ELSEVIER

A Refined Comorbidity Measurement Algorithm for Claims-Based Studies of Breast, Prostate, Colorectal, and Lung Cancer Patients

CARRIE N. KLABUNDE, PHD, JULIE M. LEGLER, SCD, JOAN L. WARREN, PHD, LAURA-MAE BALDWIN, MD, MPH, AND DEBORAH SCHRAG, MD

PURPOSE: We evaluated (i) how combining comorbid conditions identified from Medicare inpatient or physician claims into a single comorbidity index compared with three other comorbidity indices and (ii) the need for comorbid condition weights that are specific to different cancer sites.
METHODS: This observational study used the SEER-Medicare linked database, from which four cohorts of cancer patients were derived: breast ($n = 26,377$), prostate ($n = 53,503$), colorectal ($n = 26,460$), and lung ($n = 33,975$). We calculated two established (Charlson; NCI) and two new (NCI Combined; Uniform Weights) comorbidity indices, and used Cox proportional hazards models to assess their predictive ability. We also used a pooled dataset to examine the inclusion of cancer site-specific condition weights.
RESULTS: The four comorbidity indices all significantly predicted mortality, but the NCI and new NCI Combined indices showed the greatest contribution to model fit. The new NCI Combined index is simpler to use and statistically more efficient than the NCI index. Modeling further demonstrated the utility of cancer site-specific weights.
CONCLUSIONS: Our results support the need for cancer site-specific comorbidity measures that employ empirically-derived condition weights. The new NCI Combined index is a refined, easier to implement comorbidity measurement algorithm appropriate for investigators using administrative claims databases to study four commonly-occurring cancers.
Ann Epidemiol 2007;17:584–590. © 2007 Elsevier Inc. All rights reserved.

KEY WORDS: Comorbidity, Data Sources, Medicare, SEER Program, Neoplasms, Health Services Research.

INTRODUCTION

The Charlson index (1) is a widely used measure of comorbidity, particularly among investigators conducting epidemiologic and outcomes research studies using administrative claims (2). The measure comprises 19 comorbid conditions, each assigned a weight according to its potential for influencing 1-year mortality. The index is the sum of the weighted comorbidities and accounts for the number and seriousness of the conditions. Originally developed to predict mortality among hospitalized patients, the index has been used to predict other patient outcomes, including treatment and costs. It has been adapted for use with administrative claims by mapping *International Classification of Diseases, revision* 9 (ICD-9) diagnostic and procedure codes to the conditions identified by Charlson as prognostically important (3, 4).

Since the publication of Charlson's original study, the proportion of medical care provided in the outpatient setting has increased considerably (5). Earlier studies that have used Medicare data generally have relied on codes obtained from inpatient claims to calculate a Charlson index (2). These analyses tend to classify a large proportion of cohort members as having no comorbid conditions simply because only a small proportion of the population is hospitalized in a given year.

An earlier study of comorbidity among Medicare beneficiaries with cancer used Medicare physician (Part B) claims and inpatient data to construct an adaptation of the Charlson index (6). This approach, hereafter referred to as the National Cancer Institute (NCI) index, involved calculation of two separate comorbidity indices from inpatient and physician claims. The analysis demonstrated improved prediction of noncancer mortality and treatment choice with the inclusion of physician claims for breast and prostate cancer patients. The authors obtained comorbid condition weights that differed from Charlson's (1) and showed that condition weights differed by cancer site. Although the NCI index (6) has been used in several claims-based analyses of cancer treatment and outcomes (7–14), it may be

From the Health Services and Economics Branch, Applied Research Program, National Cancer Institute, Bethesda, MD (C.N.K., J.L.W.); Department of Mathematics, Statistics, and Computer Science, Northfield, MN (J.M.L.); Department of Family Medicine, University of Washington School of Medicine, Seattle, WA (L.-M.B.); and Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY (D.S.).

Address correspondence to: Carrie N. Klabunde, PhD, Applied Research Program, National Cancer Institute, Executive Plaza North Room 4005, 6130 Executive Boulevard, Bethesda, MD 20892-7344. Tel.: 301-402-3362; fax: 301-435-3710. E-mail: ck97b@nih.gov.

