Revista Colombiana de Estadística
Junio 2010, volumen 33, no. 1, pp. 43 a 61

Appraisal of Several Methods to Model Time to Multiple Events per Subject: Modelling Time to Hospitalizations and Death

Revisión de varios métodos para modelar tiempo a múltiples eventos por sujeto: modelamiento de tiempo a hospitalizaciones y muerte

Javier Castañeda[a], Bart Gerritse[b]

Clinical Outcomes, Research and Biometry, Cardiac Rhythm Disease Management, Medtronic Bakken Research Center, Maastricht, Netherlands

Abstract

During the disease-recovery process of many diseases, such as in Heart Failure (HF), often more than one type of event plays a role. Some clinical trials use the combined endpoint of death and a secondary event; for instance, HF-related hospitalizations. This is often analyzed with time-to-first-event survival analysis which ignores possible subsequent events, such as several HF-related hospitalizations. Accounting for multiple events provides more detailed information on the disease-control process, and allows a more precise understanding of the prognosis of patients.

In this paper we explore and illustrate several modelling strategies to study time to repeated events of disease-related hospitalizations and death per subject. Marginal models are revised in order to account for intra-subject correlation and competing risks. Finally, we recommend a Multi-state model which allows a flexible modelling strategy that incorporates important features in the analysis of HF-related hospitalizations and death, and at the same time extends relevant characteristics of the Andersen & Gill (1982), Wei et al. (1989) and Prentice et al. (1981) models.

***Key words*:** Survival analysis, Competing risks, Correlated observations, Marginal models.

Resumen

Algunos ensayos clínicos para estudiar el efecto de nuevos tratamientos en pacientes con insuficiencia cardiaca (IC) se basan en la evaluación de hospitalizaciones relacionadas con IC y muerte. Frecuentemente el análisis se enfoca en el tiempo a la primera ocurrencia de alguno de estos dos

[a]Biostatistician. E-mail: javier.castaneda@medtronic.com
[b]Principal Statistician. E-mail: bart.gerritse@medtronic.com

desenlaces. Este tipo de análisis ignora importantes eventos como nuevas hospitalizaciones relacionadas con IC, que permiten una mejor descripción y compresión del proceso de recuperación de estos pacientes.

En este trabajo se describen y exploran varias estrategias para el análisis de tiempo a repetidas hospitalizaciones relacionadas con IC y tiempo a la muerte. Se estudian modelos marginales para incorporar la correlación intra-sujeto y riesgos competitivos propios de este tipo de ensayos clínicos. Finalmente, se recomienda un modelo multi-estado como una estrategia sencilla y flexible que incorpora elementos importantes en el análisis de hospitalizaciones relacionadas con IC y muerte, y a la vez extiende características relevantes de los modelos de Andersen & Gill (1982), Wei et al. (1989) and Prentice et al. (1981).

***Palabras clave*:** análisis de sobrevida, riesgos competitivos, observaciones correlacionadas, modelos marginales.

1. Introduction

Some clinical trials use combined endpoints to evaluate the effect of new therapies. For instance, in the treatment of Heart Failure (HF) patients, a common combined endpoint is death and HF-related hospitalizations (Gheorghiade et al. 2005), and this is often analyzed with a time-to-first-event analysis. In the case of a first event being a hospitalization, this analytical approach ignores subsequent hospitalizations or death. Despite the simplicity of time-to-first-event analysis, this strategy has a severe drawback: the waste of information.

As discussed by Gheorghiade et al. (2005) and Solomon et al. (2007), subsequent events provide detailed information on the disease-control process and are worth modelling to get a more precise understanding of patients' prognoses. The objective of this paper is to explore and formulate a simple and flexible modelling strategy for the joint analysis of survival and time to disease-related hospitalizations. Several marginal models are explored in order to illustrate statistical methods that account for intra-subject correlation. Finally, we propose a multistate model as a flexible modelling strategy for the combined analysis of survival and time to disease-related hospitalizations.

Statistical methods for repeated events survival analysis are illustrated using a HF-dataset derived from the PROSPECT study (Predictors of Response to Cardiac Resynchronization Therapy, study described by Yu et al. 2005) and results published by Chung et al. (2008). The HF-dataset incorporates relevant features and information on HF-related hospitalizations and death for 426 patients, randomly assigned to two treatment groups (G1 and G2).

