

Semi-Competing Risks Data Analysis Accounting for Death as a Competing Risk When the Outcome of Interest Is Nonterminal

Sebastien Haneuse, PhD; Kyu Ha Lee, PhD

Abstract—Hospital readmission is a key marker of quality of health care. Notwithstanding its widespread use, however, it remains controversial in part because statistical methods used to analyze readmission, primarily logistic regression and related models, may not appropriately account for patients who die before experiencing a readmission event within the time frame of interest. Toward resolving this, we describe and illustrate the semi-competing risks framework, which refers to the general setting where scientific interest lies with some nonterminal event (eg, readmission), the occurrence of which is subject to a terminal event (eg, death). Although several statistical analysis methods have been proposed for semi-competing risks data, we describe in detail the use of illness–death models primarily because of their relation to well-known methods for survival analysis and the availability of software. We also describe and consider in detail several existing approaches that could, in principle, be used to analyze semi-competing risks data, including composite end point and competing risks analyses. Throughout we illustrate the ideas and methods using data on N=49 763 Medicare beneficiaries hospitalized between 2011 and 2013 with a principle discharge diagnosis of heart failure. (*Circ Cardiovasc Qual Outcomes*. 2016;9:322-331. DOI: 10.1161/CIRCOUTCOMES.115.001841.)

Key Words: death ■ heart failure ■ readmission ■ risk assessment ■ survival analysis

Hospital readmission is a key marker of quality of health-care; it has been used to investigate variation in quality among patients in a broad range of clinical contexts^{1–10} and has become an important policy measure.^{11,12} Notwithstanding its widespread use, however, readmission remains controversial as a measure of quality.^{13–16} Among the concerns raised, whether and how patient deaths are handled in the analyses and reported represents an important challenge. Joynt and Jha, for example, note that “Factoring a hospital’s mortality rate into its readmission-penalty calculation could ensure that the best institutions ... are not inappropriately penalized.”¹⁵

In the literature, analyses of readmission have generally been conducted with the use of logistic regression models for the binary outcome of readmission within 30 days.¹⁷ Such models are also used by Centers for Medicare and Medicaid Services in their calculation of risk-adjusted readmission rates.¹⁸ Practically, if a patient dies within the 30-day interval before experiencing a readmission event, they are counted as not having experienced the event throughout the interval. That is, person-time after death contributes to the analysis of readmission, even though a patient, by definition, cannot experience a readmission during this time. Arguably, a more appropriate approach would be to explicitly treat death as a competing force that is considered simultaneously with readmission. To this end, researchers could, in principle, use competing risks analysis.^{19–22} One key feature of these methods is that they are

specifically designed for settings where patients can experience one of the several so called terminal events, terminal in the sense that if a patient experiences an event of any given type, they cannot subsequently experience an event of any other type. Readmission, in contrast, is nonterminal, in the sense that the occurrence of a readmission event does not preclude the patient from death. Consequently, although a typical data set will contain some information on dependence between readmission and death, competing risks analyses do not make use of it.

As an alternative, this article considers embedding the analysis of readmission within the semi-competing risks framework.^{23,24} As we elaborate on, a key benefit of this framework is the explicit use of information on the timing of readmission events post death which provides a means to acknowledge and characterize dependence between the 2 events. In addition to describing semi-competing risks data and analysis, we describe a range of well-known methods that analysts may consider, including the use of a composite end point. As we emphasize throughout, key differences across the various methods we present lie in the interpretation of the model components as well as the extent to which information on the dependence between the 2 events is exploited. Although the ideas and methods are useful in a broad range of settings, we illustrate them using data on N=49 763 Medicare beneficiaries hospitalized between 2011 and 2013 with a principle discharge diagnosis of heart failure.

Received August 3, 2015; accepted February 24, 2016.

From the Department of Biostatistics, Harvard Chan School of Public Health, Boston, MA (S.H.); and Epidemiology and Biostatistics Core, The Forsyth Institute, Cambridge, MA (K.H.L.).

Correspondence to Sebastien Haneuse, PhD, Department of Biostatistics, Harvard Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115. E-mail shaneuse@hsph.harvard.edu

© 2016 American Heart Association, Inc.

Circ Cardiovasc Qual Outcomes is available at <http://circoutcomes.ahajournals.org>

DOI: 10.1161/CIRCOUTCOMES.115.001841

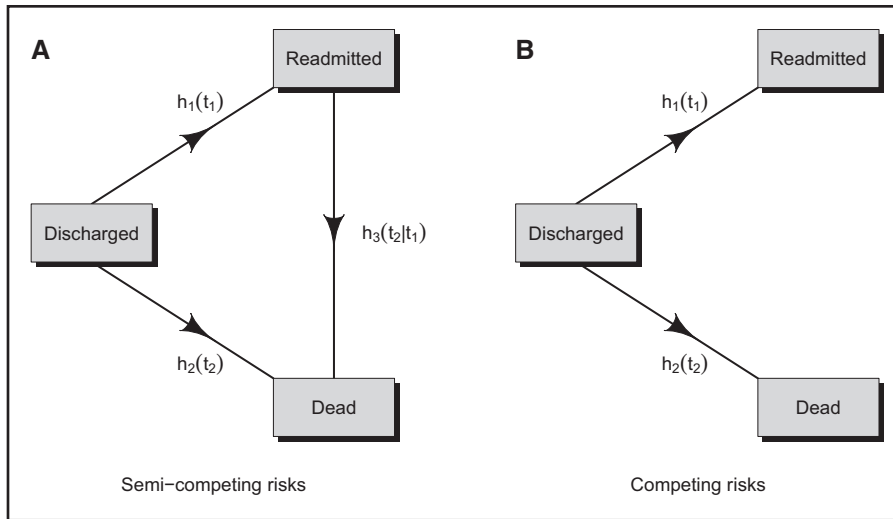


Figure 1. Graphical representation of the states that patients can transition between under the semi-competing risks (A) and competing risks (B) frameworks.

Semi-Competing Risks Data

General Structure

Semi-competing risks refers to the setting where primary interest lies with some nonterminal event, the occurrence of which is subject to a terminal event. Although we focus on hospital readmission as the nonterminal event, the notion is much more general and could refer to the incidence or recurrence of any health outcome, the discontinuation of treatment, or dropout. Typically, the terminal event is death, although it can also refer to other phenomena. For example, one could view treatment discontinuation as terminal if the nonterminal event is a treatment-related toxicity.