Received October 25, 2006; accepted March 5, 2007.

© 2007 Elsevier Inc. All rights reserved.
360 Park Avenue South, New York, NY 10010

1047-2797/07/$–see front matter
doi:10.1016/j.annepidem.2007.03.011

Selected Abbreviations and Acronyms

ICD = *International Classification of Diseases*
NCI = National Cancer Institute
SEER = Surveillance, Epidemiology, and End Results program
AJCC = American Joint Committee on Cancer

somewhat cumbersome in practice because two separate comorbidity scores must be calculated and interpreted. Moreover, weights were derived only for breast and prostate cancer patients. Other studies (15–17) have also demonstrated the need for comorbidity indices tailored to the specific population of interest.

In this report, we describe a new analysis to improve the measurement of comorbidity in cancer patients using administrative claims databases. Our objectives are to evaluate (i) how combining comorbid conditions identified from Medicare inpatient or physician claims into a single comorbidity index, unlike the two separate indices used in the NCI index (6), compares with three other comorbidity indices and (ii) the need for comorbid condition weights specific to four different cancer sites. We chose to examine large cohorts of breast, prostate, colorectal, and lung cancer patients in our study because—excluding nonmelanoma skin cancer—these are the four most commonly diagnosed cancers in the United States.

METHODS

Data Sources

We used data from the NCI's Surveillance, Epidemiology, and End Results (SEER) program, and Medicare Part A hospitalization and Part B physician, supplier, and outpatient facility claims. The SEER and Medicare data have been linked as previously described (18, 19). For this study, we excluded Part B clinical laboratory, diagnostic imaging, and durable medical equipment claims, because the reporting of diagnoses on these claims is not required, and the diagnostic codes from these providers are less likely to be complete (20).

Study Sample

Medicare beneficiaries who were 66 years of age or older, not in health maintenance organizations, enrolled in both Parts A and B, and diagnosed with stages I-IV breast, colorectal, lung, or prostate cancer between January 1, 1992, and December 31, 1996, were eligible for inclusion (n = 153,342). We excluded patients whose cancer was diagnosed only on the death certificate or by autopsy (n = 7461). We also excluded colorectal and lung cancer patients with tumors of unusual histological subtype based on the ICD-O tumor classification system reported by the SEER registries (n = 5566) (21). We thus identified analytic cohorts of 26,377 female breast, 53,503 prostate, 26,460 colorectal, and 33,975 lung cancer patients.

Identifying Comorbid Conditions

ICD-9-CM diagnostic codes recorded in Medicare claims during the 12-month period before the patient's cancer diagnosis were searched for the comorbid conditions identified by Charlson et al. (1) as significantly influencing mortality. Conditions reported during the month of diagnosis were excluded to avoid misclassifying complications or conditions directly resulting from cancer diagnosis or treatment as comorbidities. We excluded the diagnostic codes corresponding to solid tumors and leukemia/lymphoma because of our focus on studying noncancer comorbidity in cancer patients. To ensure that diagnoses recorded in Part B claims were not transient episodes, we included only those conditions that appeared on more than one physician claim, as consistent with prior work (6, 22). The algorithm we used to screen for rule-out diagnoses has been shown to more accurately identify comorbid conditions in Medicare claims (23).

Defining Comorbidity Indices and Developing Condition Weights

We evaluated four comorbidity indices: (i) the Charlson index (1), (ii) the NCI index (6), (iii) the NCI Combined index, and (iv) the Uniform Weights index. All are based on the comorbid conditions identified by Charlson et al. (1) as important predictors of mortality. The NCI index (6) is composed of two weighted comorbidity scores derived separately from inpatient and physician claims. The new NCI Combined index uses weights derived from conditions identified in either the inpatient or physician claims and is reported as a single score. The new Uniform Weights index is composed of a count of the number of conditions that appear in either the inpatient or physician claims data.

