormone replacement therapy. Such patients are likely to fare better than patients whose primary diagnosis has a poor short-term prognosis that may deter treatment of secondary conditions. This is consistent with earlier findings that sicker patients are less likely to be treated for comorbid conditions (Redelmeier, Tan, and Booth 1998), particularly if these conditions are not immediately life threatening and medications for treating these conditions have some preventive effects, for example, oral antidiabetic agents (Glynn et al. 1999b) or lipid lowering drugs (Glynn et al. 2001). These drugs are less frequently prescribed to very sick patients, and therefore users are often in fact healthier than would be suggested by their medication-based scores.

The CDS-1 weights were derived for a younger population (18 to 65 years), which may be an additional cause of the weaker performance in a Medicare population. Adjusting the weights for an elderly population and allowing negative weights for some conditions improved the performance of medication-based CDS-1 by up to 43 percent. The improvement in the Romano and CDS-1 scores was confirmed in an independent validation sample of Pennsylvania Medicare enrollees.

Even with the improved confounder adjustment from empirically derived weights for elderly, the medication-based CDS-1 did not perform as well as the diagnosis-based Romano score. Pending confirmation of these findings in other populations, this suggests that medication-based scores should be used only in situations when pharmacy claims data are of much better quality than diagnoses, or the only source of information.

Table 5: Weights Assigned to Conditions in the Chronic Disease Score (Von Korff, Wagner, and Saunders 1992) and weights derived from New Jersey Medicare Data ($N = 235{,}881$) for the Same Conditions

| *Conditions and Medications Attributed to These Conditions According to Von Korff et al. (1992)* | | | *Prevalence in NJ in %* | *Original CDS-1 Weights* | *NJ Odds Ratio Estimates** | *95% Confidence Interval* | *Assigned NJ Medicare Weights*** |
|---|---|---|---|---|---|---|---|
| Heart disease: | (1) Anti-coagulants, hemostatics, | one class | 42.3 | 3 | 1.45 | (1.39–1.51) | 1 |
| | (2) Cardiac agents, ACE inhibitors, | two classes | 16.3 | 4 | 2.53 | (2.42–2.64) | 3 |
| | (3) Loop diuretics | three classes | 2.3 | 5 | 3.14 | (2.89–3.41) | 4 |
| Respiratory illness: | (1) Isoproterenol, (2) Beta-adrenergic | one class | 7.0 | 2 | 1.49 | (1.41–1.58) | 1 |
| | (3) Xanthines, (4) Bronchodilators and mucolytic but excluding cromolyn, (5) Epinephrine | two classes | 4.6 | 3 | 1.89 | (1.77–2.02) | 2 |
| Asthma, rheumatism: | Glucocorticoids | | 8.7 | 3 | 1.32 | (1.25–1.40) | 1 |
| Rheumatoid arthritis: | Gold salts | | 0.1 | 3 | 0.80 | (0.40–1.61) | −1 |
| Cancer: | Antineoplastics | | 3.1 | 3 | 2.18 | (2.03–2.34) | 3 |
| Parkinson's disease: | L-dopa | | 1.8 | 3 | 2.04 | (1.87–2.24) | 2 |
| Hypertension: | (1) Antihypertensives (except ACEI) or calcium channel blockers | class (1) | 18.4 | 2 | 0.95 | (0.91–0.99) | 0 |
| | (2) Beta blockers, diuretics | class (2) not (1) | 10.4 | 1 | 1.03 | (0.97–1.08) | 0 |
| Diabetes: | Insulin, oral hypoglycemics | | 16.2 | 2 | 1.42 | (1.36–1.48) | 1 |
| Epilepsy: | Anticonvulsants | | 4.0 | 2 | 1.95 | (1.83–2.08) | 2 |
| Asthma, rhinitis: | Cromolyn | | 0.6 | 2 | 0.54 | (0.43–0.69) | −2 |
| Acne: | (1) Antiacne tretinoin, (2) topical macrolides | | 0.2 | 1 | 0.49 | (0.28–0.86) | −2 |
| Ulcers: | Cimetidine | | 27.7 | 1 | 1.30 | (1.25–1.34) | 1 |
| Glaucoma: | Ophthalmic miotics | | 10.1 | 1 | 0.94 | (0.89–0.99) | 0 |
| Gout, hyperuricemia: | Uric acid agents | | 3.1 | 1 | 1.16 | (1.08–1.26) | 0 |
| High cholesterol: | Antilipidemics | | 10.0 | 1 | 0.59 | (0.55–0.63) | −2 |
| Migraines: | Ergot derivates | | 0.7 | 1 | 0.84 | (0.69–1.03) | −1 |
| Tuberculosis: | Antitubercular agents | | 0.2 | 1 | 1.81 | (1.38–2.38) | 2 |

*Odds ratio of dying during the follow-up comparing patients with the condition versus subjects without.

