HHS Public Access

Author manuscript

Health Serv Outcomes Res Methodol. Author manuscript; available in PMC 2018 December 01.

Published in final edited form as:

Health Serv Outcomes Res Methodol. 2017 December ; 17(3-4): 237–255. doi:10.1007/s10742-017-0171-2.

On the Use of Summary Comorbidity Measures for Prognosis and Survival Treatment Effect Estimation

Elizabeth A. Gilbert,

Mathematics Department, Worcester State University

Robert T. Krafty,

Department of Biostatistics, University of Pittsburgh

Richard J. Bleicher, and

Department of Surgical Oncology, Fox Chase Cancer Center, Temple University Health System

Brian L. Egleston

Biostatistics and Bioinformatics Facility, Fox Chase Cancer Center, Temple University Health System, Philadelphia, Pennsylvania, 19111, U.S.A., Telephone: (215) 214-3917, Fax: (215) 728-2553

Abstract

Prognostic scores have been proposed as outcome based confounder adjustment scores akin to propensity scores. However, prognostic scores have not been widely used in the substantive literature. Instead, comorbidity scores, which are limited versions of prognostic scores, have been used extensively by clinical and health services researchers. A comorbidity is an existing disease an individual has in addition to a primary condition of interest, such as cancer. Comorbidity scores are used to reduce the dimension of a vector of comorbidity variables into a single scalar variable. Such scores are often added to regression models with other non-comorbidity variables such as age and sex, both for analyzing prognosis and for confounder adjustment when analyzing treatment effects. Despite their widespread use, the properties of and conditions under which comorbidity scores are valid dimension reduction tools in statistical models is largely unknown. In this article, we show that under relatively standard assumptions, comorbidity scores can have equal prognostic and confounder-adjustment abilities as the individual comorbidity variables, but that biases can occur if there are additional effects, such as interactions, of covariates beyond that captured by the comorbidity score. Simulations were performed to illustrate empirical properties and a data example using breast cancer data from the SEER Medicare Database demonstrates the application of these results.

Keywords

Comorbidity scores; Confounding adjustment; Prognostic scores; Summary measures

The corresponding author is Brian L. Egleston.

Compliance with ethical standards: Ethical approval: The Fox Chase Cancer Center Institutional Review Board has determined that this project meets the exempt status. Hence, this article does not contain any studies with human participants or animals performed by any of the authors.

1 Introduction

Prognostic scores have been proposed for the use of confounding adjustment in health services research, such as Hansen (2008) and DuGoff et al. (2014). However, searches of publication indices suggest that prognostic scores have not been widely incorporated into the substantive literature. Instead, comorbidity scores, which are diminutive versions of prognostic scores, have been used in tens of thousands of research papers. Comorbidity scores are summary variables of multiple health conditions, and do not contain information on non-health related variables such as age or sex. Two of the most widely used summary comorbidity measures are the Charlson Comorbidity Index, developed using hazard ratios by Charlson et al. (1987), and the Elixhauser Score, developed by Elixhauser et al. (1998), and later formed into a score by van Walraven et al. (2009). Despite the popularity of comorbidity scores in the substantive literature, the conditions under which they can be appropriately used in statistical models has not been extensively studied. In this paper, we establish the properties of summary comorbidity measures and provide thorough guidance for their use in prognosis and in confounder control.

The little that is known about the use of summary comorbidity measures within statistical models focuses on prognosis, or in the ability of summary comorbidity scores to provide as much information as the individual comorbidities when predicting the risk of a time-to-event survival outcome (Austin et al., 2015, Egleston et al., 2015). We add new insights into the use of summary comorbidity scores in prognosis by formally showing that summary comorbidity measures can provide valid dimension reduction tools both when formulated nonparametrically via a hazard and when formulated semiparametrically in proportional hazards models, which are widely used in practice. We also provide guidance on how best to estimate them. For example, while Hansen (2008) discussed the estimation of prognostic scores in general, questions about whether prognostic scores should be estimated from historical cohorts are still unanswered. We address some of these questions for comorbidity scores.

While demonstrating the validity of a summary comorbidity measure when used alone for prognostic purposes is important, often in practice, comorbidity scores are used in conjunction with other covariates to control for confounding when analyzing treatment effects. In addition to the previously mentioned insights into the use of summary comorbidity measures for prognosis, this article presents some of the first results on their use for confounder control when analyzing treatment effects. It is shown that, although summary comorbidity measures can provide valid dimension reduction tools for confounder control, care must be taken in the presence of additional contemporary effects of individual comorbidities or interactions with other covariates.

2 Examples

2.1 Charlson Comorbidity Index

To illustrate summary comorbidity scores, in this subsection we discuss one of the most popular. The Charlson Comorbidity Index is a summary measure that was first developed for prognosis and predicting risk of death from medical records data. However, due to its

widespread use and success, it has been adapted to health claims data, and is also being investigated for its use with questionnaire-based and physician claims-based data. Charlson et al. (1987) proposed a summary score that is based on over 20 medical conditions diagnosed from hospital data that have a wide range of severity. Each condition is represented by a score of 1 to 6, with 6 being the most severe conditions, such as a malignant tumor or AIDS, and 1 being less severe conditions, such as an ulcer or diabetes. If a condition is present in an individual, then they are given the score that is assigned to that condition. When all of an individual's scores are summed, the Charlson index has been calculated and can be used to predict mortality.

The Charlson Index is different from a prognostic score, which Hansen (2008) explains is a measure that summarizes covariates' association with potential outcomes. A prognostic score includes all confounders, while the Charlson Index is calculated using a select set of comorbidities weighted based on historical cohort estimates. The Charlson Comorbidity Index was first developed using ICD-9 (International Classification of Diseases) codes, but when the new ICD-10-CA codes were implemented, Quan et al. (2005) showed that this new administrative data performed as well, if not out-performed, the older ICD-9 codes.

For patients diagnosed with cancer, the Charlson Comorbidity Index is calculated by adding up 15 indicator variables related to non-cancer comorbidities weighted on their severity. Table 1 shows a breakdown of these variables and their corresponding weights.

2.2 SEER Medicare Data

To explore the use of comorbidity scores in practice, we will consider breast cancer data from the SEER Medicare Database. The Surveillance, Epidemiology, and End Results (SEER) Program is run by the National Cancer Institute (NCI) and reports on cancer incidence in the United States. The SEER program is a population-based cancer registry program that currently collects data in geographic areas with 28% of the United States population (<http://seer.cancer.gov/> as accessed on May 23, 2017). SEER collects patient demographic and cancer related information such as stage of cancer and survival time. Medicare data can be linked to SEER to identify Medicare based claims related to comorbidities. For this paper we identified early stage breast cancer cases diagnosed between 1992 and 2009. We only included individuals 66 years or older with non-HMO Medicare Parts A and B coverage for at least one year prior to diagnosis. To be considered a comorbidity, an individual generally had to have at least two claims for a relevant comorbidity at least 30 days apart. The requirement of having two claims was to ensure that “rule-out” diagnostic claims were not counted as comorbidities. In total, 155,106 patients are included in this dataset, approximately 99.0% female, and 61% of all of the patients were alive. We coded comorbidities using the algorithms of Deyo et al. (1992), Romano et al. (1993), and Klabunde et al. (2000) as implemented in an adapted program available at <https://healthcaredelivery.cancer.gov/seermedicare/considerations/calculation.html> (as accessed on May 23, 2017).