Figure 1 displays the repeated events nature of the dataset. In 18 patients (7 in G1 and 11 in G2) the first event was death, and for 73 patients (33 in G1 and 40 in G2) the first event was hospitalization. Twenty seven patients presented a second hospitalization (15 in G1 and 12 in G2) and only 6 had a third hospitalization (3 in each group). Ten (3 in G1 and 7 in G2), six (3 in each group) and two (1 in each group) patients died after the first, second and third hospitalization, respectively.

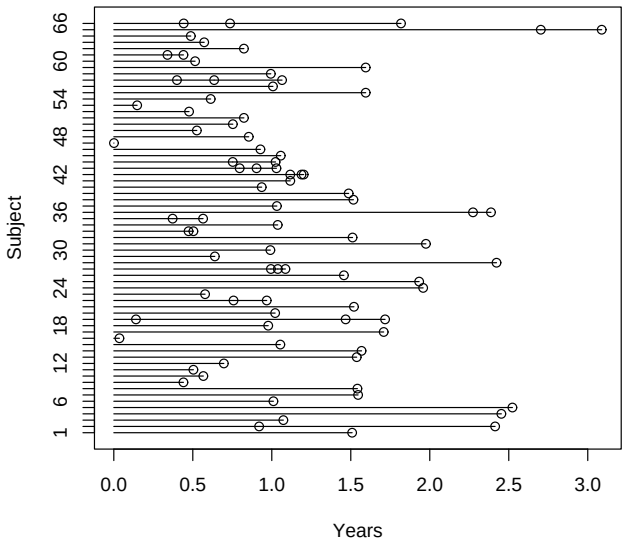

FIGURE 1: Multiple events for 66 patients in the HF-dataset.

In Section 2 several methods and analysis strategies are defined, among them, time to first event Cox proportional hazards model; marginal models for multiple events: Andersen & Gill (1982), Wei et al. (1989) and Prentice et al. (1981) models; and the multi-state model (Andersen & Gill 1982). In Section 3 the pros and cons of these modelling strategies are illustrated using the HF-dataset, analyzing the time to death and/or disease-related hospitalizations. Finally, in Section 4 a discussion is presented and the use of a multi-state model is recommended for the analysis of the HF-dataset. These methods are described next (the descriptions of the Cox and marginal models are based primarily on Therneau & Grambsch (2000)).

2. Methods

Statistical methods for survival analysis, such as the Kaplan-Meier estimator, log-rank test and Cox regression model, can be rewritten as stochastic integrals with respect to counting processes and martingale theory. The counting processes approach is used in this section for the presentation of the different methods (for instance in the description of the Cox model) and the reason for this, as explained by Fleming & Harrington (1991), Therneau & Grambsch (2000) and Andersen et al. (1993), is that counting processes provide a single, elegant, solid basis for survival analysis, which allows flexible ways of modelling (allowing for extensions of the basic survival analysis models to more general multi-state models applicable for event history data) and lead to a unified framework for studying both small sample and asymptotic properties of survival analysis statistics. A complete discussion on

counting processes can be found in the books by Fleming & Harrington (1991) and Andersen et al. (1993).

2.1. The Cox Model

Let $X_{ij}(t)$ be the jth covariate of the ith subject, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$ and X_i denotes the covariate vector for subject i. The hazard for individual i is specified as $\lambda_i(t) = \lambda_0(t)e^{X_i(t)\beta}$, where λ_0 is an unspecified nonnegative function (the baseline hazard), and β is a vector of coefficients. For untied failure time data, Cox (1972) proposed the estimation of β based on the partial likelihood function:

$$PL(\beta) = \prod_{i=1}^{n} \prod_{t \geq 0} \left\{ \frac{Y_i(t) r_i(\beta, t)}{\sum_j Y_j(t) r_j(\beta, t)} \right\}^{dN_i(t)}$$

where $Y_i(t)$ is the indicator function that subject i is still under observation at time t, $N_i(t)$ is the number of observed failures for subject i and $dN_i(t)$ denotes the increment in $N_i(t)$ over the infinitesimal time interval $[t, t + dt)$. $r_i(\beta, t)$ is the risk score for subject i, $r_i(\beta, t) = \exp[X_i(t)\beta] \equiv r_i(t)$. The product integral is defined such that only time points where patient i is at risk ($Y_i(t) = 1$) generate a contribution. Therefore, it is convenient to write the partial likelihood function as:

$$PL(\beta) = \prod_{i=1}^{n} \prod_{t : Y_i(t)=1} \left\{ \frac{r_i(\beta, t)}{\sum_j Y_j(t) r_j(\beta, t)} \right\}^{dN_i(t)}$$