Intuitively, one can think of patients in the semi-competing risks setting as transitioning through a series of states. For example, in the Medicare data, we present that patients are initially in a discharged state (see Figure 1A). As time progresses, a patient could transition into the state of being readmitted or of being dead. If a patient transitions into the readmitted state, they could subsequently transition into the dead state, although these transitions cannot occur in the reverse order.

Medicare Data

Table 1 reports on $N=49\,763$ Medicare fee-for-service beneficiaries aged ≥ 65 who were hospitalized between 2011 and 2013 and subsequently discharged with a principle diagnosis of heart failure at one of the 141 hospitals in the 6 New England states (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont). Additional eligibility criteria for inclusion were that patients had at least 12 months of prior enrollment in Medicare Part A, had not transferred to some other acute care facility, and had been discharged to their home, an intensive care, or skilled nursing facility or to hospice care. Applying these criteria, some patients had >1 eligible hospitalization between 2011 and 2013; for the purposes of this article, we selected one hospitalization at random.

Table 1 also provides a summary of 30-day outcomes, overall and within levels of the factors reported. From the first row, we see 2.8% of patients experienced a readmission event and subsequently died within 30 days. We also see that 15.7% experienced a readmission event without subsequently dying within 30 days, so that the overall 30-day readmission rate was 18.5%. Similarly, combining the patients who experienced both events within 30

days with the 7.0% who died with 30 days without a readmission event, the overall 30-day mortality rate was 9.8%. Focusing on the 7.0% of patients who died within 30 days without a readmission event, Figure 2 provides a histogram of the death times for this group; nearly half of the 3504 patients (48.2%) died within 1 week and more than two thirds (69.6%) within 2 weeks.

Finally, we note that 74.4% of patients did not experience either event during the 30-day interval post discharge and were therefore censored. Most of these patients were censored administratively at 30 days, although a small percentage (2.2%) lost their fee-for-service eligibility before experiencing either event and before the 30-day mark; in the analyses presented below, person-time for these patients was censored accordingly.

Beyond the overall rates, Table 1 reveals substantial variation in the distribution of the 4 outcome types across levels of certain factors. We see, for example, that the rates at which patients are readmitted within 30 days post discharge decrease from 18.1% among patients aged 65 to 74 years to 12.0% among patients aged ≥ 95 years. Furthermore, this decrease is accompanied by a substantial increase in the rate of death without readmission (from 3.3% to 14.6%) and an increase in the rate of death with readmission (from 1.9% to 3.5%). In contrast, the rates at which patients are experiencing a readmission event, a death event, or both are all increasing as the length of stay for the initial hospitalization stay increases; correspondingly, the rate at which patients are surviving to 30 days without either event decreases dramatically from 79.1% among those whose stay was 1 to 3 days to 58.7% among those whose stay was ≥ 14 days. Finally, we find considerable variation in the distribution of the 4 outcome types across levels of discharge destination. Perhaps, most dramatic is the low readmission rate among patients discharged to a hospice facility (1.8% overall), although the rate at which patients are experiencing a death event without readmission is high (71.9%). Across the other 3 major discharge destinations, considerable variation exists in the rate at which patients die without a readmission event (1.5%–9.6%), as well as in the rate at which patients experience both events (1.2%–5.1%).

The Illness–Death Model

To formalize the structure of semi-competing risks data, let T_1 denote the time to the nonterminal event and T_2 the time to

Table 1. Patient Characteristics and 30-Day Outcomes for N=49 573 Medicare Beneficiaries Hospitalized in 1 of the 6 New England States Between 2011 and 2013 and Discharged With a Principle Diagnosis of Heart Failure

	Distribution		30-Day Outcomes, %			
	N	%	Readmission and Death	Readmission and Censored	Death Without Readmission	Censored
Total	49 763	100.0	2.8	15.7	7.0	74.4
Sex						
Female	27 775	55.8	2.7	15.5	6.9	74.8
Male	21 988	44.2	3.0	15.9	7.2	74.0
Age, y						
65–74	10 590	21.3	1.9	18.1	3.3	76.8
75–84	16 160	32.5	2.6	16.0	5.7	75.8
85–94	19 851	39.9	3.5	14.7	9.0	72.8
95+	3 162	6.4	3.5	12.0	14.6	69.8
Race						
White	46 412	93.3	2.9	15.6	7.3	74.2
African American	1 979	4.0	2.0	16.9	3.4	77.8
Other	1 372	2.8	2.3	16.8	4.5	76.3
Length of stay, days						
1–3	20 528	41.3	1.7	14.6	4.5	79.1
4–6	17 847	35.9	3.0	15.8	6.6	74.6
7–13	9 626	19.3	4.3	17.1	11.5	67.1
14+	1 762	3.5	6.2	19.1	15.9	58.7
Discharge location						
Home	13 899	27.9	1.2	14.4	1.5	82.8
Home with care	17 893	36.0	2.3	17.8	3.1	76.8
ICF/SNF	16 332	32.8	5.1	15.9	9.6	69.5
Hospice	1 639	3.3	0.6	1.2	71.9	26.3

ICF indicates intensive care facility; and SNF, skilled nursing facility.

the terminal event. The central challenge in analyzing semi-competing risks data is in developing an interpretable model for T_1 while simultaneously acknowledging potential dependence between T_1 and T_2 . In the statistical literature, several methods for the analysis of semi-competing risks data have been recently proposed.^{21,23–33} In this article, we focus on one of these approaches, the illness–death model. Although other methods are mentioned briefly in the Discussion, we focus on the illness–death model because it builds naturally and intuitively on well-known methods for survival analysis in standard settings and because software is readily available.

Toward formally describing the illness–death, consider again Figure 1A. Viewing T_1 and T_2 as time-to-event outcomes, one can specify the rates at which patients transition between the 3 states using 3 transition-specific hazard functions:

$$h_1(t_1) = \lim_{\Delta \rightarrow 0} P(t_1 \leq T_1 < t_1 + \Delta | T_1 \geq t_1, T_2 \geq t_1) / \Delta, t_1 > 0 \quad (1)$$

$$h_2(t_2) = \lim_{\Delta \rightarrow 0} P(t_2 \leq T_2 < t_2 + \Delta | T_1 \geq t_2, T_2 \geq t_2) / \Delta, t_2 > 0 \quad (2)$$

$$h_3(t_2 | t_1) = \frac{\lim_{\Delta \rightarrow 0} P(t_2 \leq T_2 < t_2 + \Delta | T_1 = t_1, T_2 \geq t_2)}{\Delta, 0 < t_1 < t_2} \quad (3)$$

The first of these is the hazard rate for readmission from discharge at a given point in time t_1 , given that neither the readmission nor the death event has occurred before time t_1 . The second is the corresponding hazard rate for death from discharge at a given time t_2 , given that neither the readmission nor the death event has occurred before time t_2 . Finally, $h_3(t_2 | t_1)$ is the hazard for death from readmission at time t_2 , given that a readmission event was observed at time $T_1 = t_1$ and that a death event had not occurred before time t_2 .