The Charlson Comorbidity Index works as a type of prognostic-balancing score for the set of comorbidities included in the formation of the Index (Eggleston et al., 2015). To illustrate this property, choosing an individual comorbidity, such as cerebrovascular disease (CVD), it can

be seen that the Charlson Index balances the SEER data as in Figure 1. In the overall data, patients who did not have cerebrovascular disease had a higher survival probability than those who did have cerebrovascular disease. However, when stratifying by the Charlson score, the survival probabilities are much more similar regardless of whether an individual had cerebrovascular disease or not.

3 Summary Comorbidity Measures for Prognosis

3.1 Nonparametric Time-Dependent Formulation

Many studies have been conducted to empirically examine if it is better to use a summary measure or the individual comorbidities to predict mortality, but different investigations have produced several different results. For example, Lieffers et al. (2011) used colorectal cancer data and found that a model created using the Charlson comorbidity index produced similar results as a model of the individual comorbidities, but Sundararajan et al. (2007) came to the opposite conclusion using diverse cross national datasets. This subsection summarizes the known methodological justification and properties of comorbidity scores and subsequent subsections present new results.

Let T be a random variable denoting survival time. Let $\mathbf{X}$ be a vector of comorbidities, $\mathbf{X} = \{X_1, X_2, \dots, X_n\}'$, and $\mathbf{x}$ be a vector of the realized values, $\mathbf{x} = \{x_1, \dots, x_n\}'$. Let $b_t^0(\mathbf{X})$ be a comorbidity score formed from a historical cohort, and let $h^0(\cdot)$, $f^0(\cdot)$, and $S^0(\cdot)$ be the hazard, density, and survival of this historical cohort. We use the superscript 0 to emphasize that these correspond to functions relevant to the historical cohort. Consider the time-dependent summary comorbidity measure defined nonparametrically by the hazard

$$b_t^0(\mathbf{X}) = h^0\{t|\mathbf{X}\} = \frac{f^0\{t|\mathbf{X}\}}{S^0\{t|\mathbf{X}\}}. \quad (1)$$

The following theorem, whose proof is given in Austin et al. (2015) and Eggleston et al. (2015), provides the validity of $b_t^0(\mathbf{X})$ as a summary measure for prognosis.

Theorem 1. Let $b_t^0(\mathbf{X})$ represent a comorbidity score defined from a historical control hazard as in (1). The comorbidity score, $b_t^0(\mathbf{X})$, is valid for the hazards model in that $h^0\{t|\mathbf{X}\} = h^0\{t|b_t^0(\mathbf{X})\}$.

3.2 Semiparametric Time-Invariant Formulation

Although the nonparametric formulation provides a very general tool and a valid time-dependent summary comorbidity score, often in practice a semiparametric proportional hazards model is used and a time-invariant score is desired. This is the case with the Charlson comorbidity index. The proportional hazards model is defined as

$$h^0\{t|\mathbf{X}\} = \lambda^0(t)e^{\beta' \mathbf{X}}$$

and the hazard ratio (HR) is defined as

$$e^{\beta' \mathbf{X}} = \frac{h^0\{t|\mathbf{X}\}}{\lambda^0(t)}.$$

This is the hazard ratio that compares the hazard associated with the sum of the comorbidity weights to the hazard when the score is set to zero; the hazard when the score is zero is simply the baseline hazard, $\lambda^0(t)$.

Austin et al. (2015) and Eggleston et al. (2015) suggested that it was applicable to use the hazard ratio as opposed to the hazard in the result for the validity of summary measures under the proportional hazards model and consider

$$b^0(\mathbf{X}) = \frac{h^0\{t|\mathbf{X}\}}{\lambda^0(t)}. \quad (2)$$

Note that under the proportional hazards model, the hazard ratio does not depend on t , providing a time-invariant score. The following corollary, whose proof is provided in online supplementary material, establishes the validity of $b^0(\mathbf{X})$ as a summary measure for prognosis. We use the corollary to more clearly map Theorem 1 to the hazard ratio, as many published comorbidity summary measures are derived from functions of hazard ratios.

Corollary 1. Let $b^0(\mathbf{X})$ represent a comorbidity score defined from a historical hazard ratio as in (2). The comorbidity score, $b^0(\mathbf{X})$, is valid for the proportional hazards model in that

$$\frac{h^0\{t|\mathbf{X}\}}{\lambda^0(t)} = \frac{h^0\{t|b^0(\mathbf{X})\}}{\lambda^0(t)}.$$

Since $\lambda^0(t)$ is the baseline hazard in both hazard ratios, this is equivalent to $h^0\{t|\mathbf{X}\} = h^0\{t|b^0(\mathbf{X})\}$.

The Charlson score was created by summing estimated hazard ratios (i.e. exponentiated Cox, 1972, model coefficients) rather than summing coefficients and then exponentiating the linear predictor, as the proof indicates would be more appropriate. While summing the exponentiated coefficients might be suboptimal, Charlson and colleagues had the fortuity of assigning points of 1, 2, 3, or 6 to conditions. In practice, hazard ratios of 1.2 to < 1.5 were rounded to 1, while the other scores were directly rounded. The rounded and limited range (i.e. no scores of 4 or 5) maintained the same approximate relative importance of the log

coefficients. For example, $\log(6)/\log(1.35)$ is approximately 6, which is the same as the 6/1 relative importance given by Charlson's rounded weights.

3.3 Historical Cohorts and Contemporary Studies

In practice, comorbidity scores are based on historical cohorts, but used to analyze data from contemporary studies. For example, the Charlson Comorbidity Index was proposed in 1987 but has been used perennially since then. Let $h(\cdot)$ represent the hazard from a contemporary group, such as the SEER breast cancer population. The following theorem and corollary, whose proofs are provided in supplementary material, establish the validity of time-dependent and time-invariant comorbidity scores based on historical cohorts in accounting for comorbidities in a contemporary group under the equivalence of conditional survival distributions.

Theorem 2. Let $b_t^0(\mathbf{X})$ represent a comorbidity score defined from a hazard for a historical population as in (1). If $h\{t|\mathbf{X}\} = h^0\{t|\mathbf{X}\}$, then the comorbidity score is valid for the contemporary hazards model in that $h\{t|\mathbf{X}\} = h\{t|b_t^0(\mathbf{X})\}$.

Corollary 2. Let $b^0(\mathbf{X})$ represent a comorbidity score defined from a hazard ratio for a historical population as in (2). If $h\{t|\mathbf{X}\} = h^0\{t|\mathbf{X}\}$, then the comorbidity score, $b^0(\mathbf{X})$, is valid in the contemporary cohort under the proportional hazards assumption in that

$$\frac{h\{t|\mathbf{X}\}}{\lambda(t)} = \frac{h\{t|b^0(\mathbf{X})\}}{\lambda(t)}. \quad (3)$$

While the assumption that $h\{t|\mathbf{X}\} = h^0\{t|\mathbf{X}\}$ might seem to be a strong assumption that justifies the use of historically derived comorbidity scores in contemporary research, this is an evaluable assumption. We will demonstrate in Section 6 that these are reasonable assumptions in some cases.