The log partial likelihood is:

$$l(\beta) = \sum_{i=1}^{n} \int_{t:Y_i(t)=1} \left[\log(r_i(\beta, t)) - \log \left(\sum_j Y_j(t) r_j(\beta, t) \right) \right] dN_i(t)$$

$$= \sum_{i=1}^{n} \int_0^{\infty} Y_i(t) \left[X_i(t)\beta - \log \left(\sum_j Y_j(t) r_j(t) \right) \right] dN_i(t)$$

After differentiating the log partial likelihood with respect to β, the $p \times 1$ score vector, $U(\beta)$ is:

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\infty} [X_i(s) - \overline{x}(\beta, s)] dN_i(s)$$

where $\overline{x}(\beta, s)$ is a weighted mean of X, over those observations still at risk at time s,

$$\overline{x}(\beta, s) = \frac{\sum Y_i(s) r_i(s) X_i(s)}{\sum Y_i(s) r_i(s)}$$

$\widehat{\beta}$ is found by solving the partial likelihood equation $U\big(\widehat{\beta}\big) = 0$, and is consistent and asymptotically normally distributed with mean β, and variance $I^{-1}\big(\widehat{\beta}\big)$, the inverse of the observed information matrix.

2.2. Multiple events per subject

A major issue in extending proportional hazards regression models to multiple events per subject is the intra-subject correlation (Therneau & Grambsch 2000). A simple approach that sidesteps this is to take time to first event. Time to first event is simple and easy to interpret, but important information on the disease-recovery process is lost. Other more appropriate approaches are marginal models and multi-state models (with competing risk component). These methods are described below.

2.2.1. Marginal Models

Marginal models offer flexibility in the formation of strata and risk sets, definition of the time scale, and have a well-developed estimator of the variance. Marginal models allow for population average estimation of treatment effect. Therneau & Grambsch (2000) summarize the analysis with these models in three steps:

- Decide on a model (issues such as covariate selection, inclusion of strata, etc.) and structure the data set accordingly.
- Fit the data as an ordinary Cox model, ignoring the possible intra-subject correlation (i.e. treating multiple events from a subject as independent).
- Replace the standard variance estimate with one which is corrected for the possible correlations.

Robust Variance. When a given subject may contribute multiple events, the assumption of independent observations in the standard Cox model does not hold. Lipsitz et al. (1990) showed that marginal models can overcome this assumption for the estimation of the variance of $\widehat{\beta}$ by an appropriate correction based on a grouped jackknife estimate.

Grouped-jackknife values are defined as $J_i = \widehat{\beta} - \widehat{\beta}_{(i)}$, where $\widehat{\beta}_{(i)}$ is the result of the fit that includes all of the individuals except individual i. It is denominated as grouped because in the multiple event case, one individual contributes several observations, and removing a subject implies removing a group of observations. Therneau & Grambsch (2000) describe a way to compute the grouped-jackknife values directly at the Newton-Raphson iteration. The change in the estimated coefficient vector can be expressed in the following way, $\Delta\beta = 1'(UI^{-1}) \equiv 1'D$, where U is the matrix of score residuals. Thus, the change in $\widehat{\beta}$ at each iteration is the column sum of a matrix D, defined as the score residual scaled by I^{-1} (the variance of $\widehat{\beta}$).

This grouped jackknife can be used to derive a robust estimate of the variance for the Cox model. If J is the matrix of grouped-jackknife values (i.e. the ith row of J is $\widehat{\beta} - \widehat{\beta}_{(i)}$), then the grouped jackknife estimate of the variance can be written as the matrix product $V_J = \frac{n-1}{n}\big(J - \overline{J}\big)'\big(J - \overline{J}\big)$, where $\overline{J}$ is the matrix

of column means of J. A natural approximation that is preferred is $D'D$, the matrix product of the approximate jackknife variances (ignoring the $\frac{n-1}{n}$ term). Writing $D'D = I^{-1}(U'U)I^{-1}$, this variance can be viewed as a sandwich estimator ABA, where A is the usual variance and B a correction term. Sandwich estimates are familiar from robust variance estimation in generalized estimating equations (GEE) proposed by Liang & Zeger (1986). Although unbiased, this grouped-jackknife estimate is typically more variable than the ordinary variance of the Cox model, but it is a robust variance that adequately addresses repeat event correlation, and it is expected to be reported when fitting marginal models.