To establish weights for comorbid conditions in the Charlson, NCI, and new NCI Combined indices, we used a randomly selected half of the dataset for each cancer site (i.e., "development" half). We included dichotomous indicator variables for comorbid conditions (1 if condition present; 0 if not) in a Cox proportional hazards model to predict 2-year mortality from causes other than cancer. Consistent with Charlson's methodology (1) and other studies of cancer and comorbidity (6, 15), we chose noncancer mortality as an outcome measure because of our focus on deaths attributable to comorbid conditions. Therefore, patients who died of cancer within the 2-year follow-up period or who were alive as of December 1998 were censored. Survival was measured as the number of months surviving during the follow-up period, which began with the month of diagnosis. In addition to comorbid conditions, patient age at diagnosis (continuous), cancer stage at diagnosis, and—for the lung and

colorectal cohorts—gender were included as explanatory variables in the models. We used a dichotomous measure of early versus late stage at diagnosis. For the breast and lung cancer cohorts, we used the SEER American Joint Committee on Cancer (AJCC) stage at diagnosis variable to define early stage disease as I-II and late stage as III-IV. We defined early stage colorectal cancer as I-III and late stage as stage IV. For the prostate cancer cohort, we used the SEER summary staging variable to define early stage as local/regional, and late stage as distant disease (24).

Calculating Comorbidity Indices

The estimated hazard ratios or coefficients obtained from the models described previoulsy were used as condition weights in constructing comorbidity indices for each patient. For the Charlson index, we retained only conditions with adjusted hazard ratios of 1.2 or greater and rounded these adjusted hazard ratios to the nearest integer. The integer weights were multiplied by their corresponding dichotomous condition indicators and summed to create the index (1).

For the NCI index, estimated coefficients for conditions were multiplied by their corresponding condition indicators and summed to create separate comorbidity scores for inpatient and physician claims (6). The new NCI Combined index was constructed by multiplying each condition indicator identified in inpatient or physician claims by its corresponding estimated coefficient from the Cox proportional hazards model and then summing over all conditions to create a single score (25). For the NCI and NCI Combined indices, we summed the estimated coefficients rather than exponentiated coefficients (i.e., hazard ratios), because the scale for comorbidity measures constructed in this way is consistent with the Cox proportional hazards and logistic regression models in which the measures are used. We calculated the new Uniform Weights index from conditions identified in either inpatient or physician claims. Each condition was assigned a weight of 1.0. The index is the sum of the uniformly-weighted conditions. Table 1 provides a summary of the analytic approach used to create each comorbidity index.

Assessing Comorbidity Indices—Predicting Mortality

To assess each index's predictive ability, we compared a series of models fit to the "test" half of each cohort's dataset (n = 13,108 breast, n = 26,737 prostate, n = 13,235 colorectal, n = 17,077 lung). We first estimated a base Cox proportional hazards model with age, stage, and gender as explanatory variables and 2-year noncancer mortality as the dependent variable. We then estimated models that included these variables plus the comorbidity index of interest. In all models, patients who died of cancer within the 2-year follow-up period or who were alive at the end of this period were treated as censored cases. Survival was measured in months. The validity of the proportional hazards assumption was assessed by plotting residuals and examining survival curves; the proportional odds assumption was met for the final models. We used the reduction in twice the log likelihood from the baseline models to compare nested models.

Assessing the Need for Cancer Site-Specific Condition Weights

In addition to comparing different comorbidity indices, we assessed the need for cancer site-specific condition weights

TABLE 1. Summary of analytic approach used to construct four comorbidity indices

Index name	Medicare data source	Method for assigning weights	Method for constructing index
Charlson (1)	Hospital claims	Obtain hazards ratio (HR) for each condition from survival estimation model. Retain conditions with HR > 1.2. Round HR values to the nearest integer, and multiply by the condition indicator.[a]	Sum the weighted conditions to establish a single index.
NCI (2)	Hospital and physician claims	Condition indicators are specific to each claims data source. Estimated coefficients for conditions identified in each claims source are obtained from survival estimation model. Coefficients are multiplied by condition indicators.[a]	Sum the weighted conditions to establish two separate indices: inpatient and physician claims.
NCI Combined	Hospital and physician claims	Conditions are identified from either claims data source. Estimated coefficients for conditions identified in either claims source are obtained from survival estimation model. Coefficients are multiplied by condition indicators.[a]	Sum the weighted conditions to establish a single index.
Uniform weights	Hospital and physician claims	Each condition found to be present in either claims source is assigned a value of 1.	Conditions are not differentially weighted. Sum to establish a single index.