Collectively, the 3 hazard functions given by Equations 1–3 can be shown to define the joint distribution of (T_1, T_2) as consisting of 2 components.^{24,29} From Figure 3, the first component corresponds to instances where both events are experienced. That is, it indicates the distribution on the so called upper wedge on which $T_1 < T_2$ (ie, a readmission event is observed before a death event). The second component corresponds to instances where only a death event is experienced. Mathematically, because the induced joint distribution on the upper wedge does not integrate to 1.0, the balance of probability is attributed along the line $T_1 = \infty$. As discussed by Xu et al,²⁹ this strategy is, arguably, preferable to analysis strategies

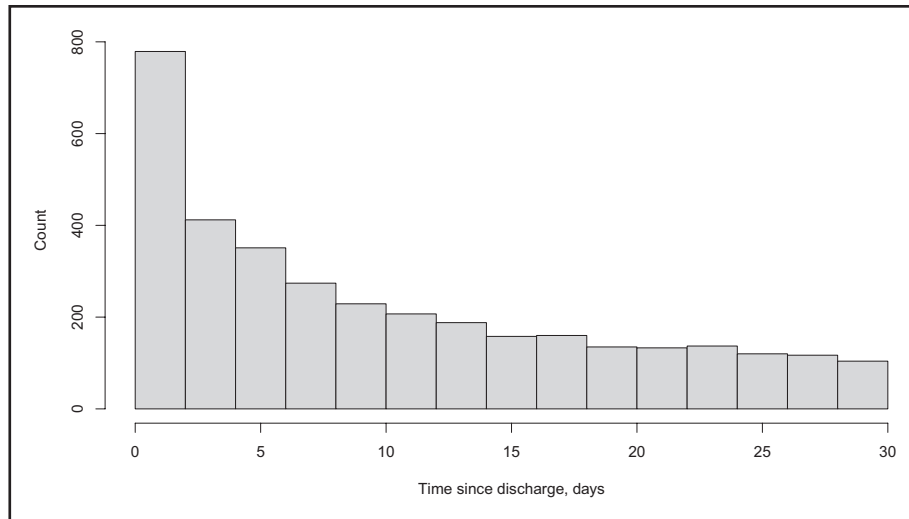


Figure 2. Distribution of the time to death for the 3504 patients who died within 30 days of being discharged and without experiencing a readmission event.

that assume some latent distribution for (T_1, T_2) on the lower wedge for which $T_1 > T_2$.

Although expressions 1–3 formally define the 3 hazard functions, as in standard survival analyses, one can adopt model structure for each to examine covariate effects and to obtain estimates of absolute risk. By far, the most common structure is the Cox model.³⁴ In the semi-competing risks context, the basic Cox model set up can be extended as follows:

$$h_1(t_1 | \gamma_i, \mathbf{X}_i) = \gamma_i h_{01}(t_1) \exp(\mathbf{X}_i^T \beta_1), t_1 > 0 \quad (4)$$

$$h_2(t_2 | \gamma_i, \mathbf{X}_i) = \gamma_i h_{02}(t_2) \exp(\mathbf{X}_i^T \beta_2), t_2 > 0 \quad (5)$$

$$h_3(t_2 | t_1, \gamma_i, \mathbf{X}_i) = \gamma_i h_{03}(t_2 | t_1) \exp(\mathbf{X}_i^T \beta_3), 0 < t_1 < t_2 \quad (6)$$

where γ_i is a patient-specific frailty and \mathbf{X}_i is a vector of patient-specific covariates.^{24,29} In expression 4, $h_{01}(t_1)$ denotes the baseline hazard function (ie, the hazard for a

population of patients with $\mathbf{X}=0$) for readmission from discharge, whereas β_1 denotes the vector of log hazard ratios (HRs) that characterize the effects of \mathbf{X} on the hazard for readmission from discharge. More specifically, the j th component of $\exp(\beta_1)$ can be interpreted as the HR for readmission corresponding the j th covariate in \mathbf{X} conditional on the fact that a death event has not occurred, holding the remaining elements of \mathbf{X} and γ_i fixed. Similarly, $h_{02}(t_2)$ in expression 5 denotes the baseline hazard function for death from discharge, whereas the components of $\exp(\beta_2)$ can be interpreted as HRs for death corresponding to the covariates in \mathbf{X} conditional on the fact that a readmission event has not occurred. Finally, $h_{03}(t_2 | t_1)$ in expression 6 corresponds to the conditional baseline hazard function for death, given that a readmission event occurred at time t_1 and $\exp(\beta_3)$ denotes the analogous HR parameters.

Additionally, each transition-specific model includes a shared patient-specific frailty, denoted γ_i . These frailties are analogous to random effects in mixed effects models for

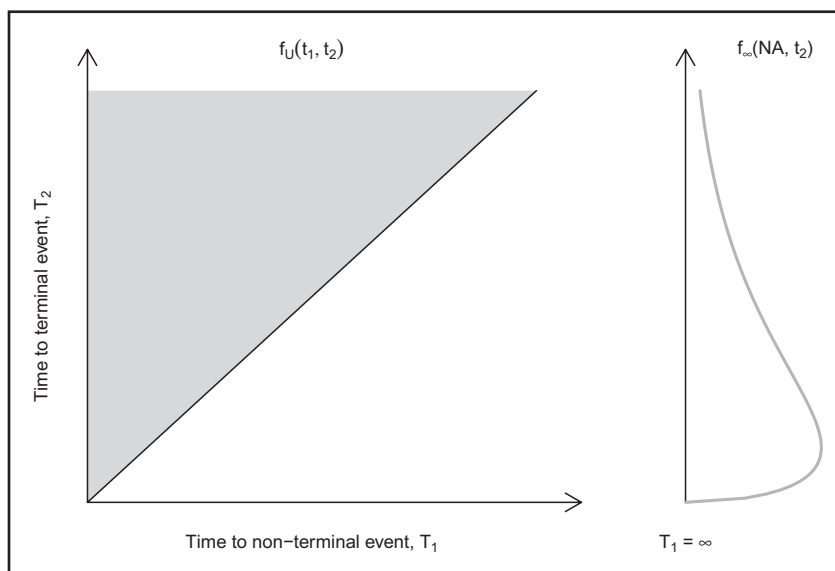


Figure 3. Graphical representation of the joint distribution of (T_1, T_2) based on the illness-death model.