3.4 Estimation of $h\{t|\mathbf{X}\}$ in the presence of censoring

In this section, we considered the hazard with respect to the survival time T . In actual practice, many only observe the survival time in the presence of censoring, as not all participants of a study may have died by the end of the study observation period. Typically, those who estimate comorbidity scores, such as Charlson and colleagues, assume that censoring is independent of survival time given covariates, such as $\mathbf{X}$. Implicit in the estimation section below, we assume that censoring is similarly non-informative. Future work can elucidate the conditions under which dependent censoring can impact the utility of a comorbidity summary measure (e.g. Jackson et al., 2014).

4 Comorbidity Measures for Confounder Control in Treatment Effect Estimation

In clinical research, although predicating prognosis is important, understanding the effect of treatment is often the primary goal. Let us consider a treatment variable, Z , and a vector of additional covariates, $\mathbf{W}$, such as age, sex, and race, as are often included in statistical models. We use $\mathbf{W}$ for notation to distinguish the non-comorbidity covariates (i.e. $\mathbf{W}$) from the comorbidity covariates (i.e. $\mathbf{X}$). For the breast cancer data example to follow, $Z = 1$ indicates that the patient had treatment and $Z = 0$ indicates that the patient did not. To facilitate the formulation of a treatment effect, we utilize the potential outcomes framework, which was developed by Neyman (1923) and extended to non-randomized observational studies by Rubin (1974) and Holland (1986). Under the potential outcomes framework, there are two random variables for survival for each subject: T_0 is survival under control and T_1 is survival under treatment. Data will only provide information about the realization of one of these variables for a given subject, depending on what treatment is given. We let $h_0(\cdot)$, $f_0(\cdot)$, and $S_0(\cdot)$ and $h_1(\cdot)$, $f_1(\cdot)$, and $S_1(\cdot)$ represent the hazard, density, and survival of T_0 and T_1 , respectively for the contemporary population. Let $\{\cdot \perp \cdot \mid \cdot\}$ represent the independence of two variables conditional on a third variable.

The causal hazard ratio treatment effect is given by the ratio $h_1(t)/h_0(t)$ and if randomization occurs, this can be identified by $h_1(t|Z=1)/h_0(t|Z=0)$. Since we want to estimate the hazard ratio treatment effects conditional on the covariates, $\mathbf{X}$, and additional covariates, $\mathbf{W}$, we can use expectations as follows

$$\frac{E_{\mathbf{X}, \mathbf{W}}[h_1(t|Z=1, \mathbf{X}, \mathbf{W})]}{E_{\mathbf{X}, \mathbf{W}}[h_0(t|Z=0, \mathbf{X}, \mathbf{W})]},$$

where $h_1\{t|Z=1, \mathbf{X}, \mathbf{W}\}$ can be estimated using the treatment group and $h_0\{t|Z=0, \mathbf{X}, \mathbf{W}\}$ can be estimated using the control group.

We investigate the properties of summary comorbidity measures for analyzing treatment effects under assumptions. We make the commonly used Stable Unit Treatment Value Assumption (SUTVA) of causal inference, given by Rubin (1978). This states that the binary treatments are well defined (i.e. there are not multiple types of treatments grouped together), and that individuals' treatment decisions and outcomes are independent (i.e. no treatment interference among patients that may cause correlated outcomes). We further make a more formalized strong ignorability (Rosenbaum and Rubin, 1983) assumption as follows.

Assumption 1. $\{T_0, T_1\} \perp Z \mid \mathbf{X}, \mathbf{W}$.

Next, we assume that the conditional hazards can be partitioned as an underlying baseline hazard that only depends on the comorbidities $\mathbf{X}$, in addition to further proportional effects that depend on either $\mathbf{W}$ (the non-comorbidity variables) or $\mathbf{X}$.

Assumption 2. There exist positive functions α_0 , α_1 , ξ_0 , ξ_1 such that

$$h_j(t|\mathbf{X}, \mathbf{W}) = \alpha_j(\mathbf{W}) \xi_j(\mathbf{X}, \mathbf{W}) h^0(t|\mathbf{X}), j=0, 1.$$

In this parameterization, α_0 and α_1 represent the main effects of $\mathbf{W}$ and ξ_0 and ξ_1 account for additional contemporary main effects of $\mathbf{X}$ and possible interaction effects between $\mathbf{X}$ and $\mathbf{W}$ in the risk of death under control and treatment, respectively. We partition the additional effects of $\mathbf{X}$ and $\mathbf{W}$ into the α_j and ξ_j functions so that we can more easily examine identifiability in the presence of additional $\mathbf{W}$ -related main effects and $\mathbf{X}$ related main and interaction effects separately. Without loss of generality, to aid identifiability, it is assumed that ξ_j cannot be a function of $\mathbf{W}$ alone, or that $\xi_j(\mathbf{X}, \mathbf{W}) = \theta_j(\mathbf{W})$ for some function θ_j only if $\xi_j(\mathbf{X}, \mathbf{W}) = 1$ for all $\mathbf{X}, \mathbf{W}$.

Theorem 3. Let $b_t^0(\mathbf{X})$ represent a comorbidity score defined from a hazard for a historical population as in (1). Under Assumptions 1 and 2,

$$h_j(t|\mathbf{X}, \mathbf{W}) = h_j\{t|b_t^0(\mathbf{X}), \mathbf{W}\} B_j(t|\mathbf{X}, \mathbf{W}), j = 0, 1, \text{ where}$$

$$B_j(t|\mathbf{X}, \mathbf{W}) = \frac{\xi_j(\mathbf{X}, \mathbf{W})}{E\{\xi_j(\mathbf{X}, \mathbf{W})|b_t^0(\mathbf{X}), \mathbf{W}\}}$$

Proof.

$$\begin{aligned} h_j\{t|b_t^0(\mathbf{X}), \mathbf{W}\} &= \frac{f_j\{t|b_t^0(\mathbf{X}), \mathbf{W}\}}{S_j\{t|b_t^0(\mathbf{X}), \mathbf{W}\}} \\ &= \int \frac{f_j\{t|\mathbf{X}=\mathbf{x}, b_t^0(\mathbf{X}), \mathbf{W}\}}{S_j\{t|\mathbf{X}=\mathbf{x}, b_t^0(\mathbf{X}), \mathbf{W}\}} f\{\mathbf{x}|b_t^0(\mathbf{X}), \mathbf{W}\} d\mathbf{x} \\ &= \int \frac{f_j\{t|\mathbf{X}=\mathbf{x}, \mathbf{W}\}}{S_j\{t|\mathbf{X}=\mathbf{x}, \mathbf{W}\}} f\{\mathbf{x}|b_t^0(\mathbf{X}), \mathbf{W}\} d\mathbf{x} \\ &= \int h_j(t|\mathbf{X}=\mathbf{x}, \mathbf{W}) f\{\mathbf{x}|b_t^0(\mathbf{X}), \mathbf{W}\} d\mathbf{x} \\ &= \int \alpha_j(\mathbf{W}) \xi_j(\mathbf{X}=\mathbf{x}, \mathbf{W}) h^0(t|\mathbf{X}=\mathbf{x}) f\{\mathbf{x}|b_t^0(\mathbf{X}), \mathbf{W}\} d\mathbf{x} \\ &= E\{\alpha_j(\mathbf{W}) \xi_j(\mathbf{X}, \mathbf{W}) b_t^0(\mathbf{X}) | b_t^0(\mathbf{X}), \mathbf{W}\} \\ &= \alpha_j(\mathbf{W}) b_t^0(\mathbf{X}) E\{\xi_j(\mathbf{X}, \mathbf{W}) | b_t^0(\mathbf{X}), \mathbf{W}\} \\ &= \alpha_j(\mathbf{W}) h^0(t|\mathbf{X}) \{\xi_j(\mathbf{X}, \mathbf{W}) / \xi_j(\mathbf{X}, \mathbf{W})\} E\{\xi_j(\mathbf{X}, \mathbf{W}) | b_t^0(\mathbf{X}), \mathbf{W}\} \\ &= h_j(t|\mathbf{X}, \mathbf{W}) E\{\xi_j(\mathbf{X}, \mathbf{W}) | b_t^0(\mathbf{X}), \mathbf{W}\} / \xi_j(\mathbf{X}, \mathbf{W}). \end{aligned}$$

