

Why Summary Comorbidity Measures Such As the Charlson Comorbidity Index and Elixhauser Score Work

Steven R. Austin,* Yu-Ning Wong, MD,† Robert G. Uzzo, MD,‡ J. Robert Beck, MD,§
and Brian L. Egleston, PhD||

Background: Comorbidity adjustment is an important component of health services research and clinical prognosis. When adjusting for comorbidities in statistical models, researchers can include comorbidities individually or through the use of summary measures such as the Charlson Comorbidity Index or Elixhauser score. We examined the conditions under which individual versus summary measures are most appropriate.

Methods: We provide an analytic proof of the utility of comorbidity summary measures when used in place of individual comorbidities. We compared the use of the Charlson and Elixhauser scores versus individual comorbidities in prognostic models using a SEER-Medicare data example. We examined the ability of summary comorbidity measures to adjust for confounding using simulations.

Results: We devised a mathematical proof that found that the comorbidity summary measures are appropriate prognostic or adjustment mechanisms in survival analyses. Once one knows the comorbidity score, no other information about the comorbidity variables used to create the score is generally needed. Our data example and simulations largely confirmed this finding.

Conclusions: Summary comorbidity measures, such as the Charlson Comorbidity Index and Elixhauser scores, are commonly used for clinical prognosis and comorbidity adjustment. We have provided a theoretical justification that validates the use of such scores under many conditions. Our simulations generally confirm the utility of the summary comorbidity measures as substitutes for use of the individual comorbidity variables in health services research. One caveat is that a summary measure may only be as good as the variables used to create it.

From the *Whiting School of Engineering Undergraduate Student, Johns Hopkins University, Baltimore, MD; Departments of †Medical Oncology; ‡Surgery, Fox Chase Cancer Center; §Academic Affairs, Fox Chase Cancer Center, Temple University Health System; and ||Biostatistics and Bioinformatics Facility, Fox Chase Cancer Center, Temple University Health System, Philadelphia, PA.

Supported by the National Institutes of Health, National Cancer Institute, grants P30CA006927 and R03CA152388.

The authors declare no conflict of interest.

Reprints: Brian L. Egleston, PhD, Biostatistics and Bioinformatics Facility, Fox Chase Cancer Center, Temple University Health System, 333 Cottman Avenue, Philadelphia, PA 19111, E-mail: brian.egleston@fccc.edu.

Copyright © 2013 Wolters Kluwer Health, Inc. All rights reserved.
ISSN: 0025-7079/15/5309-0e65

Key Words: Charlson Comorbidity Index, Elixhauser score, comorbidity adjustment, prognostic models, comorbidity summary measures, SEER-Medicare, kidney cancer

(*Med Care* 2015;53: e65–e72)

Baseline comorbidity adjustment is an important component of health services research and clinical prognosis. Researchers have widely used summary measures for comorbidity adjustment in outcome studies that use administrative health data.^{1–3} When adjusting for comorbidities, researchers may consider comorbidities individually or through the use of summary measures such as the Charlson Comorbidity Index^{4–6} or the Elixhauser comorbidity measures.^{7,8}

In statistical models, investigators might incorporate comorbidities, such as diabetes or heart disease, by including indicator covariates to denote whether the condition is present (the indicator equals 1 if the condition is present, 0 otherwise). In contrast, summary measures, such as the Charlson Comorbidity Index, attach weights to each condition, and then sum the weights of those conditions that are present in an individual.⁴ The Charlson Comorbidity Index is based on a number of conditions that are each assigned an integer weight from 1 to 6, with a weight of 6 representing the most severe morbidity. The summation of the weighted comorbidity scores results in a summary score.

In this paper, we use the Charlson Comorbidity Index as the main example of a comorbidity summary measure because of its widespread use. A Web of Science search finds that the original and derivative papers concerning the Charlson Comorbidity Index have been cited over 8800 times. Although initially developed for use with medical records data, the Charlson Comorbidity Index has been adapted for use with health claims data.^{5,6,9} The validity of the Charlson Comorbidity Index as well as its adaptations have been investigated in multiple studies.^{10–12} The success of the index has prompted inquiry into further adaptations of the Charlson Comorbidity Index using questionnaire-based and physician claims–based indices.^{13–15}

Although the Charlson Comorbidity Index is commonly used, competitor comorbidity measures have been developed. As an additional example, we also investigate properties of the more recently developed Elixhauser score.⁸

Like the Charlson score, the Elixhauser score was derived using regression estimates.