longitudinal,³⁵ in that they serve as a device for accounting for residual correlation between T_1 and T_2 because of the collective impact of patient-specific covariates not included in \mathbf{X}_i . Although useful for the purposes of inference, it is important to note that their inclusion has an impact on the interpretation of the regression coefficients. Specifically, in addition, acknowledging the cause-specific nature of hazards 4–6, one should acknowledge that β_1 , β_2 , and β_3 are patient-specific and, therefore, distinct from their marginal counterparts. From a practical perspective, the patient-specific frailties are typically assumed to arise from a Gamma(θ^{-1} , θ^{-1}) distribution, parameterized so $E[\gamma]=1$ and $V[\gamma]=\theta$. As in standard mixed effects models, the variance of the frailties dictates the extent of induced correlation between T_1 and T_2 and can be estimated and reported along with the other components of the model.

Finally, analyses for the illness–death model can proceed either within the frequentist²⁹ or Bayesian paradigms,²⁴ allowing analysts to either adopt a parametric distribution for the 3 baseline hazard functions (eg, a Weibull) or specify them nonparametrically. As a function of 2 continuous variables, $h_{03}(t_2|t_1)$ has the potential to be complex and therefore difficult to specify parametrically or estimate nonparametrically. In practice, analysts resolve this by making 1 of 2 simplifying assumptions. In the first, $h_{03}(t_2|t_1)$ is assumed to be independent of t_1 so that expression 6 becomes:

$$h_3(t_2|t_1, \gamma_i, \mathbf{X}_i) = \gamma_i h_{03}(t_2) \exp(\mathbf{X}_i^T \beta_3), 0 < t_2 \quad (7)$$

This representation is often referred to as the Markov model because it does not depend on the timing of the readmission event. Intuitively, one can interpret this model as stating that a patient's risk for death at any given time post discharge, say t_2 , is initially governed by the hazard given in expression 5; if and when they experience a readmission, their risk for death is subsequently governed by the hazard given in expression 7. An alternative to the Markov assumption is one where expression 6 is replaced by

$$h_3(t_2|t_1, \gamma_i, \mathbf{X}_i) = \gamma_i h_{03}(t_2 - t_1) \exp(\mathbf{X}_i^T \beta_3), 0 < t_1 < t_2 \quad (8)$$

so that the risk of death at a given time post discharge for a patient who has experienced a readmission event at time t_1 depends on the time since readmission (ie, $t_2 - t_1$, the so called sojourn time), rather than the time since discharge. This representation is referred to as the Semi-Markov model. Generally, when choosing between these specifications, the primary consideration will be the extent to which model 7 or model 8 better characterizes risk post readmission from either the perspective of scientific plausibility or in terms of goodness-of-fit, or possibly both.

Alternative Analysis Strategies

To clarify the importance of specific features of the illness–death model, it is instructive to consider alternative modeling strategies that analysts will generally be familiar with and may use to analyze semi-competing risks data.

Univariate Logistic Regression

As mentioned in the Introduction, analyses of readmission are typically conducted using logistic regression models. To formalize this, let $Y_i=0/1$ indicate whether the i th patient was readmitted within 30 days post discharge. The logistic regression model relating \mathbf{X}_i to Y_i is given by:

$$\text{logit } P(Y_i = 1 | \mathbf{X}_i) = \mathbf{X}_i^T \alpha \quad (9)$$

In the presence of death as a competing risk, however, the interpretation of results based on this model is complicated by the fact that it specifies a patient's risk as being linear on the logit scale throughout the 30-day window, even though such an assumption is not well-defined if a patient has died before the 30-day mark. Consequently, even though one can fit the model (in the sense of running some computer code), the estimates correspond to some mixture of the effect of \mathbf{X} for patients who were at risk throughout and the effect of \mathbf{X} for patients who could not have experienced the event during portions of person-time.

Univariate Survival Analysis

To explicitly acknowledge when a patient is at risk to experience the nonterminal event during the 30-day interval, one could use standard survival analysis methods. In terms of regression analyses, one way forward would be to build a model for the marginal hazard of readmission:

$$h^*(t_1) = \lim_{\Delta \rightarrow 0} P(t_1 \leq T_1 < t_1 + \Delta | T_1 \geq t_1) / \Delta, t_1 > 0 \quad (10)$$

Note, we refer to this hazard as marginal in the sense that it does not condition on death as a competing risk and is therefore distinct from the hazard given by expression 1. Typically, researchers model marginal hazard functions with a univariate Cox model, such as

$$h^*(t_1 | \mathbf{X}_i) = h_0^*(t_1) \exp(\mathbf{X}_i^T \beta^*) \quad (11)$$

In addition to not conditioning on death as a competing risk, we note that, in contrast to expression 4 for the illness–death model, the specific choice of expression 11 does not include a patient-specific frailty. Hence, although β^* and β_1 both speak to the association between \mathbf{X} and readmission, their precise interpretations differ, and they may or may not be numerically the same.

Operationally, estimation for model 11 typically proceeds on the basis of the usual partial likelihood treating death as a censoring mechanism. An important drawback of this approach, however, is that if there is dependence between readmission and death (as noted in Table 1), then the resulting estimate will not be guaranteed to be consistent for β^* , although assuming proportional hazards holds, it will be consistent for its cause-specific counterpart (see later). In this sense, caution is needed to ensure an appropriate interpretation of the results.

Composite End Points

Instead of treating death as a censoring mechanism, analysts may consider incorporating it as a component of the outcome by defining and analyzing a composite end point.³⁶ Toward this, let $T_c = \min(T_1, T_2)$ denote the time from discharge to the

first of readmission or death. One could perform a standard survival analysis for T_c by fitting the model:

$$h(t_c | \mathbf{X}_i) = h_{0c}(t_c) \exp(\mathbf{X}_i^T \beta_c) \quad (12)$$

Assuming that the remaining forms of censoring are independent, estimation and inference for β_c could proceed as usual. The main drawback of this approach, however, is that the interpretation of β_c requires viewing it as a mixture of the effects of \mathbf{X} on both readmission and death, so that scientific attention is shifted away from readmission as the primary outcome. Consequently, in attempting to resolve a statistical issue, the primary scientific questions have been altered.