[a]Condition indicators are dichotomous variables denoting the presence or absence of specific comorbid conditions. The condition indicator is set to 1 if the condition is present and to 0 if it is not.

by conducting a subsidiary analysis in which we first examined the pattern of comorbid conditions among the different cohorts by pooling the entire data set (all four cohorts; test and development halves) and fitting separate logistic regression models, with the individual comorbid conditions ascertained from either the inpatient or physician claims sources as dependent variables. Three binary indicators for cancer site (lung, prostate, colorectal), one indicator for gender (male = 1), and one indicator for stage (late = 1) were included as explanatory variables. The early stage breast cancer cohort thus served as the baseline. Interaction terms involving conditions and cancer site, gender, and stage were included and Wald-type tests performed to determine if the age-adjusted odds of individual conditions varied by cancer site, gender, and stage.

To further assess the need for cancer site-specific weights, we pooled the lung and colorectal cancer cohorts, and fit Cox proportional hazards models with 2-year noncancer mortality as the dependent variable and the individual comorbid conditions, cancer site indicators, and interaction terms as explanatory variables. We compared models with and without the condition by site interaction terms using twice the difference in the log-likelihoods. We used a similar approach to assess the need for gender- and stage-specific condition weights.

RESULTS

Description of Study Cohorts

Table 2 displays summary statistics for each cohort by stage at diagnosis, including the number of noncancer deaths and crude mortality rates, which are not age-adjusted. Stage distributions for the breast, prostate, and colorectal cancer cohorts are approximately 80% early and 20% late but are nearly reversed for the lung cancer cohort. The percent of patients with late-stage disease who died from cancer during the study period approached 85% for the lung and colorectal cohorts but was considerably lower for the breast and prostate cohorts. These differences in the lethality of cancers by site affect the amount of time patients are at risk for a noncancer death, the event of interest in this study. Because of the high proportion of lung cancer patients diagnosed with late stage disease, there is less time for them to be at risk for a noncancer death.

Comorbid Condition Prevalences

Table 3 shows the unadjusted prevalences for each comorbid condition by cohort. Chronic pulmonary disease, diabetes, and cerebrovascular disease are the most prevalent conditions, for all cohorts. For certain comorbid conditions, prevalences vary substantially by cancer site. For example, 7% of women with breast cancer and 15% of women with lung cancer have comorbid pulmonary disease. Table 3 also shows that physician claims are an important source of comorbidity information, particularly for conditions such as chronic pulmonary disease, diabetes, peripheral vascular disease, and rheumatologic disease, which generally require long-term medical monitoring. Because few cohort members had moderate/severe liver disease and none had AIDS, these conditions were excluded from the modeling.

Empirically Derived Weights

Table 4 displays results for the Cox proportional hazards models fit on the development half of each cohort's dataset. The coefficient estimates used as the empirically derived weights for the new NCI Combined index appear in the first column of results for each site. The corresponding hazard

TABLE 2. Cohort summaries[a]

Cancer site and stage	Number (%)	Males % of cohort	Females % of cohort	Number of noncancer deaths	Crude 2-year noncancer mortality per 1000 persons	% of cohort that died of cancer
Breast	26,377 (100)	0	100	1,733	65.7	14
Early[b]	23,094 (88)	0	100	1,414	61.2	9
Late[c]	3,283 (12)	0	100	319	97.2	53
Prostate	53,503 (100)	100	0	4,214	78.8	13
Early[b]	41,872 (78)	100	0	2,673	63.8	8
Late[c]	11,631 (22)	100	0	1,541	132.5	28
Colorectal	26,460 (100)	46	54	2,780	105.1	37
Early[b]	21,277 (80)	46	54	2,329	109.5	25
Late[c]	5,183 (20)	47	53	451	87.0	85
Lung	33,975 (100)	57	43	3,530	103.9	76
Early[b]	8,103 (24)	56	44	948	117.0	49
Late[c]	25,872 (76)	57	43	2,582	99.8	85