Whether it is better to use the Charlson Comorbidity Index or the individual comorbidities separately in statistical models is an open question. For example, using ICD-10 data from a multinational group of patients, Sundararajan et al¹⁶ investigated the use of 17 comorbidities as individual variables to predict mortality. The authors found that the individual comorbidities consistently had better prognostic ability than the Charlson summary score. Conversely, Lieffers et al¹ found that the addition of the Charlson comorbidities as individual binary variables generated a model that was no better than a comparison model created using the Charlson Comorbidity Index in a sample of patients with colorectal cancer. Further, some investigators have advocated that researchers develop weights for comorbidity indices using their own data rather than published weights. For example, Ghali et al¹⁷ found that the prognostic ability of summary measures created using one's own data was superior to that of measures derived using published algorithms in a sample of patients undergoing coronary artery bypass graft surgery. Using data from patients receiving angiotensin-converting enzyme inhibitors or calcium-channel blockers, Schneeweiss and Maclure¹⁸ similarly suggested that confounding can better be controlled by deriving study-specific weights.

Hansen¹⁹ provided general theoretical justification for prognostic scores (the Charlson Comorbidity Index is a prognostic score), by demonstrating that comorbidity summary measures can have properties similar to propensity scores^{20,21} in removing confounding in observational studies. However, as with the propensity score, Hansen suggests that the summary measures, or prognostic scores as he terms them, be estimated using a researcher's own data.¹⁹ Further, Hansen suggests that prognostic scores be estimated only using data from the control group in a 2-arm study.¹⁹

Here, we examine analytic conditions under which summary measures are appropriate. Unlike others who examined which comorbidity scores are most useful in specific contexts,²² we focus on general approaches to comorbidity summary measure development. In particular, we present a mathematical proof that justifies the use of appropriately constructed summary measures in place of the individual components. Unlike Hansen,¹⁹ we consider comorbidity measures developed similarly to the Charlson Comorbidity Index. To investigate some finite sample (real world) characteristics of the Charlson Comorbidity Index and Elixhauser score, we used a data example and simulations.

METHODS

We developed a mathematical proof that generally validates the use of summary measures. For ease of presentation, we assume baseline (ie, pretreatment) comorbidities.

Mathematical Justification

We provide a proof that a comorbidity summary score based on a hazard is heuristically a balancing score when used in survival analyses. That is, if one knows the hazard of death conditional on the comorbidity score, then knowledge of the

covariates used to create the score does not provide additional information about the hazard. In the Appendix, we provide a similar proof for comorbidity scores estimated by linear regressions when used in subsequent linear regressions.

The proof provides technical audiences with a rigorous justification for the use of comorbidity scores. We examine a hazard, as the Charlson Comorbidity Score was estimated using summed hazard ratios from a Cox regression. Our proof is more directly applicable to measures like the Elixhauser score in which regression coefficients are summed. Let T represent the survival time, $f_T(t)$ its probability density function, and $S_T(t) = P(T > t)$. Let $h(\cdot)$ represents the hazard, whereas X is a vector of covariates, $X = \{X_1, \dots, X_n\}'$. Also, x is a vector of the realized values, $X = \{x_1, \dots, x_n\}'$.

For ease of notation let $b(X)$ represent a comorbidity score derived from a hazard rate; $b(X) = h(t|X) = \frac{f_T(t|X)}{S_T(t|X)}$. To prove that survival time is independent of the covariates given the comorbidity score (ie, the balancing property), it is necessary to show that:

$$h(t|X, b(X)) = h(t|b(X)).$$

As $b(X)$ is a function of X , it follows that $h(t|X, b(X)) = h(t|X)$. Thus, for the proof it is sufficient to show that $h(t|X) = h(t|b(X))$. Now,

$$\begin{aligned} h(t|b(X)) &= \frac{f_T(t|b(X))}{S_T(t|b(X))} \\ &= \int_{x_n \in \text{Range } X_n} \dots \int_{x_1 \in \text{Range } X_1} \frac{f_T(t|X=x, b(X))}{S_T(t|X=x, b(X))} \times \\ &\quad f(x_1, \dots, x_n | b(X)) dx_1 \dots dx_n \\ &= \int_{x_n \in \text{Range } X_n} \dots \int_{x_1 \in \text{Range } X_1} \frac{f_T(t|X=x)}{S_T(t|X=x)} \times \\ &\quad f(x_1, \dots, x_n | b(X)) dx_1 \dots dx_n \\ &= \int_{x_n \in \text{Range } X_n} \dots \int_{x_1 \in \text{Range } X_1} b(x) f(x_1, \dots, x_n | b(X)) dx_1 \dots dx_n \\ &= E[b(X) | b(X)], \text{ where } E[\cdot] \text{ denotes the expectation} \\ &= b(X) \\ &= h(t|X). \end{aligned}$$

Hence, it is shown that

$$\begin{aligned} h(t|X, b(X)) &= h(t|X) \\ &= h(t|b(X)). \end{aligned}$$

This demonstrates the balancing property of measures derived analogously to the Charlson Comorbidity Index when used in survival analyses. Of note is that this proof assumes that we know the true comorbidity score, $b(X)$. In practice, we only know an estimate of $b(X)$ on the basis of a model. Although we often assume that the estimator of $b(X)$ converges to the truth as the sample size grows (an asymptotic result), there may be some bias or efficiency effects of using the estimate in small samples. We explore this in the simulation section.