As with the properties for prognosis investigated in Section 3, results similar to those given in Theorem 3 for the nonparametric time-dependent scores exist for the semiparametric time-invariant scores under a proportional hazard model.

Corollary 3. Let $b^0(\mathbf{X})$ represent a comorbidity score defined by a hazard ratio as in (2). Under Assumptions 1 and 2,

$$\frac{h_j(t|\mathbf{X}, \mathbf{W})}{\lambda(t)} = \frac{h_j\{t|b^0(\mathbf{X}), \mathbf{W}\}}{\lambda(t)} B_j(\mathbf{X}, \mathbf{W}), j=0, 1, \quad (4)$$

$$\text{where } B_j(\mathbf{X}, \mathbf{W}) = \frac{\xi_j(\mathbf{X}, \mathbf{W})}{E\{\xi_j(\mathbf{X}, \mathbf{W}) | b^0(\mathbf{X}), \mathbf{W}\}}$$

Note that $B_j(\cdot)$ is not a function of t in Corollary 3. The results of Theorem 3 and its corollary point to two distinct settings when considering the use of summary comorbidity measures in analyzing treatment effects that depend on the presence of interactions. First, if $\xi_j(\mathbf{X}, \mathbf{W})$ is only a function of $\mathbf{X}$ through the historical comorbidity score, $b_t^0(\mathbf{X})$, then $B_j(t | \mathbf{X}, \mathbf{W}) = 1$, and the summary comorbidity measures are sufficient in estimating risk for both treatment and control in that $h_j(t | \mathbf{X}, \mathbf{W}) = h_j\{t | b_t^0(\mathbf{X}), \mathbf{W}\}$. It immediately follows that, in the absence of additional effects of $\mathbf{X}$ beyond the effect through $b_t^0(\mathbf{X})$, the historical comorbidity score is also valid in estimating the treatment effect where

$$\frac{h_1\{t | \mathbf{X}, \mathbf{W}\}}{h_0\{t | \mathbf{X}, \mathbf{W}\}} = \frac{h_1\{t | b_t^0(\mathbf{X}), \mathbf{W}\}}{h_0\{t | b_t^0(\mathbf{X}), \mathbf{W}\}}. \quad (5)$$

It should be noted that observed data can be used to estimate the treatment effect since, under Assumption 1, Equation (5) is equal to

$$\frac{h_1\{t | Z=1, \mathbf{X}, \mathbf{W}\}}{h_0\{t | Z=0, \mathbf{X}, \mathbf{W}\}} = \frac{h_1\{t | Z=1, b_t^0(\mathbf{X}), \mathbf{W}\}}{h_0\{t | Z=0, b_t^0(\mathbf{X}), \mathbf{W}\}},$$

where $h_1\{t | Z=1, \mathbf{X}, \mathbf{W}\}$ and $h_1\{t | Z=1, b_t^0(\mathbf{X}), \mathbf{W}\}$ can be estimated using the treatment group, while $h_0\{t | Z=0, \mathbf{X}, \mathbf{W}\}$ and $h_0\{t | Z=0, b_t^0(\mathbf{X}), \mathbf{W}\}$ can be estimated using the control group.

Next, consider the case where there is an interaction such that $\xi_j(\mathbf{X}, \mathbf{W}) \neq 1$ for some $\mathbf{X}, \mathbf{W}$. If the function only depends on $\mathbf{X}$ through $b_t^0(\mathbf{X})$, then $B_j(t | \mathbf{X}, \mathbf{W}) = 1$, and the summary comorbidity score is valid for the hazard and for the treatment effect. A simple example when W is one-dimensional is when $\xi_j(\mathbf{X}, \mathbf{W}) = \exp\{\tau b_t^0(\mathbf{X})W\}$ for some $\tau \neq 0$. If the function does not depend on the comorbidities only through the summary comorbidity score, then $B_j(t | \mathbf{X}, \mathbf{W}) \neq 1$ for some $\mathbf{X}$ and $\mathbf{W}$, and the use of summary comorbidity scores will result in biased estimates of the hazard. An example is when $b_t^0(\mathbf{X})$ is a function of all the comorbidities but $\xi_j(\mathbf{X}, \mathbf{W})$ is only a function of one of the comorbidities. However, if the additional effect of the comorbidities is the same under treatment and control such that $\xi_0(\mathbf{X}, \mathbf{W}) = \xi_1(\mathbf{X}, \mathbf{W})$, and more specifically, $B_1(t | \mathbf{X}, \mathbf{W}) / B_0(t | \mathbf{X}, \mathbf{W}) = 1$, then the bias terms negate one another so that Equation 5 still holds. Consequently, although estimates of the hazards are biased, summary comorbidity scores are valid in the estimation of the treatment effect.

5 Simulation Studies

5.1 Comorbidity Measures with Treatment Variable and Additional Co-variates

The first simulation study explores the use of a summary comorbidity measure with additional covariates, $\mathbf{W}$, as well as with an added treatment variable, Z . The baseline

results, without additional covariates, for this simulation study are given in Austin et al. (2015). Let X_1, X_2, X_3, X_4 be four comorbidities, W be an additional covariate, such as age, sex, or race, and Z be the treatment variable. The treatment variable, Z , is calculated with an inverse logit model where

$$\log \left[\frac{P(Z=1|\mathbf{X}, W)}{1 - P(Z=1|\mathbf{X}, W)} \right] = -0.25 + 0.25X_1 + 0.25X_2 + 0.25X_3 + 0.25X_4 + 0.25W$$

if there is a weak relationship between the covariates and their treatment assignment and

$$\log \left[\frac{P(Z=1|\mathbf{X}, W)}{1 - P(Z=1|\mathbf{X}, W)} \right] = -2 + 2X_1 + 2X_2 + 2X_3 + 2X_4 + 2W$$

if there is a strong relationship between the covariates and their treatment assignment. The covariates, X_1, X_2, X_3, X_4 , and the additional covariate, W , are distributed as

$$P[W=1] = 0.23$$

$$\text{logit}[P(X_1=1|W)] = -1.2 + 1.0W \quad (6)$$

$$\text{logit}[P(X_2=1|X_1, W)] = -1.9 + 1.1W + 1.4X_1 \quad (7)$$

$$\text{logit}[P(X_3=1|X_1, X_2, W)] = -1.6 + 0.4W + 1.0X_1 + 0.5X_2 \quad (8)$$

$$\text{logit}[P(X_4=1|X_1, X_2, X_3, W)] = -1.2 + 1.8W + 1.6X_1 + 0.2X_2 + 0.6X_3 \quad (9)$$