Ordered events. One important issue is to distinguish between data sets where the multiple events have a distinct ordering and those where they do not. In the particular case of the HF-dataset, the outcomes hospitalizations and death are correlated and ordered. Death can happen either as the first event or after hospitalization; there is a specific ordering in this case, obviously after the event of a death it is not possible to have a hospitalization. The most common approaches for correlated ordered outcomes are: the independent increments (Andersen & Gill 1982), marginal (Wei et al. 1989), or PWP (Prentice et al. 1981) models. All three are "marginal" regression models in that $\widehat{\beta}$ is determined from a fit that ignores the correlation between the events followed by a correction of the variance, but differ considerably in their creation of the risk sets.

Andersen and Gill (AG) model. This method is the simplest, but makes the strongest assumptions. Each subject is represented as a series of observations (rows of data) with time intervals as: (entry time, first event], (first event, second event],..., (mth event, last follow up]. The intensity process for subject i is:

$$Y_i(t)\lambda_0(t)\exp(X_i(t)\beta)$$

The difference with the standard Cox model lies in the definition of the at-risk indicator $Y_i(t)$. For survival data, the individual ceases to be at risk when an event occurs and $Y_i(t)$ takes value zero, but for the AG model for recurrent events, $Y_i(t)$ remains one as events occur. Of course the at-risk indicator does not remain one if the event observed is Death. No extra strata or strata by covariate interaction terms are introduced for the multiple events (Therneau & Grambsch 2000). The Andersen-Gill formulation of the Cox proportional hazards model has a number of advantages, including the ability to accommodate left-censored data, time-varying covariates, multiple events, and discontinuous intervals of risks. Some of these practical advantages are discussed in an applied framework by Johnson et al. (2004).

Wei, Lin and Weissfeld (WLW) model. In this model, the ordered outcome dataset is treated as if it were an unordered competing risk case. The number of strata in the analysis will be equal to the maximum number of events a patient reports in the study. Every subject will have one observation in each stratum.

The hazard function for the jth event for subject i is:

$$Y_{ij}(t)\lambda_{0j}(t)\exp(X_i(t)\beta_j)$$

Unlike the AG model, this model allows a separate underlying hazard for each event and for strata by covariate interactions, as shown by the notation β_j. In the WLW model the at-risk indicator for the jth event, $Y_{ij}(t)$, is one until the occurrence of the jth event, unless, of course, some external event causes censoring. When either of those occurs, it becomes zero, indicating that subject is no longer at risk after the last given event (Therneau & Grambsch 2000, Wei et al. 1989). A frequently raised concern is the method's risk set, where each individual is considered to be at risk of all recurrent events from the start of the observation period, and often this method gives estimates that exceed those provided by alternative approaches. By simulation studies, Metcalfe & Thompson (2007) have shown that the WLW model infringes on the proportional hazards assumption when applied to recurrent events data, but the bias this may cause is not behind the distinctive effect estimates. Metcalfe & Thompson (2007) discuss that the analyses of medical data indicate that the infringement of the proportional hazards assumption is not necessarily greater than that experienced with other applications of proportional hazards regression and need not prohibit the application of WLW's method to recurrent events data.

Prentice, Williams and Peterson (PWP) models. This model clearly defines the order of the events. A subject is not at risk for the kth event until he/she has experienced event $k-1$st. Like in the AG model, time intervals are defined as: (entry time, first event], (first event, second event],. . ., (mth event, last follow up], but each event is assigned to a separate stratum (Prentice et al. 1981). The use of time-depending strata means that the underlying hazard function may vary from event to event, unlike the AG model, which assumes that all events are identical. The hazard function for the jth event for subject i is:

$$Y_{ij}(t)\lambda_{0j}(t)\exp(X_i(t)\beta_j)$$

The primary difference between the WLW and PWP models is in the definition of the at-risk indicator and the definition of the strata in the analysis. In the PWP model the at-risk indicator, $Y_{ij}(t)$, is defined as zero until the $j-1$st event and only then becomes one. Once the jth event occurs, $Y_{ij}(t)$ becomes 0 again. The PWP model can be seen as a stratified AG model with event-specific baseline hazards and a restricted risk set. By means of a simulation study, Kelly & Lim (2000) have illustrated that the naive and robust standard errors for the PWP model appear to be similar regardless of the within-subject correlation.