Data Example

For the data example, we used Surveillance Epidemiology and End Results (SEER) data that had been linked to Medicare claims data.²³ The SEER database is maintained by the National Cancer Institute and currently has data on demographic and tumor characteristics about incident cancer cases in approximately 25% of the United States. SEER data can be linked to Medicare claims to find additional information on treatment and comorbidities. Medicare covers almost all individuals over 65 years of age in the SEER database. Fee-for-service claims from Medicare Part A and Part B provide a record of treatments obtained before and after cancer diagnosis.

We included patients over 66 years old diagnosed from 1995 to 2007 with surgically treated early-stage (localized) kidney cancer in the SEER-Medicare data. We restricted the sample to those over 66 years old so that individuals would have at least 1 year prediagnosis Medicare claims data. We chose a group with localized kidney cancer, as such tumors are often slow growing, and many patients are likely to die from their comorbidities rather than the cancer itself.²⁴

We coded the Charlson Comorbidity Index using an algorithm for claims data.²³ We used a similar algorithm to identify Elixhauser measures.²⁵ As everyone in the sample had a kidney cancer diagnosis, cancer diagnosis was not used to calculate the scores. Also, the Elixhauser program does not calculate the cardiac arrhythmia indicator because of Dr Elixhauser's "concerns about reliability."²⁵ To be considered a comorbidity and not a "rule-out" diagnosis, 2 Medicare claims at least 30 days apart had to be found in the 1-year period before kidney cancer diagnosis.

For comparison purposes, we examined the discriminative ability of using the Charlson and Elixhauser scores and the respective individual comorbidities in Cox proportional hazards regressions. In all models we included the following baseline characteristics: age at diagnosis, year of diagnosis, diameter of the tumor, sex, race/ethnicity (5 categories: Hispanic or non-Hispanic black, white, Asian, other), and marital status (married/not married). We used the Harrell concordance index (C-index) to compare the 2 methods of incorporating comorbidities.²⁶ We examine the concordance statistic as it is often of interest to health service researchers. A C-index of 0.5 indicates that a model is not useful in predicting who will have longer survival among pairs of individuals, whereas a value of 1.0 indicates that the model has perfect discriminatory power.

Simulations

In the SEER-Medicare example, we examine prognostic characteristics in a surgically treated sample. We use simulated data to more generally explore the degree to which Charlson-type summary measures can adjust for confounding when assessing treatment effects. Here, we assume that comorbidities are related both to treatment assignment and to the outcome.

We generated simulated data under the following algorithm using the *rexp()* and *rbinom()* random number generators in the R programming language.²⁷ We performed simulations twice, assuming sample sizes of 250 or 2500. We used 2000 simulation iterations using the following algorithm.

Step 1. Variable Generation: Generate 4 comorbidity variables as Bernoulli with conditional probabilities $P(X_1=1)=0.23$, $P(X_2=1|X_1)=\text{expit}(-1.2+1.0X_1)$, $P(X_3=1|X_1, X_2)=\text{expit}(-1.9+1.1X_1+1.4X_2)$, $P(X_4=1|X_1, X_2, X_3)=\text{expit}(-1.6+0.4X_1+1.0X_2+0.5X_3)$, where $\text{expit}(q)=e^q/(1+e^q)$.

The first probability of 23% corresponds to the proportion of diabetes without complications in the dataset. The other coefficients were derived from logistic regressions using SEER-Medicare data (X_1 = diabetes, X_2 = congestive heart failure, X_3 = chronic renal failure, X_4 = cerebrovascular disease). A constant was added to the intercepts from the regressions to reduce sparse data in simulations (ie, iterations when certain comorbidities were not expressed and hence models would not converge).

Sensitivity analysis 1: We also examined simulations in which we multiplied the slopes by 1.5.

Step 2: Generate a random treatment indicator, Z , as Bernoulli ($Z=1$ if in treatment arm, 0 if in control arm) under 2 assumptions:

Coefficient set A: Weaker association of comorbidities with treatment assignment:

$$P(Z=1|X)=\text{expit}(-0.25+0.25X_1+0.25X_2+0.25X_3+0.25X_4).$$

Coefficient set B: Stronger association of comorbidities with treatment assignment:

$$P(Z=1|X)=\text{expit}(-2+2X_1+2X_2+2X_3+2X_4).$$

Step 3: Generate a survival time, T , in months as exponential with a density of $f(t|X,Z)=\lambda(X,Z)\exp(-\lambda(X,Z)t)$. We will examine our models with parameters:

$$\lambda(X,Z)=\exp(-1+2Z-2X_1-2X_2-2X_3-2X_4). \quad (1)$$

Setting all slopes equal to 2 (ie, hazard ratio = $\exp(2)=7.4$) ensured that both the comorbidities and treatment had a strong relationship with survival time.