[a]Data source: SEER-Medicare linked database.
[b]Early stage is defined as SEER AJCC stage I-II for breast and lung cancer and stage I-III for colorectal cancer, and as SEER summary stage local and/or regional for the prostate cancer cohort.
[c]Late stage is defined as SEER AJCC stage III-IV for breast and lung cancer, IV for colorectal cancer, and as SEER summary stage distant for the prostate cancer cohort.

TABLE 3. Prevalence of comorbid conditions derived from Medicare inpatient or physician claims, and percentage of conditions identified in physician claims only, by cancer site and gender (ordered by decreasing overall prevalence)

	Breast(n = 26,377)		Prostate(n = 53,503)		Colorectal-female (n = 14,328)		Colorectal-male (n = 12,132)		Lung-female (n = 14,647)		Lung-male (n = 19,328)	
Condition	All claims (%)	% MD	All claims (%)	% MD	All claims (%)	% MD	All Claims	% MD	All claims (%)	% MD	All claims (%)	% MD
Chronic pulmonary disease	7.2	65	16.2	62	4.7	56	4.8	53	15.0	57	19.1	56
Diabetes	10.2	78	17.4	79	6.4	71	5.4	73	4.8	68	7.4	69
Congestive heart failure	5.7	55	9.8	53	5.1	49	3.6	49	5.2	45	6.7	43
Cerebrovascular disease	3.6	52	7.4	44	2.4	42	2.2	43	3.0	42	3.7	44
Peripheral vascular disease	2.1	74	4.6	61	1.5	67	1.5	59	2.3	60	3.8	55
Old myocardial infarction	0.8	20	2.9	29	0.5	22	1.0	22	0.7	13	1.5	22
Rheumatologic disease	1.5	79	1.7	80	0.9	72	0.4	78	1.6	69	1.0	69
Acute myocardial infarction	0.6	24	2.1	21	0.6	16	0.6	13	0.5	20	1.0	20
Moderate/severe renal disease	0.7	52	2.0	55	0.5	39	0.6	48	0.5	40	1.1	49
Diabetes with complications	1.0	64	1.5	69	0.6	62	0.5	63	0.5	56	0.7	69
Dementia	1.1	56	1.3	39	0.9	43	0.5	31	0.7	38	0.6	41
Ulcer disease	0.5	65	1.5	47	0.6	40	0.6	43	0.6	48	0.8	46
Paralysis	0.4	25	1.0	19	0.3	17	0.4	16	0.4	14	0.5	17
Mild liver disease	0.2	52	0.3	58	0.1	56	0.2	48	0.2	44	0.3	42
Moderate/severe liver disease	0.0	30	0.1	20	0.0	9	0.1	7	0.0	17	0.1	22
AIDS	0.0	100	0.0	0	0.0	0	0.0	100	0.0	0	0.0	0

MD = physician claims.

ratios from the new NCI Combined models are provided to facilitate comparison with empirically derived weights for the Charlson Index. Examination of the weights reveals differences in both ordering and relative magnitude by condition, cancer site, and comorbidity measure. Only the coefficients for the new NCI Combined index models are shown because, as described below, this index is generally preferred over the NCI index. Researchers who wish to use the NCI index because their analysis involves exclusively inpatient or exclusively outpatient claims files may obtain this measure's empirically derived coefficients from the corresponding author.