**A 35% increase in risk of dying is reflected in a one-point increase in weights (see appendix for scoring rule).

Earlier studies have also shown that the number of distinct medications received during the previous year is a better predictor of mortality than the Chronic Disease Score (CDS-1). Adding the number of medications to the Romano score or the CDS-1 improved the performance of both scores for one-year mortality and for a range of health care utilization endpoints (Schneeweiss et al. 2001). Although for some conditions the CDS-1 can capture the use of multiple drug therapy versus monotherapy, it fails to do so for other diagnoses and does not account for medication changes during the progression of disease; these would all be accounted for by the number of distinct medications received during the previous year. This suggests that the number of different medications used should therefore be included to improve the adjustment for comorbidity when possible.

Modeling comorbidity by including all component conditions of the Romano score or the Elixhauser system as indicator terms did slightly improve performance, but with the price of 30 additional parameters to estimate. Inclusion of this many covariates may not be feasible and may seriously affect the statistical power when analyzing small patient subgroups with rare exposures or outcomes (Peduzzi et al. 1996). The application is further limited by computational constraints if comorbidity is modeled as a time-varying covariate, which is an increasingly popular analytic option for longitudinal claims data.

With our improved comorbidity measures, one-year mortality could be predicted with a c-statistic of 0.80, a value that is considered as excellent prediction (Hosmer and Lemeshow 2000). This model adjusted only for three sociodemographic variables (age, gender, nursing home residency) and two comorbidity measures (Romano with new weights, number of distinct prescription drugs). This does not include a specific primary diagnosis or other study specific markers that would likely further improve control for confounding in addition to comorbidity adjustment.

Although it had been shown in a simplified model how predictive validity can be translated into confounder adjustment (Schneeweiss et al. 2001), some benchmarks are helpful to appreciate this performance. Hannan et al. (1992) reported a c-statistic of 0.742 for the prediction of in-hospital mortality in 22,827 patients with bypass surgery in New York State based on detailed discharge data that included demographics and up to four comorbidities per patient. After including important clinical predictors including ejection fraction, $>90$ percent narrowing of the left main vessel, and reoperation, the c-statistic improved to 0.790. The National Cholesterol Education Program guidelines I and II predict cardiovascular mortality with c-statistics between 0.72 and 0.74 (Grover, Coupal, and Hu 1995). Predicting

Table 6: Improvement in Performance of Comorbidity Scores by Using Medicare Weights in the Development and Validation Samples

| *Population Sample* | *Score** | *Old Weights* | *New Weights* | *Performance Improvement in %* |
|---|---|---|---|---|
| Development sample (New Jersey/ PAAD) | CDS-1 | 0.703 (0.699;0.707) | 0.727 (0.723;0.731) | 35 % |
| | Romano | 0.771 (0.767;0.775) | 0.780 (0.776;0.783) | 7.4 % |
| | Romano + number of Rx | 0.775 (0.771;0.779) | 0.783 (0.779;0.786) | |
| Validation sample (Pennsylvania/PACE) | CDS-1 | 0.695 (0.691;0.699) | 0.715 (0.711;0.719) | 43 % |
| | Romano | 0.757 (0.754;0.761) | 0.765 (0.762;0.769) | 8.3 % |
| | Romano + number of Rx | 0.760 (0.756;0.764) | 0.767 (0.763;0.770) | |

*Logistic regression models included age and gender in addition to the scores.

coronary heart disease in the Framingham Heart Study using clinically measured risk factors and comorbidities produced c-statistics between 0.68 and 0.77 (Wilson et al. 1998). From this and other examples it appears that large investments yield only small numeric gains in c-statistics above 0.75. Whether those gains are worth their price depends on the benefits of a "truer" analysis and the costs of error, which are unique to each problem. Earlier studies on comorbidity scores found c-statistics ranging between 0.64 and 0.87 for the different adaptations of the Charlson Index for use with claims data (Schneeweiss and Maclure 2000). However, these values were derived from diverse populations. Because population characteristics as well as the type of outcome are important determinants of predictive validity besides the inherent performance of a score, the usefulness of these numbers is limited.