Competing Risks Analysis

Finally, analysts could use the competing risks framework, specifically building models for the 2 hazards of the subdistribution or 2 cause-specific hazard functions.^{19–22,37} Focusing on the cause-specific hazard functions, an analysis could proceed by fitting the following 2 models:

$$h_1(t_1 | \mathbf{X}_i) = h_{01}^*(t_1) \exp(\mathbf{X}_i^T \beta_1^*), t_1 > 0 \quad (13)$$

$$h_2(t_2 | \mathbf{X}_i) = h_{02}^*(t_2) \exp(\mathbf{X}_i^T \beta_2^*), t_2 > 0 \quad (14)$$

for the 2 cause-specific hazards given by expressions 1 and 2. As with β^* in expression 11, the superscript $*$ has been added to emphasize the distinction between the components in Equations 13 and 14 from the corresponding components in Equations 4 and 5. For example, although both β_1 and β_1^* are cause-specific parameters, the interpretation of β_1^* also requires acknowledging the fact that the model conditions on the patient-specific frailties. As such, β_1 and β_1^* may not be equivalent numerically. Nevertheless, one could proceed with estimation and inference for either cause-specific hazard function by using the standard partial likelihood treating the other event as a censoring mechanism. The primary drawback of applying the competing risks framework to the study of readmission, however, is that postreadmission death events and their timing are ignored (see Figure 1B). That is, information that could be used to inform an understanding of dependence between readmission and death, and hence joint distribution of the 2 events, is essentially thrown away.

Results

Here we present results from a series of analyses that apply the methods discussed earlier to N=49763 Medicare beneficiaries. For all analyses, we considered the factors listed in Table 1, together with an indicator of the New England state in which the hospitalization occurred. For simplicity, we take the baseline hazard functions for all survival models to be structured according to a Weibull distribution, although note that current methods and software permit a more flexible specification.²⁴ Furthermore, we restrict attention to results for the illness–death model based on the Markov specification for $h_3(t_2 | t_1)$, given by expression 7. Finally, all of the analyses

were conducted using the freely available SemiCompRisks package for R.³⁸

Focusing on the univariate analyses, Table 2 shows that the general conclusions based on the logistic regression analysis and the survival analysis with readmission as the primary outcome are very similar (see the first 2 sets of columns). Both analyses, for example, indicate small-to-moderate associations for sex and age and a strong association for length of stay. In regard to discharge destination, although there is little evidence of a meaningful difference in risk for readmission between patients discharged to home without care and those discharged to either a intensive care facility or skilled nursing facility (eg, HR 1.05; 95% confidence interval 1.00–1.10), patients discharged to their home without care have significantly lower risk of readmission than those with care (eg, HR 0.77; 95% confidence interval 0.72–0.81). This is, arguably, because of unadjusted case-mix differences, specifically in the distribution of comorbid conditions, patients being discharged to their home without care are likely healthier than those discharged to their home with care and, therefore, less likely to be readmitted.

From the results for the composite end point analysis (see the third set of columns), we see a dramatic shift in the hospice association (HR 5.95; 95% confidence interval 5.58–6.36), suggesting that death plays an important role in the analysis of these patients. A similar but less dramatic shift is also seen for the intensive care facility/skilled nursing facility association: in contrast to the nonsignificant association when readmission is taken as the outcome, the association is statistically significant under the composite end point analysis (HR 1.28; 95% 1.22–1.33).

Although a comparison across the univariate analyses provides insight into the role that death plays in the analysis of readmission, the results from the illness–death model provide a more formal assessment. Focusing on the variance component, there is clear evidence of between-patient heterogeneity that is not accounted for by \mathbf{X} and, consequently, that T_1 and T_2 are positively correlated: θ is estimated to be 2.60 (95% confidence interval 2.34, 2.89), and a likelihood ratio test for the inclusion of the patient-specific frailties into the model³⁹ was found to be highly statistically significant ($P < 0.001$). The practical impact of this can be seen in several of the risks factors, most notably the length of stay and the discharge destination variables. We see, for example, substantive differences in the effects of a 7- to 13-day stay (HR 1.47 versus 1.31) and a stay of 14 days or longer (HR 2.03 versus 1.63). We also find substantive differences in the effects of being discharged to an intensive care facility/skilled nursing facility; under the univariate analysis, the HR is estimated to be 1.05 and is not statistically significant; under the semi-competing risks analysis, it is estimated to be 1.13 and is statistically significant. Also of note is that for several of the risk factors, there are meaningful differences in the associations for death pre- and post-readmission. This is most evident for the length of stay and discharge destination variables but also manifests for African Americans. That these differences exist also speaks to the dependence between readmission and death as outcomes.

Finally, Figure 4 provides results for the baseline survivor function corresponding to $h_{01}(t_1)$ in expression 4 and denoted

Table 2. Estimates of Odds Ratio (OR) and Hazard Ratio (HR) Parameters, Together With Associated 95% Confidence Intervals (CI), From a Series of Analysis Applied to the Medicare Heart Failure Data*