TABLE 4. Coefficient estimates and hazard ratios for comorbid conditions identified by the NCI Combined and Charlson methods, by cancer site, for the development half of the datasets (ordered by decreasing combined risk scores)

	Breast[a]			Prostate[a]			Colorectal[b]			Lung[b]		
	n = 13,247 women			n = 26,766 men			n = 13,186			n = 16, 829		
	841 noncancer deaths			2,122 noncancer deaths			1,385 noncancer deaths			1756 noncancer deaths		
	NCI Combined		Charlson	NCI Combined		Charlson	NCI Combined		Charlson	NCI Combined		Charlson
	Coef	HR	HR	Coef	HR	HR	Coef	HR	HR	Coef	HR	HR
Moderate/severe renal disease	1.188	3.28	3	0.678	1.97	2	0.966	2.63	3	0.804	2.23	2
Congestive heart failure	0.845	2.33	2	0.874	2.40	2	0.836	2.31	2	0.770	2.16	2
Dementia	1.192	3.29	3	0.777	2.17	2	0.596	1.82	2	0.650	1.92	2
Chronic pulmonary disease	0.471	1.60	2	0.725	2.06	2	0.471	1.60	2	0.334	1.40	1
Cerebrovascular disease	0.713	2.04	2	0.266	1.30		0.545	1.73	2	0.341	1.41	1
Paralysis	0.210	1.23	2	0.393	1.48	1	0.295	1.34		0.504	1.65	2
Diabetes	0.450	1.57	2	0.239	1.27	1	0.443	1.56	2	−0.008	0.99	
Diabetes with complications	0.023	1.02	1	0.440	1.55	2	0.302	1.35	2	0.407	1.50	2
Peripheral vascular disease	0.224	1.25		0.359	1.43	1	0.284	1.33	1	0.231	1.26	
Rheumatologic disease	0.751	2.12	3	0.091	1.09		−0.020	0.98		0.082	1.09	1
Acute myocardial infarction	0.201	1.22	1	0.242	1.27		−0.021	0.98		0.120	1.13	1
Old myocardial infarction	0.396	1.49	2	0.054	1.06		−0.185	0.83		−0.089	0.91	1
Ulcer disease	0.228	1.26	2	−0.247	0.78		0.080	1.08	2	−0.046	0.95	

HR = hazard ratio.
[a]Breast and prostate include age at diagnosis and late-stage indicator.
[b]Colorectal and lung models include age at diagnosis, male, and late-stage indicators.

Performance of Comorbidity Indices

Table 5 shows the reduction in twice the log likelihood from the baseline models obtained by including each comorbidity index. Larger values indicate better-fitting models. All models were highly significantly improved with the addition of any of the indices. But, for all four cancer sites, the models with the NCI and new NCI Combined indices showed the greatest reduction. Moreover, greater efficiency is achieved with the use of the new NCI Combined compared with the NCI index, because the NCI Combined index adds one parameter to the baseline models, whereas the NCI index requires adding two. Although the models that included the new Uniform Weights index also resulted in considerable improvements compared with the baseline model, the reduction is not as great as that achieved by the NCI or NCI Combined indices, in which empirically derived condition weights are applied. However, it is greater than that achieved by the Charlson Index, which may reflect the smaller number of conditions in the Charlson Index as the result of exclusion of conditions with adjusted hazard ratios of less than 1.2.

Need for Cancer Site-Specific Weights

Logistic regression results for the pooled dataset (not shown) demonstrated highly statistically significant differences in the age-adjusted odds of each comorbid condition by cancer site ($p < 0.0001$). There also was significant model improvement with the inclusion of site by condition indicators, demonstrating that age-adjusted odds ratios differ significantly by cancer site for each condition when controlling for cancer stage and gender. Models using the Combined lung and colorectal cohorts showed statistically significant differences in condition weights by cancer site ($p = 0.003$) and stage ($p = 0.005$), but not by gender ($p = 0.477$). We tested a three-way interaction term indicative of different weights for each site by stage, but it was not statistically significant ($p = 0.983$). On the basis of these results, we fit development models for condition weights separately by cancer site, as our analyses did not provide strong statistical evidence for either stage- or gender-specific weights.