The hazard models with the individual covariates are given by

$$\frac{h_1\{t|\mathbf{X}, W\}}{\lambda_1(t)} = e^{1-2(\sum_{p=1}^4 X_p) - 2W}, \quad \frac{h_0\{t|\mathbf{X}, W\}}{\lambda_0(t)} = e^{-1-2(\sum_{p=1}^4 X_p) - 2W},$$

so that the log-treatment effect is 2. Note that this setting includes no interactions, or in the notation of Assumption 2, $\xi_j(\mathbf{X}, \mathbf{W}) = 1, j = 0, 1$. Survival times were generated using an exponential distribution,

$$f_T(t) = \left[\lambda(t) e^{2Z-1-2(\sum_{p=1}^4 X_p)-2W} \right] \exp \left[-\lambda(t) t e^{2Z-1-(\sum_{p=1}^4 X_p)-2W} \right].$$

Censoring times were generated using an exponential distribution with a rate of 0.3. Approximately 35% of the data was censored. For the simulations, we set $\lambda(t) = \lambda_0(t) = \lambda_1(t) = 1$.

We considered several settings under this framework and, for each setting, ran 500 simulations. For each simulation, we first simulated a “historical cohort”, which was used to estimate $b^0(\mathbf{X})$. The estimate, $\hat{b}^0(\mathbf{X})$ came from the linear predictor of a Cox regression that only used the comorbidities as covariates. We then simulated 250 “contemporary samples” and fit the Cox proportional hazard model

$$\frac{h\{t|\hat{b}^0(\mathbf{X}), W\}}{\lambda(t)} = e^{\delta Z + \beta \hat{b}^0(\mathbf{X}) + \omega W},$$

in which we used $\hat{b}^0(\mathbf{X})$ for $b^0(\mathbf{X})$, as is done, for example, when one uses a published comorbidity measure. Both the historical and contemporary samples were subject to censoring. The log-treatment effect (Log TE) was the estimate $\hat{\delta}$.

Simulations were run with varying sample sizes and coefficient sizes, as well as varying number of coefficients. The results are displayed in Table 2, where a ‘coefficient size’ of 1 represents unscaled covariates and a ‘coefficient size’ of 1.5 denotes that the coefficients, with the exception of the intercepts, in (6) through (9) are multiplied by a constant factor of 1.5 when developed. The category ‘number of coefficients’ is denoted as a 4 if all four covariates, $\mathbf{X} = \{X_1, X_2, X_3, X_4\}$, are used and as a 2 if only two of the covariates, $\mathbf{X} = \{X_1, X_2\}$, are used. The category ‘treatment set’ is denoted as ‘weak’ if there is a weak relationship of covariates with treatment and as ‘strong’ if there is a strong relationship of covariates with treatment. The ‘sample size’ used in these simulations is either 250 or 2500.

The table shows that the estimated log-treatment effects for each model are approximately 2, which is in agreement with the model of the individual covariates. The table also shows the concordance estimate. A pair is said to be concordant if the hazard of death is smaller for the patient that lived longer in the pair. Harrell et al. (1996) developed a C-Statistic that is the proportion of all patient pairs that have concordant outcomes. Here, we use the concordance probability estimate (CPE) developed by Gönen and Heller (2005) because it is an asymptotically unbiased estimator due to being a function of the regression parameters and the covariate distribution only. We see that the discriminatory power and predictive accuracy is high for each model at approximately 0.80.

5.2 The Effect of Interaction

This section illustrates the effect of interaction when interactions are different for each treatment. Five hundred simulations were run with varying sample sizes and coefficient

sizes, as well as varying number of coefficients, as described previously. The survival model for the individual covariates with different interactions for different treatment is given by

$$\frac{h_1\{t|\mathbf{X}, W\}}{\lambda_1(t)} = e^{1-2(\sum_{p=1}^4 X_p) - 2W - 8X_1W}, \quad \frac{h_0\{t|\mathbf{X}, W\}}{\lambda_0(t)} = e^{-1-2(\sum_{p=1}^4 X_p) - 2W + 8X_1W}$$

We estimated the comorbidity measure in our “historical” cohort using the linear predictor from a Cox regression that only included the four X_p variables. We then used the estimate of the comorbidity score for $b^0(\mathbf{X})$ in the model below which was estimated in the contemporary samples.

$$\frac{h_1\{t|b^0(\mathbf{X}), W\}}{\lambda(t)} = e^{\gamma Z + \beta b^0(\mathbf{X}) + \omega W + \tau b^0(\mathbf{X})W}$$

The treatment effect was estimated by $\hat{\gamma}$. The results can be seen in Table 3 for each combination of varying categories. The Concordance estimate is still high for each model at approximately 0.80. However, the table now shows that the estimated log treatment effect for each model are no longer approximately 2, which illustrates the bias incurred when using summary comorbidity scores in the presence of different interaction effects. Simulations that illustrate the lack of bias when interactions are the same within the treatment groups are presented in supplementary material.

6 Deciding whether to use a historically derived comorbidity score

Here, we demonstrate with data how one may evaluate the conditions set forth by the theorems and corollaries above. In this way, a researcher can decide whether a previously published or historically derived comorbidity score is appropriate for one's own endeavor. In this example, we will again use the linked SEER Medicare breast cancer database (see Section 2.2) where treatment is coded as $Z = 1$ if a person received chemotherapy treatment reimbursable by Medicare parts A or B and $Z = 0$ if they did not. For this example, let $\mathbf{W}$ be a vector of additional covariates which includes the categorical variables of sex, race/ethnicity, AJCC stage, grade, and the continuous variables of year of diagnosis, age, tumor size, and number of positive nodes. Survival time, T , was given in years and months. Sex is an indicator variable denoting male or female. Race was defined as denoting a patient as Hispanic or non-Hispanic White, Black, Asian, or other/unknown. Grade was defined using the SEER differentiation qualifiers of well, moderate, poor, undifferentiated, or unknown.

We will demonstrate the applicability of the Charlson Comorbidity Index in our SEER Medicare breast cancer sample. We will use proportional hazard models as they are commonly used by clinical investigators. Let $h\{t|b^0(\mathbf{X})\}$ represent the hazard in our SEER breast cancer population conditional on a comorbidity score derived from a historical control; note that this is a different population than the population used to develop the Charlson Comorbidity Index.

6.1 Evaluating Theorem 2 and Corollary 2

In Figure 1, we demonstrated how one could investigate the prognostic balance of the Charlson score using stratified Kaplan Meier curves. The balance achieved in the figures after conditioning on the comorbidity score provides preliminary evidence that

$h\{t|\mathbf{X}\} = h\{t|b^0(\mathbf{X})\}$, indicating that Theorem 2 may be applicable, at least for cerebrovascular disease. A more complete analysis would examine the balance of each comorbidity in $\mathbf{X}$.