	Univariate Analyses						Semi-Competing Risks Analysis					
	Logistic Regression		Survival Analysis									
	Readmission		Readmission		Composite End Point		Readmission		Death Before Readmission		Death After Readmission	
	OR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Sex												
Female	1.00		1.00		1.00		1.00		1.00		1.00	
Male	1.06	(1.01, 1.11)	1.07	(1.02, 1.11)	1.11	(1.07, 1.15)	1.11	(1.05, 1.17)	1.34	(1.23, 1.46)	1.29	(1.15, 1.45)
Age, y												
65–74	1.00		1.00		1.00		1.00		1.00		1.00	
75–84	0.89	(0.84, 0.95)	0.91	(0.86, 0.96)	0.98	(0.93, 1.03)	0.90	(0.84, 0.97)	1.40	(1.2, 1.62)	1.36	(1.13, 1.64)
85–94	0.86	(0.80, 0.91)	0.89	(0.84, 0.94)	1.01	(0.96, 1.07)	0.89	(0.83, 0.96)	1.93	(1.67, 2.23)	1.98	(1.66, 2.36)
95+	0.71	(0.64, 0.80)	0.78	(0.70, 0.86)	1.07	(0.99, 1.16)	0.82	(0.72, 0.93)	3.30	(2.75, 3.97)	2.52	(1.93, 3.30)
Race												
White	1.00		1.00		1.00		1.00		1.00		1.00	
African American	0.93	(0.83, 1.05)	0.93	(0.83, 1.03)	0.88	(0.80, 0.97)	0.89	(0.78, 1.02)	0.60	(0.45, 0.8)	0.72	(0.51, 1.02)
Other	1.01	(0.88, 1.15)	1.02	(0.90, 1.15)	0.98	(0.88, 1.09)	1.05	(0.89, 1.23)	0.82	(0.60, 1.12)	0.85	(0.58, 1.24)
Length of stay, d												
1–3	1.00		1.00		1.00		1.00		1.00		1.00	
4–6	1.13	(1.07, 1.19)	1.12	(1.06, 1.17)	1.11	(1.06, 1.16)	1.15	(1.08, 1.22)	1.18	(1.06, 1.31)	1.34	(1.15, 1.55)
7–13	1.31	(1.23, 1.40)	1.31	(1.24, 1.39)	1.41	(1.34, 1.47)	1.47	(1.37, 1.59)	1.97	(1.76, 2.22)	1.64	(1.39, 1.93)
14+	1.63	(1.45, 1.83)	1.63	(1.47, 1.80)	1.73	(1.59, 1.88)	2.03	(1.76, 2.34)	2.79	(2.31, 3.38)	2.01	(1.55, 2.59)
Discharge destination												
Home	1.00		1.00		1.00		1.00		1.00		1.00	
Home with care	0.75	(0.71, 0.80)	0.77	(0.72, 0.81)	0.75	(0.71, 0.79)	0.70	(0.65, 0.75)	0.51	(0.44, 0.61)	0.71	(0.58, 0.86)
ICF/SNF	1.02	(0.97, 1.08)	1.05	(1.00, 1.10)	1.28	(1.22, 1.33)	1.13	(1.06, 1.21)	2.95	(2.65, 3.29)	2.54	(2.22, 2.91)
Hospice	0.07	(0.05, 0.10)	0.18	(0.13, 0.26)	5.95	(5.58, 6.36)	0.63	(0.43, 0.93)	151.8	(126.2, 182.6)	15.3	(7.04, 33.2)

ICF indicates intensive care facility; and SNF, skilled nursing facility.

*Adjusted for state.

$S_{01}(t_1)$. Note that this function is the readmission-specific survivor function for a white, female patient aged 65 to 74 years, admitted to a hospital in MA for heart failure, hospitalized for no longer than 3 days, and subsequently discharged alive to their home with care. Overall, the univariate survival analysis generally indicates better (baseline) survival, with respect to readmission, across the first 30 days post discharge relative to that indicated by the semi-competing risks analysis. This overestimation is likely because of the univariate analysis erroneously assuming death to be an independent censoring mechanism. Given the strong evidence of dependence between readmission and death, this assumption is likely unrealistic. Nevertheless, the univariate analysis, intuitively, takes patients who do not die to be representative of those who do die, a consequence being that the higher (latent) risk for readmission among those who die is not accounted for adequately. In the semi-competing risks analysis, however, the dependence

between readmission and death is formally part of the model specification, so that the higher (latent) risk for readmission among those who die is taken into account; hence, the baseline survivor function for readmission is estimated to be lower.

Discussion

In this article, we reviewed the semi-competing risks framework as a means investigating variation in risk for a non-terminal event in settings where occurrence of the event is subject to a terminal event. We have also described a range of approaches that researchers may use to analyze semi-competing risks data. In principle, each of these approaches estimates different quantities (as highlighted by the notation we adopt), so that they could be viewed as addressing different scientific questions. As we have elaborated on, however, the potentially informative role that death plays as a dependent competing risk, both in terms of the impact on parameter

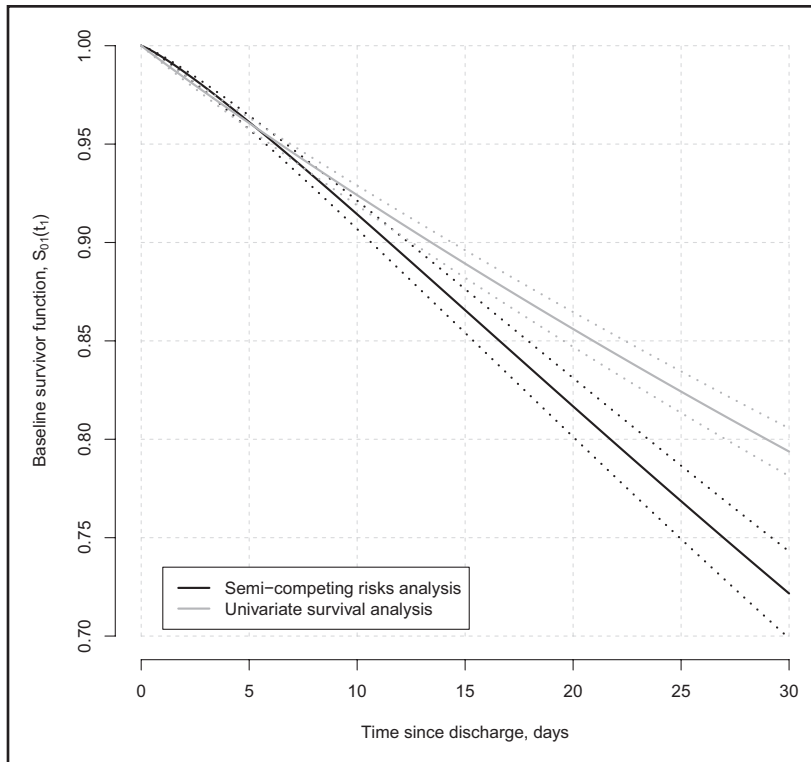


Figure 4. Estimates and pointwise 95% confidence intervals for the readmission-specific baseline survivor function from the Markov illness–death model.

interpretation and estimation, requires specific consideration. Although our goal is not to be prescriptive in the analysis of semi-competing risks data (arguably, the science should play the most important role in how to move forward), we think that key benefits of the semi-competing risks framework are that (1) estimates of risk factor associations for the nonterminal event explicitly account for the terminal event as a dependent competing risk; (2) statistical inference accounts for within-patient correlation between the 2 event times via the use of patient-specific frailties; and (3) the quantification of the dependence via the variance component θ and the interplay between $h_2(t_2)$ and $h_3(t_2|t_1)$.