TABLE 5. Reduction in –2 log likelihood[a] for Cox proportional hazards models for the test half of the data by cancer site, with 2-year noncancer mortality as the dependent variable

Variables in the Model[b]	Breast ($n = 13,073$)	Prostate ($n = 26,682$)	Colorectal ($n = 13,187$)	Lung ($n = 16,990$)
−2 log likelihood[a] for baseline models that include age at diagnosis, stage at diagnosis (1 = late), and gender (1 = male)	11269	26410	16147	19173
Charlson index	195	419	280	247
NCI index	338	646	394	323
NCI Combined index	338	661	406	324
Uniform Weights index	319	570	374	282

[a]Rounded to the nearest integer.
[b]All models include baseline covariates of age at diagnosis and late stage. Models for the lung and colorectal cancer cohorts also include an indicator for gender (male = 1). Each model in the table adds only one of the comorbidity indices indicated.

DISCUSSION

Researchers working with observational data require methods to appropriately adjust their analyses for underlying differences in patients' health status. The Charlson Index (1) is widely used for this purpose in health services and outcomes research studies, including those involving cancer patients (2). Charlson's work has been extended through the development of comorbidity measures that incorporate physician (Part B) claims in addition to inpatient data from Medicare files (6). This comorbidity measurement algorithm, the NCI index, requires constructing separate inpatient and outpatient comorbidity indices, an approach that some may find cumbersome. Furthermore, the NCI index is specific to breast and prostate cancer, and not other malignancies.

The present analysis expands upon this earlier work (6) by demonstrating the utility of a comorbidity index constructed from a single set of comorbid condition weights derived from the combined claims (i.e., inpatient or physician) of large cohorts of breast, prostate, colorectal, and lung cancer patients. Although our results showed comparable ability of the NCI and new NCI Combined indices to predict 2-year noncancer mortality, the new NCI Combined index is easier to implement than the prior algorithm (6). Furthermore, some efficiency is gained, both practically and statistically, with the new NCI Combined index because it introduces only one additional parameter into predictive models. We also provided statistical evidence of the need for cancer site-specific comorbid condition weights. Finally, we derived condition weights for the four most common malignancies in the adult U.S. population.

From a statistical perspective, there is theoretical justification for cancer site-specific comorbid condition weights. In the presence of multicollinearity among comorbid conditions, the values of the individual condition weights may be somewhat unstable. Patterns of multicollinearity may vary by cohort, contributing to condition weights that vary by cancer site. Moreover, comorbid conditions that vary in prevalence by cohort may induce different patterns of multicollinearity, also contributing to condition weights that vary by cancer site. Because patterns of multicollinearity and condition prevalence make it difficult to interpret coefficients individually, we believe that greater emphasis should

be placed on interpreting the magnitude and predictive value of the overall comorbidity index as opposed to comparing the individual condition weights.

Nevertheless, the philosophy and tradition of using complex algorithms for obtaining comorbid condition weights may require further examination. One recent study (26) comparing the ability of models with unweighted and weighted comorbid conditions to predict mortality and chemotherapy receipt showed them to have similar predictive power. In our study, models using the Uniform Weights index had significantly improved fit over models that did not include comorbidity. The improvements, however, were not as great as those achieved by the NCI or new NCI Combined index models, in which empirically derived weights were used. Because the empirically derived weights reported in our analysis provide better approximations to the true condition weights than do uniform weights, we recommend their use and have documented them in this report to facilitate their implementation by other investigators. We also recommend that users include the appropriate baseline covariates (age at diagnosis, stage, and gender where appropriate) when using the new NCI Combined comorbidity algorithm. Whether the effort involved in deriving condition weights for other disease-specific cohorts is warranted is a topic for future study.

In conclusion, in evaluating the new NCI Combined index, we documented a refined, easier-to-use comorbidity measurement algorithm that can be implemented by investigators working with administrative claims databases such as SEER-Medicare to study four common cancers. The statistical analyses we performed for cancer patients enhance currently available resources, and illustrate an approach that investigators using other data sources may wish to adopt.

The authors thank Nicki Schussler and Neil Rolfes of Information Management Services, Inc., Silver Spring, MD, for expert programming assistance.