Sensitivity analysis 2: we also repeated the analyses in which we dropped variables X_3 and X_4 from the models in steps 2 and 3. We kept all other parameters the same.

Step 4: Generate a censoring time, C , as exponential with $\lambda=0.03$. Let $Y=\min(T,C)$ be the observed time to censoring or death, and let $D=1$ if $T\leq C$, 0 otherwise. This created approximately 15% to 35% censored observations depending on the parameter assumptions above.

We estimated comorbidity indices in the simulations using Cox proportional hazards models as was done for the initial Charlson Comorbidity Index. To align our simulations directly with our proof, we estimate the comorbidity score by summing regression coefficients as was done for the Elixhauser score. We examined 5 simulated ways of using comorbidities for prognostic and comorbidity adjustment purposes in models that included the treatment indicator Z :

Method 1: The 4 simulated comorbidities entered as 4 untransformed covariates.

Method 2: A comorbidity score in which the coefficient weights were estimated from a single dataset and then the weights were applied in 250 randomly drawn experiments. This simulates the development of the Charlson score in which the Charlson weights were published using a single dataset, and then other researchers applied the weights in their own research. In this situation, data from individuals

in both the control and treatment groups were used for estimation.

Method 3: A comorbidity score estimated analogously to the preceding simulation (method 2) but only using data from the control group ($Z=0$). This applies the suggestion of Hansen¹⁹ that the comorbidity score only be estimated in the control group.

Method 4: A comorbidity score in which the coefficient weights were estimated with each iteration such that the same data were used to estimate the coefficient weights and estimate the adjusted treatment effect model. Data from individuals in both the control and treatment groups were used for estimation. This approach is more similar to the propensity score approach in which a model is estimated by each researcher, and researchers do not generally use propensity score weights previously published in the literature.

Method 5: A comorbidity score estimated analogously to the preceding simulation (method 4) but only using data from the control group ($Z=0$). This again takes the preceding approach but only estimates comorbidity scores using the control group as advocated by Hansen.¹⁹

RESULTS

SEER-Medicare Example

Our sample included 12,099 individuals diagnosed with localized kidney cancer (stages T1 or T2, node negative, and metastatic negative) diagnosed between 1995 and 2007 who had undergone surgical treatment for their cancer. Demographic and comorbidity information is listed in Table 1. The 15 specific noncancer Charlson comorbidities are listed in the bottom of Table 1.

In Table 2, we present the concordance statistics from 5 Cox proportional hazards regressions: (1) a model with no comorbidities; (2) a model with the 15 relevant Charlson comorbidities included as individual indicator variables; (3) a model with the Charlson Comorbidity Index as derived for health claims data^{5,6,15}; (4) a model with the 27 relevant Elixhauser comorbidities entered as indicator variables⁷; and (5) a model using an Elixhauser summary measure.⁸

The *C*-statistics in Table 2 were almost identical for the 2 models in which the Charlson score was entered as a summary measure and the Charlson comorbidities were entered individually. The Elixhauser comorbidity indicators entered individually had slightly better discriminative ability, and the Elixhauser score performed slightly worse. These results indicate that the widely used Charlson score performs approximately as well as the individual variables in predicting who will survive longer among pairs of individuals.

Simulations

In Table 3 to 6, we present the results of our simulations. The first model in each group becomes the gold standard. The true treatment effect is 2.0 [Eq. (1)]. When the sample size is relatively large ($n=2500$), we obtain the approximate treatment effect regardless of our assumptions. However, when the sample size is relatively small ($n=250$), we find that there is some bias in point estimates when using comorbidity scores, particularly when the comorbidities

TABLE 1. Characteristics of SEER-Medicare Sample [Mean or % (SD)]