Throughout this article, we have focused on the estimation of covariate effects on the risk of readmission while considering the potential role that death plays in the analysis. Although covariate effects are of interest in a broad range of settings, they are not always of primary interest. In the context that motivates the application, we present, for example, that primary interest lies with characterizing and understanding variation in hospital-specific readmission rates. Toward this, appropriate modeling of covariate effects remains crucial because its purpose is to ensure control for case-mix adjustment. Nevertheless, the extent to which the semi-competing risks framework provides improved characterization of between-hospital variation in readmission rates over, say, the competing risks framework is not known and is the subject of our on-going research.⁴⁰ Given the inherent inability of competing risks analyses to identify dependence between events, the semi-competing risks framework has substantial promise in efforts to investigate how hospital-specific readmission and mortality rates vary jointly.⁴¹

Although the Medicare data and analyses we present serve primarily to illustrate the ideas and methods, several aspects

of the example deserve comment. First, our analyses do not attempt to incorporate a comprehensive risk adjustment such as one based on using the Centers for Medicare and Medicaid Services-hierarchical condition categories model⁴² risk score, nor do they include other risk factors for readmission or death, such as whether the patient had a do-not-resuscitate order. Including these into the models would likely change the results in Table 2 and Figure 4, specifically by accounting for some of the between-patient variation currently absorbed by the patient-specific frailties. Second, the methods and analysis of the Medicare data set assumes \mathbf{X} to be the same for all 3 transition-specific hazards. This restriction is not necessary, either theoretically or with respect to use of the code available in the SemiCompRisks package. As such, analysts may choose to include different sets of covariates into the 3 models as appropriate for their specific scientific contexts. Third, for simplicity, we restrict attention to the Weibull distribution for the baseline hazard functions in all analyses. Although often considered reasonable, the Weibull distribution will not be sufficiently flexible in all settings. Recent work in the statistical literature, however, provides methods that permit researchers to either avoid specifying the baseline hazard functions²⁹ (as is typically done for the Cox model) or to specify very flexible shapes.^{24,43} Examination and comparison of model fit across alternative specifications, a critical feature of most analyses, would help in deciding modeling assumptions are reasonable. Although methods toward this have been developed in the competing risks context,⁴⁴ to the best of our knowledge, they have not in the semi-competing risks context and is therefore an important avenue for future work.

Beyond extensions that permit flexible specification of the baseline hazard function, the statistical literature is rich with methods for a broad range of settings that build on the

one presented here. For example, methods exist to accommodate more general multistate data structures, including those for settings where the time frame is relatively long and the nonterminal event is potentially recurrent (eg, if interest lies with patients' experience in the first year postdischarge then they may experience in more than one readmission)^{45,46} and there are >1 type of nonterminal event.^{30,47} Beyond the multistate approach, 2 other general frameworks have been proposed in the literature. Briefly, instead of directly modeling the transition-specific hazard functions, the first of these frameworks considers the 2 marginal event-specific survivor functions.^{23,48,49} These are then linked via a copula, a mathematical construct used to specify the joint distribution of 2 variables (and, hence, their dependence). The second framework is grounded in the principles of causal inference, specifically principle stratification, so that the focus is on defining and estimating causal effects of a particular treatment.^{50,51} As mentioned, the choice to focus on the illness-death model was primarily because its relation to well-known methods for survival analysis and the availability of software. Although we view all of these methods as being complementary in helping researchers address a broad range of scientific questions, a more comprehensive comparison of these methods represents an important ongoing extension of our work.

Sources of Funding

This work was supported by National Cancer Institute grant P01 CA134294-02 and National Institutes of Health grant R01 CA181360-01.

Disclosures

None.