Comorbidity scores are useful because they are easy to apply and they save time and resources (a major issue when analyzing massive health care databases and testing multiple hypotheses). They 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 (Maclure and Schneeweiss 2001). Even scores with improved performance impose a functional relation between comorbidities and outcome, which is likely to differ for specific subpopulations (Michels,

Greenland, and Rosner 1998; Katz and Faoxman 1993). In an ideal world, one would have enough data and well-measured variables to completely adjust statistically for differences in risk. Such adjustments would be tailored to the outcome measure (e.g., death, readmission) and the population under study. However, comorbidity scores with preassigned weights are still useful for 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.

When using comorbidity scores in elderly populations, we recommend using the diagnosis-based approach described above with our new weights. Our data suggest that this will yield the best prediction of outcomes and thus control for confounding in epidemiologic studies of those 65 years or older. This will be particularly useful in small studies or in subgroup analyses with limited sample size when it may be difficult to derive study-specific weights.

## APPENDIX

Medicare weighting rule:

Based on multivariate logistic regression estimates we increased the weights by 1 with each 0.3 increase in the ln (OR). A weight of 1 therefore refers to an exp (0.30) = 35 percent increase in risk of dying during the follow-up year.

$$\text{if } 0 \leq \ln(\text{OR}) \leq 0.15 \text{ then weight} = 0$$
$$\text{if } 0.15 < \ln(\text{OR}) \leq 0.45 \text{ then weight} = 1$$
$$\text{if } 0.45 < \ln(\text{OR}) \leq 0.75 \text{ then weight} = 2$$
$$\text{if } 0.75 < \ln(\text{OR}) \leq 1.05 \text{ then weight} = 3$$
$$\text{if } 1.05 < \ln(\text{OR}) \leq 1.35 \text{ then weight} = 4$$
$$\text{if } 1.35 < \ln(\text{OR}) \leq 1.65 \text{ then weight} = 5$$
$$\text{if } 1.65 < \ln(\text{OR}) \leq 1.95 \text{ then weight} = 6$$

$$\text{if } 0 > \ln(\text{OR}) \geq -0.15 \text{ then weight} = 0$$
$$\text{if } -0.15 > \ln(\text{OR}) \geq -0.45 \text{ then weight} = -1$$
$$\text{if } -0.45 > \ln(\text{OR}) \geq -0.75 \text{ then weight} = -2$$
$$\text{if } -0.75 > \ln(\text{OR}) \geq -1.05 \text{ then weight} = -3$$
$$\text{if } -1.05 > \ln(\text{OR}) \geq -1.35 \text{ then weight} = -4$$

## REFERENCES

American Society of Health-System Pharmacists. 1996. *AHFS Drug Information 96.* Bethesda, MD: American Society of Health-System Pharmacists.

Ash, A. S., and M. Shwartz. 1997. "Evaluating the Performance of Risk-Adjustment Methods: Dichotomous Outcomes." In *Risk Adjustment for Measuring Healthcare Outcomes*, 2nd ed, edited by L. I. Iezzoni. Chicago: Health Administration Press.

Avorn, J. 1991. "Epidemiology in Plato's Cave: Claims Data and Clinical Reality." *Journal of Clinical Epidemiology* 44: 867–9.

Bright, R. A., J. Avorn, and D. E. Everitt. 1989. "Medicaid Data as a Resource For Epidemiologic Studies: Strengths and Limitations." *Journal of Clinical Epidemiology* 42: 937–45.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40: 373–83.

Clark, D. O., M. Von Korff, K. Saunders, W. M. Baluch, and G. E. Simon. 1995. "A Chronic Disease Score with Empirically Derived Weights." *Medical Care* 33: 783–95.

Deyo, R. A., D. C. Cherkin, and M. A. Ciol. 1992. "Adapting a Clinical Comorbidity Index for Use with ICD-9-CM Administrative Databases." *Journal of Clinical Epidemiology* 45: 613–9.