N	12,099
Age (SD)	74.4 (5.7)
Year diagnosed (SD)	2002 (3.4)
Largest tumor dimension in mm (SD)	45.8 (26.2)
Female	43
Married	62
Race/ethnicity	
Asian	2
Black (non-Hispanic)	8
White (non-Hispanic)	86
Hispanic	2
Other	2
Charlson score (SD)	0.81 (1.19)
Charlson score = 0	55
Charlson score = 1	25
Charlson score = 2	11
Charlson score = 3	5
Charlson score = 4+	4
Elixhauser score (SD)	1.86 (4.16)
Elixhauser score = -1 to -9	10
Elixhauser score = 0	57
Elixhauser score = 1 to 5	18
Elixhauser score = 6 to 10	10
Elixhauser score = 11 to 40	5
Charlson comorbidities	
Myocardial infarction	3.5
Congestive heart failure	8.3
Peripheral vascular disease	5.1
Cerebrovascular disease	5.6
Chronic obstructive pulmonary disease	12.9
Dementia	0.5
Paralysis	0.5
Diabetes	23.0
Diabetes with sequelae	4.8
Chronic renal failure	5.4
Cirrhodites	0.4
Moderate-severe liver disease	0.1
Ulcers	1.6
Rheumatoid arthritis	2.5
AIDS	≤0.1
Elixhauser comorbidities	
Congestive heart failure	8.7
Valvular disease	6.4
Pulmonary circulation disease	1.1
Peripheral vascular disease	8.4
Paralysis	0.9
Other neurological disorders	3.3
Chronic pulmonary disease	12.9
Diabetes w/o chronic complications	17.8
Diabetes w/chronic complications	5.6
Hypertension uncomplicated	58.8
Hypertension complicated	9.4
Hypothyroidism	9.5
Renal failure	3.6
Liver disease	0.8
Peptic ulcer disease excluding bleeding	≤0.1
AIDS	≤0.1
Lymphoma	1.2
Rheumatoid arthritis/collagen vascular diseases	2.9
Coagulopathy	2.5
Obesity	2.2
Weight loss	1.6
Fluid and electrolyte disorders	8.1
Chronic blood loss anemia	1.4
Deficiency anemias	14.3
Alcohol abuse	0.4

(Continued)

TABLE 1. Characteristics of SEER-Medicare Sample [Mean or % (SD)] (*continued*)

Drug abuse	≤0.1
Psychoses	1.9
Depression	2.7

The proportions of individuals with Charlson comorbidities differ from those with similar Elixhauser comorbidities because of differences in claims-based codes used to identify comorbidities.

were strongly associated with treatment (Tables 3 and 4, coefficient set B, $n=250$, methods 2–5). The small sample size bias was larger when the comorbidities were more associated with each other (Table 4 compared with Table 3, $n=250$). However, there was also some slight upward bias (ie, the estimated coefficient was too large) when entering the comorbidities into the model as 4 separate indicator (yes/no) variables (Tables 3 and 4, method 1 when sample size = 250).

Using data only from the control group seemed to increase the small sample size bias (Tables 3 and 4, methods 3 and 5 compared with methods 2 and 4). Also, using one's own sample to estimate comorbidity score coefficients did not substantially affect the bias in the small sample setting (Tables 3 and 4, methods 4 and 5 compared with methods 2 and 3, $n=250$). In fact, using one's own data and estimating comorbidity score coefficients only in the control group was substantially worse when the sample size was small and the comorbidities were highly associated with treatment assignment (Table 3, $n=250$, coefficient set B, method 5).

When using 2 comorbidities rather than 4 in all steps above (Tables 5 and 6), the bias in the small sample size setting ($n=250$) was reduced.

In all cases, the concordance statistics were unaffected by the simulation design within coefficient sets. However, the concordance statistics were reduced when the association of comorbidities with treatment was strengthened (coefficient set B).

DISCUSSION

Our findings suggest why comorbidity summary measures have been so useful in health services research.

Foremost, our mathematical proof confirms the utility of comorbidity scores such as the Charlson Comorbidity

TABLE 2. Performance of Summary Measures Versus Indicator Variables Using SEER-Medicare Data

Model	C-Statistic
Base model only	0.615
Base model + individual Charlson indicator variables	0.667
Base model + Charlson score	0.664
Base model + individual Elixhauser indicator variables	0.672
Base model + Elixhauser score	0.652

The base model included age at diagnosis, year of diagnosis, diameter of the tumor, sex, race/ethnicity (5 categories: Hispanic or non-Hispanic black, white, Asian, other), and marital status (married/not married).

TABLE 3. Empirical Results From Simulations. All 4 Covariates Used, Comorbidity Slopes Derived Using Estimates From Data Example. The True Treatment Effect Coefficients Equal 2

Method	Treatment Effect	C-Statistic
Coefficient set A: Weak relationship of comorbidities with treatment (sample size = 250)		
1	2.05	0.83
2	1.96	0.83
3	1.94	0.83
4	1.96	0.83
5	1.93	0.83
Coefficient set B: Strong relationship of comorbidities with treatment (sample size = 250)		
1	2.05	0.78
2	1.86	0.77
3	1.83	0.77
4	1.88	0.77
5	1.84	0.77
Coefficient set A: Weak relationship of comorbidities with treatment (sample size = 2500)		
1	2.00	0.83
2	2.00	0.83
3	1.99	0.83
4	2.00	0.83
5	1.99	0.83
Coefficient set B: Strong relationship of comorbidities with treatment (sample size = 2500)		
1	2.01	0.77
2	1.99	0.77
3	1.99	0.77
4	1.99	0.77
5	1.99	0.77

The true point estimate is 2.0.