References

- Krumholz HM, Parent EM, Tu N, Vaccarino V, Wang Y, Radford MJ, Hennen J. Readmission after hospitalization for congestive heart failure among Medicare beneficiaries. *Arch Intern Med*. 1997;157:99–104.
- Keenan PS, Normand SL, Lin Z, Drye EE, Bhat KR, Ross JS, Schuur JD, Stauffer BD, Bernheim SM, Epstein AJ, Wang Y, Herrin J, Chen J, Federer JJ, Mattera JA, Wang Y, Krumholz HM. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. *Circ Cardiovasc Qual Outcomes*. 2008;1:29–37. doi: 10.1161/CIRCOUTCOMES.108.802686.
- Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med*. 2009;360:1418–1428. doi: 10.1056/NEJMs0803563.
- Krumholz HM, Merrill AR, Schone EM, Schreiner GC, Chen J, Bradley EH, Wang Y, Wang Y, Lin Z, Straube BM, Rapp MT, Normand SL, Drye EE. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. *Circ Cardiovasc Qual Outcomes*. 2009;2:407–413. doi: 10.1161/CIRCOUTCOMES.109.883256.
- Krumholz HM, Lin Z, Drye EE, Desai MM, Han LF, Rapp MT, Mattera JA, Normand SL. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. *Circ Cardiovasc Qual Outcomes*. 2011;4:243–252. doi: 10.1161/CIRCOUTCOMES.110.957498.
- Joynt KE, Orav EJ, Jha AK. Thirty-day readmission rates for Medicare beneficiaries by race and site of care. *JAMA*. 2011;305:675–681. doi: 10.1001/jama.2011.123.
- Epstein AM, Jha AK, Orav EJ. The relationship between hospital admission rates and rehospitalizations. *N Engl J Med*. 2011;365:2287–2295. doi: 10.1056/NEJMs01101942.
- Tsai TC, Joynt KE, Orav EJ, Gawande AA, Jha AK. Variation in surgical-readmission rates and quality of hospital care. *N Engl J Med*. 2013;369:1134–1142. doi: 10.1056/NEJMs01303118.
- Bradley EH, Yakusheva O, Horwitz LI, Sipsma H, Fletcher J. Identifying patients at increased risk for unplanned readmission. *Med Care*. 2013;51:761–766. doi: 10.1097/MLR.0b013e3182a0f492.
- Girotti ME, Shih T, Revels S, Dimick JB. Racial disparities in readmissions and site of care for major surgery. *J Am Coll Surg*. 2014;218:423–430. doi: 10.1016/j.jamcollsurg.2013.12.004.
- Centers for Medicaid and Medicare Services. Hospital Inpatient Quality Reporting Program. 2013. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html>. Accessed August 1, 2015.
- Centers for Medicaid and Medicare Services. Readmissions Reduction Program. 2013. <http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html>. Accessed August 1, 2015.
- Axon RN, Williams MV. Hospital readmission as an accountability measure. *JAMA*. 2011;305:504–505. doi: 10.1001/jama.2011.72.
- Joynt KE, Jha AK. Thirty-day readmissions—truth and consequences. *N Engl J Med*. 2012;366:1366–1369. doi: 10.1056/NEJMp1201598.
- Joynt KE, Jha AK. A path forward on Medicare readmissions. *N Engl J Med*. 2013;368:1175–1177. doi: 10.1056/NEJMp1300122.
- Hawn MT. Unintended consequences of the hospital readmission reduction program. *Ann Surg*. 2015;261:1032–1033. doi: 10.1097/SLA.0000000000001150.
- Normand S-LT, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of Medical Care: issues and applications. *J Am Stat Assoc*. 1997;92:803–814.
- Ash A, Fienberg S, Louis T, Normand S-LT, Stukel T, Utts J, COPSS-CMS White Paper Committee. Statistical issues in assessing hospital performance. 2012.
- Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496–509.
- Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*. 2005;61:223–229. doi: 10.1111/j.0006-341X.2005.031209.x.
- Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*. 2007;26:2389–2430. doi: 10.1002/sim.2712.
- Dignam JJ, Zhang Q, Kocherginsky M. The use and interpretation of competing risks regression models. *Clin Cancer Res*. 2012;18:2301–2308. doi: 10.1158/1078-0432.CCR-11-2097.
- Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001;88:907–919.
- Lee KH, Haneuse S, Schrag D, Dominici F. Bayesian semi-parametric analysis of semi-competing risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *J R Stat Soc Ser C Appl Stat*. 2015;64:253–273. doi: 10.1111/rssc.12078.
- Dignam JJ, Wieand K, Rathouz PJ. A missing data approach to semi-competing risks problems. *Stat Med*. 2007;26:837–856. doi: 10.1002/sim.2582.
- Peng L, Fine JP. Regression modeling of semicompeting risks data. *Biometrics*. 2007;63:96–108. doi: 10.1111/j.1541-0420.2006.00621.x.
- Hsieh JÄ, Wang W, Adam Ding A. Regression analysis based on semicompeting risks data. *J R Stat Soc Ser B*. 2008;70:3–20.
- Meira-Machado L, de Uña-Alvarez J, Cadarso-Suárez C, Andersen PK. Multi-state models for the analysis of time-to-event data. *Stat Methods Med Res*. 2009;18:195–222. doi: 10.1177/0962280208092301.
- Xu J, Kalbfleisch JD, Tai B. Statistical analysis of illness-death processes and semicompeting risks data. *Biometrics*. 2010;66:716–725. doi: 10.1111/j.1541-0420.2009.01340.x.
- Barrett JK, Siannis F, Farewell VT. A semi-competing risks model for data with interval-censoring and informative observation: an application to the MRC cognitive function and ageing study. *Stat Med*. 2011;30:1–10. doi: 10.1002/sim.4071.
- van den Hout A, Fox JP, Klein Entink RH. Bayesian inference for an illness-death model for stroke with cognition as a latent time-dependent risk factor. *Stat Methods Med Res*. 2015;24:769–787. doi: 10.1177/0962280211426359.
- Zeng D, Chen Q, Chen MH, Ibrahim JG; AMGEN RESEARCH GROUP. Estimating treatment effects with treatment switching via semicompeting risks models: an application to a colorectal cancer study. *Biometrika*. 2012;99:167–184. doi: 10.1093/biomet/asr062.

33. Zhang Y, Chen MH, Ibrahim JG, Zeng D, Chen Q, Pan Z, Xue X. Bayesian gamma frailty models for survival data with semi-competing risks and treatment switching. *Lifetime Data Anal.* 2014;20:76–105. doi: 10.1007/s10985-013-9254-8.
34. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. Springer Science & Business Media; 2000.
35. Diggle P, Heagerty P, Liang K-Y, Zeger S. *Analysis of Longitudinal Data*. Oxford, United Kingdom: Oxford University Press; 2002.
36. Gómez G, Lagakos SW. Statistical considerations when using a composite endpoint for comparing treatment groups. *Stat Med.* 2013;32:719–738. doi: 10.1002/sim.5547.
37. Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J Clin Epidemiol.* 2013;66:648–653. doi: 10.1016/j.jclinepi.2012.09.017.
38. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012; 2014.
39. Self SG, Liang K-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc.* 1987;82:605–610.
40. Lee KH, Dominici F, Schrag D, Haneuse S. Hierarchical models for semi-competing risks data with application to quality of end-of-life care for pancreatic cancer. *J Am Stat Assoc.* 2016. In press.
41. Krumholz HM, Lin Z, Keenan PS, Chen J, Ross JS, Drye EE, Bernheim SM, Wang Y, Bradley EH, Han LF, Normand SL. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. *JAMA.* 2013;309:587–593. doi: 10.1001/jama.2013.333.
42. Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Lezzoni LI, Ingber MJ, Levy JM, Robst J. Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev.* 2004;25:119–141.
43. Rondeau V, Mazroui Y, Gonzalez JR. frailtypack: an R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *J Stat Softw.* 2012;47:1–28.
44. Beyersmann J, Allignol A, Schumacher M. *Competing Risks and Multistate Models With R*. New York, NY: Springer Science & Business Media; 2011.
45. Ghosh D, Lin D. Marginal regression models for recurrent and terminal events. *Stat Sin.* 2002;12:663–688.
46. Kneib T, Hennerfeind A. Bayesian semi parametric multi-state models. *Stat Modelling.* 2008;8:169–198.
47. Siannis F, Farewell VT, Head J. A multi-state model for joint modelling of terminal and non-terminal events with application to Whitehall II. *Stat Med.* 2007;26:426–442. doi: 10.1002/sim.2342.
48. Lakhal L, Rivest LP, Abdous B. Estimating survival and association in a semicompeting risks model. *Biometrics.* 2008;64:180–188. doi: 10.1111/j.1541-0420.2007.00872.x.
49. Chen YH. Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime Data Anal.* 2012;18:36–57. doi: 10.1007/s10985-011-9202-4.
50. Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by death. *J Educ Behav Stat.* 2003;28:353–368.
51. Tchetgen Tchetgen EJ. Identification and estimation of survivor average causal effects. *Stat Med.* 2014;33:3601–3628. doi: 10.1002/sim.6181.