D'Hoore, W., A. Bouckaert, and C. Tilquin. 1996. "Practical Considerations on the Use of the Charlson Index with Administrative Data Bases." *Journal of Clinical Epidemiology* 49: 1429–33.

D'Hoore, W. C. Sicotte, and C. Tilquin. 1993. "Risk Adjustment in Outcome Assessment: The Charlson Comorbidity Index." *Methods of Information in Medicine* 32: 382–7.

Elixhauser, A., C. Steiner, R. Harris, and R. M. Coffey. 1998. "Comorbidity Measures for Use with Administrative Data." *Medical Care* 36: 8–27.

Fisher, E. S., F. S. Whaley, M. Krushat, D. J. Malenka, C. Fleming, J. A. Baron, and D. C. Hsia. 1992. "The Accuracy of Medicare's Hospital Claims Data: Progress Has Been Made but Problems Remain." *American Journal of Public Health* 82: 243–8.

Fowles, J. B., A. G. Lawthers, J. P. Weiner, D. W. Garnick, D. S. Petril, and R. H. Palmer. 1995. "Agreement between Physician's Office Records and Medicare Part B Claims Data." *Health Care Finance Review* 16: 189–99.

Ghali, W. A., R. E. Hall, A. K. Rosen, A. S. Ash, and M. A. Moskowitz. 1996. "Searching for an Improved Clinical Comorbidity Index for Use with ICD-9-CM Administrative Data." *Journal of Clinical Epidemiology* 49: 273–8.

Glynn, R. J., E. L. Knight, R. Levin, and J. Avorn. 2001. "Paradoxical Relations of Drug Treatment with Mortality in Older Persons." *Epidemiology* 12: 682–9.

Glynn, R. J., M. Monane, J. H. Gurwitz, I. Choodnovskiy, and J. Avorn. 1999a. "Agreement between Drug Treatment and a Discharge Diagnosis of Diabetes." *American Journal of Epidemiology* 149: 541–9.

———. 1999b. "Aging, Comorbidity, and Reduced Rates of Drug Treatment for Diabetes Mellitus." *Journal of Clinical Epidemiology* 52: 781–90.

Greenfield, S., and E. C. Nelson. 1992. "Recent Developments and Future Issues in the Use of Health Status Assessment Measures in Clinical Setting." *Medical Care* 30: MS23–41.

Grover, S., L. Coupal, and X-P. Hu. 1995. "Identifying Adults at Increased Risk of Coronary Disease: How Well Do the Current Cholesterol Guidelines Work?" *Journal of the American Medical Association* 274: 801–6.

Hannan, E. L., H. Kilburn, M. L. Lindsey, and R. Lewis. 1992. "Clinical versus Administrative Data Bases for CABG Surgery: Does It Matter?" *Medical Care* 30: 892–907.

Hosmer, D. W., and S. Lemeshow. 2000. *Applied Logistic Regression*, 2nd ed. New York: Wiley.

Iezzoni, L. I. 1997. *Risk Adjustment for Measuring Healthcare Outcomes*, 2nd ed. Chicago: Health Administration Press.

Iezzoni, L. I., S. M. Foley, J. Daley, J. Hughes, E. S. Fisher, and T. Heeren. 1992. "Comorbidities, Complications, and Coding Bias: Does the Number of Diagnosis Codes Matter in Predicting In-Hospital Mortality?" *Journal of the American Medical Association* 267: 2197–2203.

Katz, D., and B. Foxman. 1993. "How Well Do Prediction Equations Predict? Using Receiver Operating Characteristic Curves and Accuracy Curves to Compare validity and Generalizability." *Epidemiology* 4: 319–26.

Last, J. M. 1995. *A Dictionary of Epidemiology*, 3d ed. New York: Oxford University Press.

Lessler, J. T., and B. S. H. Harris. 1984. *Medicaid Data As a Source for Postmarketing Surveillance Information, Final Report.* Research Triangle Park, NC: Research Triangle Institute.

Liebetrau, A. M. 1983. *Measures of Association, Quantitative Application in the Social Sciences*, 32. Beverly Hills, CA: Sage.

Maclure, M., and S. Schneeweiss. 2001. "Causation of Bias: The Episcope." *Epidemiology* 12: 114–22.

Melfi, C., E. Holleman, D. Arthur, and B. Katz. 1995. "Selecting a Patient Characteristics Index for the Prediction of Medical Outcomes Using Administrative Claims Data." *Journal of Clinical Epidemiology* 48: 917–26.