Next, we examine the applicability of Corollary 2. We demonstrate the use of the Charlson Comorbidity Index under a Weibull proportional hazards model where the density of survival in the contemporary population is given by

$$f\{t|x\} = \lambda \gamma t^{\gamma-1} e^{\beta'x} e^{-\lambda t^\gamma e^{\beta'x}},$$

for scale parameter $\lambda > 0$, shape parameter $\gamma > 0$, and coefficient vector β . Note that for the Weibull model, contrary to the Cox proportional hazards model, the baseline hazard is not a function of the survival time. The parameter estimates for λ , γ , and β can be found via maximum likelihood estimation with respect to censoring.

We present the maximum likelihood estimates of the Weibull proportional hazards model using the linked SEER-Medicare breast cancer data in Table 4 (n=155,106). We see that scale (λ) and shape (γ) parameters are very similar when using either the individual comorbidities themselves or the Charlson comorbidity score. This provides evidence that the denominator baseline hazard for the hazard ratio in (3) of Corollary 2 is equivalent on both sides of the equation. Of note is that estimated hazard ratios in many cases are not consistent with the Charlson weights. For example, under Charlson and colleagues' rules for rounding, congestive heart failure (CHF) might more appropriately be given a score of 2 rather than 1. Quan et al. (2011) created updated weights to the Charlson score, which are more consistent with some of the estimated hazard ratios, but not with others. Also, AIDS seems not to affect mortality, but this may partly be due to the fact that AIDS is extremely rare, affecting only 14 patients in this older population of which 99.06% are female.

While we used a Weibull model above to investigate both the hazard ratios and the parametric baseline hazard function, Charlson and colleagues used a Cox model to estimate their index. When examining the Cox regression estimates of a model with the Charlson

Comorbidity Index, $\frac{h\{t|b^0(\mathbf{X})\}}{\lambda(t)}$, versus the 15 individual comorbidities themselves, $\frac{h\{t|\mathbf{X}\}}{\lambda(t)}$, Web Table 2 shows that the estimated hazard ratios are similar to those of the Weibull model

in Table 4. The hazard ratios (HRs), $\frac{h\{t|b^0(\mathbf{X})\}}{\lambda(t)}$ and $\frac{h\{t|\mathbf{X}\}}{\lambda(t)}$, along with a 95% confidence intervals (CIs) for each factor is displayed, along with the concordance probability, or predictive accuracy, of the nonlinear model. Web Table 2 shows that Gönen and Heller's

CPE is almost equal for the first model, $\frac{h\{t|\mathbf{X}\}}{\lambda(t)}$, and for the second model, $\frac{h\{t|b^0(\mathbf{X})\}}{\lambda(t)}$. Both models had a concordance probability estimate of approximately 0.57, which means 57% of patient pairs had concordant outcomes where the patient who was predicted to survive longer actually did survive longer. Despite differences in individual hazard ratio estimates from our data versus historical data, this suggests in aggregate that (3) of Theorem 2, Corollary 2 holds.

6.2 Evaluating Theorem 3, Corollary 3

In Table 5, we examine the group of older cases diagnosed with early stage breast cancer that was AJCC stages 2 or 3, estrogen-receptor status positive (ER positive), and who had either 1) surgery and radiation combined or 2) surgery and radiation in combination with chemotherapy (triple therapy). Chemotherapy during this time period had been shown to be less effective in ER positive patients (Giordano et al., 2006). However, since our selected cases had stage 2 or 3 disease (rather than stage 1), there may have been some motivation to try chemotherapy for the low probability that there might be an individual effect. The number of cases in the relevant subgroup was 19,157; the number with complete covariate data ($\mathbf{W}$) was 18,799. Of the 18,799, 56.63% had surgery and radiation without chemotherapy, 43.37% had triple therapy.

Not surprisingly given the ambiguities in effectiveness, the decision to give chemotherapy in this already substantially treated subgroup was related to comorbidities (odds ratio of giving chemotherapy for each unit increase in the Charlson score is 0.86, 95% CI 0.83 to 0.90). In the Cox regressions, we adjust for potential confounders ($\mathbf{W}$), with the continuous variables entered via restricted cubic splines with three knots (Harrell, 2001, pp. 11-40). No one in this sample had AIDS, so that comorbidity was not included in the analysis. Before adjustment with the Charlson score, we see that chemotherapy, when added to surgery and radiation, has a beneficial effect (HR=0.91, 95% CI 0.85 to 0.97). However, when either the Charlson score or the comorbidities were added, the beneficial effect is substantially attenuated and no longer statistically significant (HR=0.96 for both models, 95% CI=0.90 to 1.02 when the Charlson score is added, 95% CI=0.90 to 1.03 when the individual comorbidities are added). The hazard ratios for the chemotherapy effect were the same when using a Cox regression, and Gönen and Heller's CPE was 0.657 for the model without comorbidities, 0.671 for the model with the Charlson score, and 0.669 for the model with the individual comorbidity variables. The equivalent inferences for the treatment effect when comparing the model with either the Charlson score or the comorbidities suggests that (4) of Corollary 3 holds and that $B_f(\mathbf{X}, \mathbf{W}) = 1$. These findings support the concept that the Charlson score represents more than a simple summary measure, but instead is also a surrogate for the significant selection bias in choosing those who did not receive chemotherapy and were less likely to benefit from it, which confounded the perceived benefit of that therapy before adjustment.

6.3 In which arm should prognostic balance be assessed for treatment studies?

Hansen (2008) discussed the degree to which the development or evaluation of prognostic scores should incorporate information from the treatment group. Here, our presentation

might further help one decide. We have conceptualized published comorbidity scores akin to comorbidity scores estimated in historical control groups. However, under our assumptions and Theorem 3, the balance can be assessed under both arms separately using methods shown above. If $B_j(t|\mathbf{X}, \mathbf{W}) = 1$ from Section 4, then

$h_j\{t|Z=j, \mathbf{X}, \mathbf{W}\} = h_j\{t|Z=j, b_t^0(\mathbf{X}), \mathbf{W}\}$ and we have a no-effect modification case similar to that described by Hansen's Propositions 3 and 5. This would suggest that prognostic balance would hold within each treatment arm since the equality holds for both j equal to zero or one. We could demonstrate this balance visually using a figure similar to Figure 1 separately for each arm.

In the presence of strong additional effects of $\mathbf{W}$, one might want to use a regression separately in each arm in which hazard ratios are examined after further controlling for $\mathbf{W}$, akin to our regressions in Tables 4 and 5. In the presence of additional effects in which $B_j(t|\mathbf{X}, \mathbf{W}) \neq 1$, but the effects cancel each other out, then assessing balance might be a bit more imprecise, and further diagnostics might need to be developed for this case in the future. However, our findings above suggest that the diagnostics we proposed worked well for our surgically treated SEER-Medicare sample, in which we examined balance using the Charlson score. Combining arms might not be reasonable, because imbalance in comorbidity distributions between treatment arms might cause diagnostics, such as Figure 1 using the combined sample, to give the impression of prognostic imbalance. Such imbalance would be due to confounding by the uncontrolled treatment effect, a confounding effect that we would remove by examining the arms separately.