Index or Elixhauser score. From a theoretical perspective, once a researcher or physician knows a patient's comorbidity score, there may be no utility in knowing other information about the variables used to create the comorbidity score. Of note is that the proof assumes that we know the true comorbidity score. In practice, we only have an estimate of the truth, which could result in some small sample bias as our simulations demonstrate.

One caveat is that the utility of a comorbidity score is only as good as the variables that are used to create it. Other variables not used to create the comorbidity score might have a better prognostic ability than the summary score. Such variables include specific diseases, as well as more detailed information about disease severity for the comorbidities that are used to create the index. For example, the Charlson Comorbidity Index does not necessarily account for the effects of fine gradations of comorbidity severity that might be reflected in continuous variables. It is possible that continuous measures of diseases might outperform the Charlson score, which is based on binary (yes/no) measures.

In some respects, comorbidity summary measures are similar to propensity scores.^{20,21} Unlike propensity scores, comorbidity summary measures are derived from models of the survival outcome, rather than treatment. Further, in our mathematical proof, we only proved the general prognostic ability of comorbidity scores. Hansen wrote more generally about how comorbidity scores can be used appropriately to adjust for potential confounding in observational studies.¹⁹

TABLE 4. Empirical Results From Simulations. All 4 Covariates Used, Slopes Used to Create the Comorbidities Multiplied By 1.5, Creating Stronger Association of Comorbidities With Each Other (Sensitivity Analysis 1)

Method	Treatment Effect	C-Statistic
Coefficient set A: Weak relationship of comorbidities with treatment (sample size = 250)		
1	2.04	0.85
2	1.96	0.85
3	1.93	0.85
4	1.96	0.85
5	1.94	0.85
Coefficient set B: Strong relationship of comorbidities with treatment (sample size = 250)		
1	2.04	0.80
2	1.86	0.79
3	1.78	0.79
4	1.89	0.80
5	1.80	0.79
Coefficient set A: Weak relationship of comorbidities with treatment (sample size = 2500)		
1	2.00	0.85
2	2.00	0.85
3	1.99	0.85
4	1.99	0.85
5	1.99	0.85
Coefficient set B: Strong relationship of comorbidities with treatment (sample size = 2500)		
1	2.01	0.79
2	1.99	0.79
3	1.98	0.79
4	1.98	0.79
5	1.99	0.79

The true point estimate is 2.0.

TABLE 5. Empirical Results From Simulations. Same As Results in Table 3, But Only X_1 and X_2 Included in Simulations. Comorbidities X_3 and X_4 Not Included in Any of the Methods (Sensitivity Analysis 2)

Method	Treatment Effect	C-Statistic
Coefficient set A: Weak relationship of comorbidities with treatment (sample size = 250)		
1	2.03	0.70
2	2.00	0.77
3	2.00	0.77
4	2.00	0.77
5	2.00	0.77
Coefficient set B: Strong relationship of comorbidities with treatment (sample size = 250)		
1	2.03	0.70
2	1.96	0.69
3	1.98	0.69
4	1.96	0.70
5	2.00	0.69
Coefficient set A: Weak relationship of comorbidities with treatment (sample size = 2500)		
1	2.00	0.77
2	2.00	0.77
3	2.00	0.77
4	2.00	0.77
5	2.00	0.77
Coefficient set B: Strong relationship of comorbidities with treatment (sample size = 2500)		
1	2.00	0.69
2	2.00	0.69
3	2.00	0.69
4	1.99	0.69
5	2.00	0.69

The true point estimate is 2.0.

Another difference between the Charlson Comorbidity Index and propensity scores is that individuals generally estimate propensity scores using their own data, whereas the Charlson Comorbidity Index was first estimated in 1987. The coefficients from the first estimation of the Charlson Comorbidity Index were used in independent investigators' subsequent work. While the Charlson score sums exponentiated coefficients (ie, hazard ratios), this is a minor difference from methods that sum untransformed coefficients due to the limited numeric range of the Charlson weights.

Despite the fact that the Charlson Comorbidity Index was estimated some time ago, our data example using more modern SEER-Medicare data demonstrates that it can still be a robust summary measure for its component variables. Our Cox regression that used the Charlson Comorbidity Index had almost the same discriminative ability as a model that used the individual comorbidities as predictors. One caveat is that we only included those with localized kidney cancer, which is generally not very aggressive. The findings might differ when examining other types of cancer.