Michels, K. B., S. Greenland, and B. A. Rosner. 1998. "Does Body Mass Index Adequately Capture the Relation of Body Composition and Body Size to Health Outcomes?" *American Journal of Epidemiology* 147 (2): 167–72.

Peduzzi, P., J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein. 1996. "A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis." *Journal of Clinical Epidemiology* 49: 1373–9.

Poses, R. M., W. R. Smith, D. K. McClish, and M. Anthony. 1995. "Controlling for Confounding by Indication for Treatment. Are Administration Data Equivalent to Clinical Data?" *Medical Care* 33: AS36–46.

Ray, W. A., C. M. Stein, K. Hall, J. R. Daugherty, and M. R. Griffin. 2002. "Non-steroidal Anti-inflammatory Drugs and Risk of Serious Coronary Heart Disease: An Observational Cohort Study." *Lancet* 359: 118–23.

Redelmeier, D. A., S. H. Tan, and G. L. Booth. 1998. "The Treatment of Unrelated Disorders in Patients with Chronic Medical Diseases." *New England Journal of Medicine* 338: 1516–20.

Romano, P. S., and D. H. Mark. 1994. "Bias in the Coding of Hospital Discharge Data and Its Implications for Quality Assessment." *Medical Care* 32: 81–90.

Romano, P. S., L. L. Roos, and J. G. Jollis. 1993a. "Adapting a Clinical Comorbidity Index for Use with ICD-9-CM Administrative Data: Differing Perspectives." *Journal of Clinical Epidemiology* 46: 1075–9.

——— . 1993b. "Further Evidence Concerning the Use of a Clinical Comorbidity Index with ICD-9-CM Administrative Data." *Journal of Clinical Epidemiology* 46: 1085–90.

Roos, L. L., S. M. Sharp, and M. M. Cohen. 1991. "Comparing Clinical Information with Claims Data: Some Similarities and Differences." *Journal of Clinical Epidemiology* 44: 881–8.

Roos, L. L., S. M. Sharp, M. M. Cohen, and A. Wajda. 1989. "Risk Adjustment in Claims-Based Research: The Search for Efficient Approaches." *Journal of Clinical Epidemiology* 42: 1193–206.

Schneeweiss, S., and M. Maclure. 2000. "Use of Comorbidity Scores for Control of Confounding in Studies Using Administrative Databases." *International Journal of Epidemiology* 29: 891–8.

Schneeweiss, S., J. Seeger, M. Maclure, P. Wang, J. Avorn, and R. J. Glynn. 2001. "Performance of Comorbidity Scores to Control for Confounding in Epidemiologic Studies Using Claims Data." *American Journal of Epidemiology* 154: 854–64.

Von Korff, M., E. H. Wagner, and K. Saunders K. 1992. "A Chronic Disease Score from Automated Pharmacy Data." *Journal of Clinical Epidemiology* 45: 197–203.

Walker, A. M. 1996. "Confounding by Indications." *Epidemiology* 7: 335–56.

Wang, P. S., D. H. Solomon, H. Mogun, and J. Avorn. 2000. "HMG-CoA Reductase Inhibitors and the Risk of Hip Fractures in Elderly Patients." *Journal of the American Medical Association* 283: 3211–6.

Wang, P. S., A. Walker, M. Tsuang, E. J. Orav, R. Levin, and J. Avorn. 2000. "Strategies for Improving Comorbidity Measures Based on Medicare and Medicaid Claims Data." *Journal of Clinical Epidemiology* 53: 571–8.

Wilson, P. W. F., R. B. D'Agostino Sr., L. Levy, A. Belanger, H. Silbershatz, and W. B. Kannel. 1998. "Prediction of Coronary Heart Disease Using Risk Factor Categories." *Circulation* 97: 1837–47.

Yuan, Z., G. S. Cooper, D. Einstadter, R. D. Cebul, and A. A. Rimm. 2000. "The Association between Hospital Type and Mortality and Length of Stay." *Medical Care* 38: 231–45.

Zhang, J. X., T. J. Iwashyna, and N. A. Christakis. 1999. "The Performance of Different Lookback Periods and Sources of Information for Charlson Comorbidity Adjustment in Medicare Claims." *Medical Care* 37: 1128–39.