7 Concluding Remarks

Despite their widespread use as a dimension reduction tool in statistical models, the properties of summary comorbidity measures were widely unknown. We have demonstrated that they can indeed provide a valid dimension reduction tool, both for prognosis and for confounder control in the estimation of treatment effects, but that caution must be taken if the effect of any individual comorbidity on survival in the contemporary cohort differs from the historical controls used to develop the comorbidity score. As an example, our simulations demonstrated that interaction effects can invalidate the use of comorbidity scores. If such additional effects of the comorbidities in the contemporary sample are present and differ among the treatment groups, summary comorbidity scores from historical controls will lead to biased estimates and should be avoided. In such situations, one could formulate and use a prognostic score (Hansen, 2008). However, such scores combine the comorbidities and other covariates into a single measure, inhibiting one from evaluating their individual effects. Our data example provided a framework for how researchers might evaluate whether assumptions are applicable to a given dataset. We also demonstrated some diagnostic tools for evaluating whether it might be appropriate for a researcher to use a published comorbidity score for their own contemporary work. A topic of future substantive research will be the development of summary comorbidity scores that can provide a valid dimension reduction tool when interactions among comorbidities and other covariates exist, or when time dependent effects are present.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This was funded in part by NIH/NCI grants R03CA152388 & R21CA202130 (PI Egleston), P30CA006927 (Fox Chase Cancer Center Support Grant), and NIH/NIGMS grant R01GM113243 (PI Krafty).

Conflicts of Interest: There are no conflicts of interest related to this paper beyond the NIH grant funding support listed above.

References

- Austin S, Wong YN, Uzzo R, Beck J, Egleston B. Why summary comorbidity measures such as the Charlson Comorbidity Index and Elixhauser score work. *Medical Care*. 2015; 53:65–72. [PubMed: 25373404]
- Charlson M, Pompei P, Ales K, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*. 1987; 40:373–383. [PubMed: 3558716]
- Cox D. Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*. 1972; 34:187–220.
- Deyo R, Cherkin D, Ciol M. Adapting a clinical comorbidity index for use with icd-9-cm administrative databases. *Journal of Clinical Epidemiology*. 1992; 45:613–619. [PubMed: 1607900]
- DuGoff E, Schuler M, Stuart E. Generalizing observational study results: Applying propensity score methods to complex surveys. *Health Services Research*. 2014; 49:284–303. [PubMed: 23855598]
- Egleston B, Uzzo R, Beck J, Wong Y. A simple method for evaluating within sample prognostic balance achieved by published comorbidity summary measures. *Health Services Research*. 2015; 50:1179–1194. [PubMed: 25523400]
- Elixhauser A, Steiner C, Harris D, Coffey R. Comorbidity measures for use with administrative data. *Medical Care*. 1998; 36:8–27. [PubMed: 9431328]
- Giordano S, D Z, Kuo Y, Hortobagyi G, Goodwin J. Use and outcomes of adjuvant chemotherapy in older women with breast cancer. *Journal of Clinical Oncology*. 2006; 24:2750–256. [PubMed: 16782915]
- Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*. 2005; 92:965–970.
- Hansen B. The prognostic analogue of the propensity score. *Biometrika*. 2008; 95:481–488.
- Harrell F, Lee K, Mark D. Tutorial in biostatistics: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*. 1996; 15:361–387. [PubMed: 8668867]
- Harrell, FJ. *Regression Modeling Strategies*. Springer; New York: 2001.
- Holland PW. Statistics and causal inference. *Journal of the American Statistical Association*. 1986; 81:945–960.
- Jackson D, White I, Seaman S, Evans H, Baisley K, Carpenter J. Relaxing the independent censoring assumption in the cox proportional hazards model using multiple imputation. *Statistics in Medicine*. 2014; 33:4681–4694. [PubMed: 25060703]
- Klabunde C, Potosky A, Legler J. Development of a comorbidity index using physician claims data. *Journal of Clinical Epidemiology*. 2000; 53:1258–1267. [PubMed: 11146273]
- Lieffers J, Baracos V, Winget M, Fassbender K. A comparison of charlson and elixhauser comorbidity measures to predict colorectal cancer survival using administrative health data. *Cancer*. 2011; 117:1957–1965. [PubMed: 21509773]
- Neyman J. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes. Excerpts reprinted in English. *Statistical Science*. 1923; 5:463–472.

- Quan H, Sundararajan V, Halfon P. Coding algorithms for defining comorbidities in icd-9-cm and icd-10 administrative data. *Medical Care*. 2005; 43:1130–1139. [PubMed: 16224307]
- Romano P, Roos L, Jollis J. Adapting a clinical comorbidity index for use with icd-9-cm administrative data: differing perspectives. *Journal of Clinical Epidemiology*. 1993; 46:1075–1079. [PubMed: 8410092]
- Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effect. *Biometrika*. 1983; 70:41–55.
- Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*. 1974; 66:688–701.
- Rubin D. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*. 1978; 6:34–58.
- Sundararajan V, Quan H, Halfon P, Fushimi K, Luthi J, Burnand B, Ghali W. Cross-national comparative performance of three versions of the ICD-10 Charlson Index. *Medical Care*. 2007; 45:1210–1215. [PubMed: 18007172]
- van Walraven C, Austin P, Jennings A, Quan H, Forster A. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Medical Care*. 2009; 47:626–633. [PubMed: 19433995]

Figure 1. Charlson Index as a Balancing Score

Table 1
Comorbidities of the Charlson Comorbidity Index

Indicator	Comorbidity	Weight
1	Myocardial Infarct	1
2	Congestive Heart Failure (CHF)	1
3	Peripheral Vascular Disease (PVD)	1
4	Cerebrovascular Disease (CVD)	1
5	Chronic Obstructive Pulmonary Disease (COPD)	1
6	Dementia	1
7	Diabetes but not Diabetes with sequelae	1
8	Various Cirrhodites, but not Moderate-Severe Liver Disease	1
9	Ulcers	1
10	Rheumatologic disease	1
11	Paralysis	2
12	Diabetes with sequelae	2
13	Chronic Renal Failure	2
14	Moderate-Severe Liver Disease	3
15	AIDS	6

Table 2
Simulation Results for Different Treatments and Additional Covariates

Coef Size	Number of Coef	Treatment Set	Sample Size	Log TE _(SE)	Concordance
1	4	weak	250	1.958 _(.053)	0.861
1	4	strong	250	1.891 _(.143)	0.813
1	4	weak	2500	1.983 _(.012)	0.860
1	4	strong	2500	1.973 _(.030)	0.810
1.5	4	weak	250	1.962 _(.047)	0.862
1.5	4	strong	250	1.876 _(.130)	0.813
1.5	4	weak	2500	1.983 _(.012)	0.860
1.5	4	strong	2500	1.975 _(.028)	0.810
1	2	weak	250	2.005 _(.030)	0.807
1	2	strong	250	1.982 _(.084)	0.738
1	2	weak	2500	1.998 _(.006)	0.804
1	2	strong	2500	1.994 _(.016)	0.734
1.5	2	weak	250	2.003 _(.035)	0.807
1.5	2	strong	250	1.982 _(.091)	0.738
1.5	2	weak	2500	1.998 _(.008)	0.804
1.5	2	strong	2500	1.995 _(.021)	0.734