REFERENCES

1. 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;40:373–383.
2. Klabunde CN, Warren JL, Legler JM. Assessment of comorbidity using claims data: an overview. Med Care. 2002;40(Suppl):26–35.
3. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol. 1992;45:613–619.
4. Romano PS, Roos LL, Luft HS, Jollis JG, Doliszny K. A comparison of administrative versus clinical data: coronary artery bypass surgery as an example. J Clin Epidemiol. 1994;47:249–260.
5. Bernstein AB, Hing E, Burt CW, Hall MJ. Trend data on medical encounters: tracking a moving target. Health Aff. 2001;20:58–72.
6. Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. J Clin Epidemiol. 2000;53:1258–1267.
7. Hershman DL, Wang X, McBride R, Jacobson JS, Grann VR, Neugut AI. Delay of adjuvant chemotherapy initiation following breast cancer surgery among elderly women. Breast Cancer Res Treat. 2006;99:313–321.
8. Doyle JJ, Neugut AI, Jacobson JS, Grann VR, Hershman DL. Chemotherapy and cardiotoxicity in older breast cancer patients: a population-based study. J Clin Oncol. 2005;23:8597–8605.
9. Haggstrom DA, Quale C, Smith-Bindman R. Differences in the quality of breast cancer care among vulnerable populations. Cancer. 2005;104:2347–2358.
10. Piccirillo JF, Spitznagel EL Jr, Vermani N, Costas I, Schnitzler M. Comparison of comorbidity indices for patients with head and neck cancer. Med Care. 2004;42:482–486.
11. Gold HT, Dick AW. Variations in treatment for ductal carcinoma in situ in elderly women. Med Care. 2004;42:267–275.
12. Shavers VL, Brown M, Klabunde CN, Potosky AL, Davis W, Moul J, et al. Race/ethnicity and the intensity of medical monitoring under 'watchful waiting' for prostate cancer. Med Care. 2004;42:239–250.
13. Shavers VL, Brown ML, Potosky AL, Klabunde CN, Davis WW, Moul JW, Fahey A. Race/ethnicity and the receipt of watchful waiting for the initial management of prostate cancer. J Gen Intern Med. 2004;19:146–155.
14. Zeliadt SB, Potosky AL, Etzioni R, Ramsey SD, Penson DF. Racial disparity in primary and adjuvant treatment for nonmetastatic prostate cancer: SEER-Medicare trends 1991 to 1999. Urology. 2004;64:1171–1176.
15. Fleming ST, Rastogi A, Dmitrienko A, Johnson KD. A comprehensive prognostic index to predict survival based on multiple comorbidities: a focus on breast cancer. Med Care. 1999;37:601–614.
16. Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. Int J Epidemiol. 2000;29:891–898.
17. Fleming ST, Pearce KA, McDavid K, Pavlov D. The development and validation of a comorbidity index for prostate cancer among Black men. J Clin Epidemiol. 2003;56:1064–1075.
18. Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG. Potential for cancer related health services research using a linked Medicare-tumor registry database. Med Care. 1993;31:732–748.
19. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Med Care. 2002:40 (Suppl): IV-3–IV-18
20. Centers for Medicare and Medicaid Services. Provider Reimbursement Manual. Available at: http://www.cms.hhs.gov/manuals/cmsindex.asp. Accessed April 10, 2007.
21. Fritz A, Percy C, Jack A, Sobin L, Parkin M. In: International Classification of Diseases for Oncology. 3rd ed. Geneva: World Health Organization; 2000.
22. National Cancer Institute, Health Services and Economics Branch. SEER-Medicare: Calculation of Comorbidity Weights. Available at: http://healthservices.cancer.gov/seermedicare/program/comorbidity.html. Accessed April 10, 2007.
23. Klabunde CN, Harlan LC, Warren JL. Data sources for measuring comorbidity: a comparison of hospital records and Medicare claims for cancer patients. Med Care. 2006;44:921–928.
24. Young JL. SEER *Summary Staging Manual 2000: Codes and Coding Instructions*. Bethesda, MD: NIH Publication No. 01-4969; 2000.
25. Fleming TR, Harrington DP. Counting Processes and Survival Analysis. New York: John Wiley & Sons; 1991.
26. Baldwin LM, Klabunde CN, Green P, Barlow W, Wright GE. In search of the perfect comorbidity measure for use with administrative claims data: does it exist? Med Care. 2006;44:745–753.