The Elixhauser score was not quite as good as using the individual Elixhauser predictors or the Charlson score. Although the individual Elixhauser comorbidities used in a regression gave slightly better results overall, there are more Elixhauser comorbidities than Charlson comorbidities. It is possible that a comorbidity measure that uses more information may have better discriminative ability than one

that uses less information. However, it is also possible that a comorbidity measure with more variables can lose more information than one with a smaller number of variables in finite sample sizes, as demonstrated by our simulations.

Our simulations suggest that the performance of comorbidity scores estimated using coefficients from an independent dataset may be adequately comparable to using coefficients from one's own data. This is again consistent with our SEER-Medicare analyses in which using the Charlson Comorbidity Index derived in 1987 had almost identical discriminative ability in a Cox regression as using the indicator variables. Contrary to the recommendation of Hansen,¹⁹ our simulations did not find that estimating comorbidity scores in control groups only, rather than using data from both arms, results in scores that are substantially better at reducing bias. We also found that comorbidity scores appropriately controlled for confounding with larger sample sizes regardless of the method used. The ability of comorbidity scores to adjust for confounding in larger sample sizes even when the score is estimated in a control group (methods 3 and 5 in Table 3) provides some reassurance about the use of the Charlson Comorbidity Index. For example, the Charlson score effectively acts as a comorbidity score estimated in an out-of-sample control group for investigators who use it.

There are limitations to the use of comorbidity scores published in the literature. If an investigator's patient sample

TABLE 6. Empirical Results From Simulations. Same As Results in Table 4, But Only X_1 and X_2 Included in Simulations. Comorbidities X_3 and X_4 Not Included in Any of the Methods. The Slopes Used to Create X_1 and X_2 Are Multiplied By 1.5 (This Combines Sensitivity Analyses 1 and 2)

Method	Treatment Effect	C-Statistic
Coefficient set A: Weak relationship of comorbidities with treatment (sample size = 250)		
1	2.02	0.78
2	2.00	0.78
3	1.99	0.78
4	1.99	0.78
5	2.00	0.78
Coefficient set B: Strong relationship of comorbidities with treatment (sample size = 250)		
1	2.03	0.70
2	1.96	0.70
3	1.98	0.70
4	1.97	0.70
5	1.98	0.70
Coefficient set A: Weak relationship of comorbidities with treatment (sample size = 2500)		
1	2.00	0.78
2	2.00	0.78
3	2.00	0.78
4	2.00	0.78
5	2.00	0.78
Coefficient set B: Strong relationship of comorbidities with treatment (sample size = 2500)		
1	2.01	0.70
2	2.00	0.70
3	2.00	0.70
4	2.00	0.70
5	2.00	0.70

The true point estimate is 2.0.

is very different from the sample used to construct the comorbidity summary measure, then the summary measure might perform less well than using the individual comorbidities as prognostic variables or confounding adjustment variables. An ad hoc approach to investigate whether a researcher's sample is too different from the population used to create a published comorbidity score would be to estimate the comorbidity score coefficients in the researcher's own sample and then examine whether the coefficients approximately match those published. In the case of the Charlson score,⁴ this can be done by using a Cox regression to re-estimate the comorbidity weights. If the findings suggest sufficient similarity in comorbidity weights, then using the Charlson score might make the results more generalizable.

Another limitation is that comorbidity summary measures might not effectively control for comorbidity confounding in observational treatment effectiveness studies if patients' probabilities of being treated (ie, the propensity score) are too close to 0 or 1 (related simulation findings not shown). In practical terms, this means that the Charlson or Elixhauser scores might not be useful if there are certain people included in a sample with characteristics that would make them ineligible to be assigned to 1 arm of a study. This is again consistent with Hansen's findings.¹⁹

Overall, summary comorbidity measures can be useful statistics for condensing comorbidity information into easy-

to-use metrics. Rather than trying to interpret the significance of multiple individual comorbidities, a summary measure can give clinicians and researchers a single number that captures the information. Our work suggests that once a comorbidity score is known, knowing details on the individual comorbidities used to create the index may give little additional information about a patient's prognosis. This may be one reason why summary measures such as the Charlson Comorbidity Index have been so useful over the decades.

ACKNOWLEDGMENT

We thank Dr Samuel Litwin for his comments.

APPENDIX

Mathematical Justification for Linear Regression

Let Y represent the continuous outcome, and $E[Y]$ be its expected (mean) value. Let X be a vector of covariates, $X = \{X_1, \dots, X_n\}'$. Also, x is a vector of the realized values, $x = \{x_1, \dots, x_n\}'$. For example, if there were 2 comorbidities of interest, such as diabetes or congestive heart failure, we might set $X_1 = 1$ if diabetes is present, 0 otherwise, and $X_2 = 1$ if congestive heart failure is present, 0 otherwise. A typical multiple linear regression would set the conditional expectation to: $E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$. In such a case, the outcome of interest, Y , might be blood pressure at a given time.