Table 3
Simulation Results for the Effect of Interaction with Different Interactions for Different Treatments

Coef Size	Number of Coef	Treatment Set	Sample Size	Log TE _(SE)	Concordance
1	4	weak	250	1.457 _(.054)	0.817
1	4	strong	250	1.764 _(.132)	0.812
1	4	weak	2500	1.482 _(.015)	0.816
1	4	strong	2500	1.851 _(.032)	0.810
1.5	4	weak	250	1.472 _(.056)	0.840
1.5	4	strong	250	1.844 _(.132)	0.838
1.5	4	weak	2500	1.494 _(.016)	0.839
1.5	4	strong	2500	1.935 _(.027)	0.836
1	2	weak	250	1.298 _(.041)	0.730
1	2	strong	250	1.387 _(.138)	0.710
1	2	weak	2500	1.296 _(.012)	0.728
1	2	strong	2500	1.414 _(.040)	0.708
1.5	2	weak	250	1.164 _(.038)	0.730
1.5	2	strong	250	1.449 _(.125)	0.732
1.5	2	weak	2500	1.168 _(.011)	0.728
1.5	2	strong	2500	1.479 _(.037)	0.730

Table 4
Parameter Estimates for the Weibull Distribution Data Example. HR=Hazard Ratio, SE=Standard Error

$f_0(t \mathbf{X})$					
Parameter	Charlson Weight	Quan Weight	Estimated HR	Estimated Coef _(SE)	p-value
λ				.036	
γ				1.24	
Myocardial Infarction β	1	0	1.07	.066 _(.028)	.003
CHF β	1	2	2.29	.827 _(.014)	<.001
PVD β	1	0	1.44	.366 _(.021)	<.001
CVD β	1	0	1.52	.422 _(.018)	<.001
COPD β	1	1	1.45	.372 _(.013)	<.001
Dementia β	1	2	3.15	1.146 _(.028)	<.001
Diabetes β	1	0	1.28	.247 _(.012)	<.001
Various Cirrhodites β	1	2	1.80	.588 _(.068)	<.001
Ulcers β	1	0	1.21	.188 _(.036)	<.001
Rheumatologic Disease β	1	1	1.22	.201 _(.026)	<.001
Paralysis β	2	2	1.26	.228 _(.050)	<.001
Diabetes with sequelae β	2	1	1.37	.315 _(.022)	<.001
Chronic Renal Failure β	2	1	1.94	.662 _(.028)	<.001
Liver Disease β	3	4	2.68	.988 _(.114)	<.001
AIDS β	6	4	1.4	.359 _(.010)	.343
$f_0(t \theta^0(\mathbf{X}))$					
λ				.037 _(.0003)	<.001
γ				1.23 _(.003)	<.001
Charlson Index β			1.44	.36 _(.003)	<.001

Table 5

Regressions with treatment (chemotherapy), additional covariates, and comorbidities. COPD=Chronic Obstructive Pulmonary Disease

	No Comorbidities $h(t Z,W)$			with Charlson $h(t Z,p^0(X),W)$			with Comorbidities $h(t Z,X,W)$		
	HR _(SE)	p-val.	HR _(SE)	p-val.	HR _(SE)	p-val.			
Chemotherapy	.91 _(.03)	.004	.96 _(.03)	.189	.96 _(.03)	.251			
Female	.63 _(.06)	<.001	.69 _(.07)	<.001	.69 _(.07)	<.001			
White	Reference		Reference		Reference				
Black	1.21 _(.06)	<.001	1.08 _(.06)	.143	1.09 _(.06)	.091			
Other/Unknown	.92 _(.11)	.464	.89 _(.11)	.351	.92 _(.11)	.472			
Asian	.79 _(.09)	.041	.72 _(.08)	.005	.73 _(.08)	.006			
Hispanic	.93 _(.13)	.574	.87 _(.12)	.299	.90 _(.12)	.437			
AJCC Stage 2	Reference		Reference		Reference				
AJCC Stage 3	1.13 _(.04)	.001	1.14 _(.04)	.001	1.13 _(.04)	.002			
Grade Well	Reference		Reference		Reference				
Grade Moderate	1.18 _(.05)	<.001	1.17 _(.05)	.001	1.17 _(.05)	.001			
Grade Poor	1.45 _(.07)	<.001	1.42 _(.07)	<.001	1.43 _(.07)	<.001			
Grade Undifferentiated	1.56 _(.16)	<.001	1.61 _(.17)	<.001	1.62 _(.17)	<.001			
Grade Unknown	1.17 _(.07)	.006	1.17 _(.07)	.009	1.16 _(.07)	.011			
Year Spline 1	.98 _(.01)	.001	.97 _(.01)	<.001	.97 _(.01)	<.001			
Year Spline 2	1.01 _(.01)	.569	1.01 _(.01)	.617	1.01 _(.01)	.618			
Age Spline 1	1.07 _(.01)	<.001	1.06 _(.01)	<.001	1.06 _(.01)	<.001			
Age Spline 2	1.01 _(.01)	.498	1.01 _(.01)	.139	1.01 _(.01)	.181			
Size Spline 1	1.05 _(.005)	<.001	1.05 _(.005)	<.001	1.05 _(.005)	<.001			
Size Spline 2	.94 _(.01)	<.001	.94 _(.01)	<.001	.94 _(.01)	<.001			
Positive Nodes Spline 1	1.14 _(.01)	<.001	1.13 _(.01)	<.001	1.13 _(.01)	<.001			
Positive Nodes Spline 2	.89 _(.01)	<.001	.90 _(.01)	<.001	.90 _(.01)	<.001			
Charlson Score			1.56 _(.05)	<.001					
Myocardial Infarction					1.38 _(.13)	.001			
Congestive Heart Failure					1.65 _(.09)	<.001			

	No Comorbidities $h\{t Z, W\}$		with Charlson $h\{t Z, p^0(X), W\}$		with Comorbidities $h\{t Z, X, W\}$	
	$HR_{(SE)}$	p-val.	$HR_{(SE)}$	p-val.	$HR_{(SE)}$	p-val.
Peripheral Vascular Disease					1.39 _(.11)	<.001
Cerebrovascular Disease					1.39 _(.1)	<.001
COPD					1.52 _(.07)	<.001
Dementia					2.52	<.001
Diabetes					1.27 _(.05)	<.001
Various Cirrhotides					1.93 _(.44)	.003
Ulcers					1.55 _(.19)	<.001
Rheumatologic Disease					1.36 _(.12)	.001
Paralysis					1.13 _(.27)	.609
Diabetes with Sequelae					1.68 _(.12)	<.001
Chronic Renal Failure					1.79 _(.18)	<.001
Mod/Severe Liver Disease					5.26 _(2.37)	<.001
$\log \lambda$	.0058 _(.0009)		.0048 _(.0007)		.0049 _(.0049)	
γ	1.51 _(.02)		1.53 _(.02)		1.54 _(0.02)	