The comorbidity score is $E[Y|X]$, the linear regression of the outcome on the comorbidities. For ease of notation, define $b(X) = E[Y|X]$. To prove that the outcome is independent of the covariates given the comorbidity score (ie, heuristically a balancing property of a comorbidity score), it is necessary to show that: $E[Y|X, b(X)] = E[Y|b(X)]$.

Now, as $b(X)$ is a function of X , it follows that $E[Y|X, b(X)] = E[Y|X]$. The proof follows.

$$\begin{aligned}
 E[Y|b(X)] &= E[E[Y|X, b(X)]|b(X)], \\
 &\text{by law of iterated expectations.} \\
 &= E[E[Y|X]|b(X)], \text{ as } b(X) \text{ is a function of } X. \\
 &= E[b(X)|b(X)], \text{ by ease of notation} \\
 &\text{definition above.} \\
 &= b(X), \text{ as we condition on } b(X). \\
 &= E[Y|X].
 \end{aligned}$$

Hence, it is shown that:

$$\begin{aligned}
 E[Y|X, b(X)] &= E[Y|X] \\
 &= E[Y|b(X)].
 \end{aligned}$$

This demonstrates that a comorbidity score derived from a multiple linear regression when used in a subsequent multiple linear regression is similar to a balancing score such as the propensity score.^{20,21}

REFERENCES

- Lieffers JR, Baracos VE, Winget M, et al. A comparison of Charlson and Elixhauser comorbidity measures to predict colorectal cancer survival using administrative health data. *Cancer*. 2011;117:1957–1965.
- Lix LM, Quail J, Teare G, et al. Performance of comorbidity measures for predicting outcomes in population-based osteoporosis cohorts. *Osteoporos Int*. 2011;22:2633–2643.
- Perkins AJ, Kroenke K, Unttzer J, et al. Common comorbidity scales were similar in their ability to predict health care costs and mortality. *J Clin Epidemiol*. 2004;57:1040–1048.
- Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40:373–383.
- 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;46:1075–1079.
- Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992;45:613–619.
- Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Med Care*. 1998;36:8–27.
- van Walraven C, Austin PC, Jennings A, et al. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care*. 2009;47:626–633.
- Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43:1130–1139.
- de Groot V, Beckerman H, Lankhorst GJ, et al. How to measure comorbidity. A critical review of available methods. *J Clin Epidemiol*. 2003;56:221–229.
- D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *J Clin Epidemiol*. 1996;49:1429–1433.
- Nuttall M, van der Meulen J, Emberton M. Charlson scores based on ICD-10 administrative data were valid in assessing comorbidity in patients undergoing urological cancer surgery. *J Clin Epidemiol*. 2006;59:265–273.
- Katz JN, Chang LC, Sangha O, et al. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care*. 1996;34:73–84.
- Sangha O, Stucki G, Liang MH, et al. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum*. 2003;49:156–163.
- Klabunde CN, Potosky AL, Legler JM, et al. Development of a comorbidity index using physician claims data. *J Clin Epidemiol*. 2000;53:1258–1267.
- Sundararajan V, Quan H, Halfon P, et al. Cross-National comparative performance of three versions of the ICD-10 Charlson Index. *Med Care*. 2007;45:1210–1215.
- Ghali WA, Hall RE, Rosen AK, et al. Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. *J Clin Epidemiol*. 1996;49:273–278.
- Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol*. 2000;29:891–898.
- Hansen BB. The prognostic analogue of the propensity score. *Biometrika*. 2008;95:481–488.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for casual effects. *Biometrika*. 1983;70:41–55.
- Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc*. 1984;79:516–524.
- Sharabiani MTA, Aylin P, Bottle A. Systematic review of comorbidity indices for administrative data. *Medical Care*. 2012;50:1109–1118.
- National Cancer Institute, US National Institutes of Health: About the SEER-Medicare Database. Available at: <http://healthservices.cancer.gov/seermedicare/overview/>. Accessed August 13, 2012.
- Kutikov A, Egleston BL, Wong YN, et al. Evaluating overall survival and competing risks of death in patients with localized renal cell carcinoma using a comprehensive nomogram. *J Clin Oncol*. 2010;28:311–317.
- Elixhauser coding algorithm. Available at: http://www.hcup-us.ahrq.gov/tech_assist/faq.jsp. Accessed May 6, 2013.
- Harrell FE Jr. *Regression Modeling Strategies*. New York, NY: Springer; 2001;493.
- Foundation for Statistical Computing. *R [Computer Program]. Version 2.13.1*. Vienna, Austria: Foundation for Statistical Computing; 1998.