The AG model and the PWP model can be used in the analysis of repeated failure outcomes of the same type, while the approach by the WLW model can be applied to both multiple events of the same type and multiple events of different types as long as there is not a predetermined ordering. The WLW model has a semi-restricted risk set that allows subjects to be at risk for as many events

as the maximum number of events reported per subject in the study, even if most of the subjects only had one event, which (as reported by Kelly & Lim (2000)) leads to overestimation of the treatment effect. For all models, except the PWP model, the robust standard errors become inflated when within-subject events are not independent (Kelly & Lim 2000). When the model is correctly specified (no important covariates are omitted) the PWP model and the AG model estimate unbiased treatment effect and require similar sample size to obtain the same precision in the estimation, while the WLW model estimates biased treatment effect and requires a larger sample size. The PWP model and the AG model are considered to be more efficient than the WLW model, and require less sample size than the time to first event model (Therneau & Grambsch 2000). As noted by Wei & Glidden (1997), the appropriate modelling strategy should be chosen based on the type and nature of the multiple events structure.

2.2.2. Multi-state models

Several of the ideas presented in this section on multi-state models can be found in the 2002 (11) issue of Statistical Methods in Medical Research, entirely devoted to multi-state models. In particular, in the paper by Andersen & Keiding (2002), a multi-state process is defined as a stochastic process $(X(t), t \in T)$ with a finite state space $S = \{1, \ldots, p\}$ and with right-continuous sample paths: $X(t+) = X(t)$. Here $T = [0, \tau]$ or $[0, \tau)$ with $\tau \leq +\infty$. The process has initial distribution $\pi_h(0) = \text{Prob}(X(0) = h), h \in S$. A multi-state process $X(\cdot)$ generates a history consisting of the observation of the process in the interval $[0, t]$. Relative to this history, transition probabilities may be defined by:

$$P_{hj}(s, t) = \text{Prob}(X(t) = j \mid X(s) = h)$$

for $h, j \in S$, and $s, t \in T, s \leq t$ and transition intensities by the derivatives

$$\alpha_{hj}(t) = \lim_{\Delta t \to 0} \frac{P_{hj}(t, t + \Delta t)}{\Delta t}$$

Graphically, multi-state models may be illustrated using diagrams with boxes representing the states and with arrows between the states representing the possible transitions. A state $h \in S$ is *absorbing* if for all $t \in T, j \in S, j \neq h, \alpha_{hj}(t) = 0$; otherwise h is *transient*. The most simple multi-state model is the two-state model for survival data, which is represented in Figure 2. This model has $p = 2$ states and only one possible transition from state 0 to state 1 (state 0: alive and state 1: dead). The corresponding transition intensity $\alpha_{01}(t)$ is given by the hazard rate function $\alpha(t)$, while $\alpha_{10}(t) = 0$ for all t, that is, state 1 is absorbing. Covariates may be entered into the model using a regression model for $\alpha(\cdot)$. A throughout description of the counting process representation of the multi-state model can be found in the paper by Andersen & Keiding (2002).

In multi-state models, an individual moves from one state to another through time. It is clear that intermediate events, such as disease-related hospitalizations, provide more detailed information on the disease-recovery process and allow for

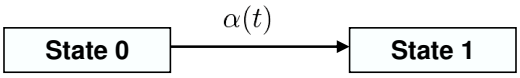

FIGURE 2: The two-state model for survival data.

more precision in predicting the prognosis of patients. These hospitalizations become intermediate events worth modelling in their natural form. These non-fatal events during the course of the disease can be seen as transitions from one state to another. The time origin is characterized by a transition into an initial transient state, such as the start of treatment (entry). Instead of survival data or time-to-event data, data on the history of events is available. Multi-state models provide a framework that allows for the analysis of such event history data (Putter et al. 2007).

If it is assumed that the future depends on the history only through the present, then the process is assumed to be Markovian. In our particular case, given the present state and the event history of a patient, the next state to be visited and the time at which this will occur will only depend on the present state. As explained by Putter et al. (2007), the counting process style of data input with time intervals of (entry time, first event], (first event, second event],..., (mth event, last follow up] can be used by a Markov model.

Andersen & Keiding (2002), Klein & Shu (2002) and Cook & Lawless (2002), discuss the use of multi-state models when the observation plan has been that of cohorts of individuals observed continuously over time. Commenges (2002) treats the situation where individuals are not observed continuously, but only at discrete time points.

3. Statistical Analysis and Results

In this section, each model description indicates how group effect is modelled and what other options are supported by each model. Special attention is given to model assumptions and practical implications for setting up of the dataset.

3.1. Time to first event

The dataset illustrated in Figure 1 was analyzed by a usual "time to first event" analysis. As a first approximation, both outcomes, "hospitalization" and "death", are treated as the same event. The dataset for this analysis has one observation per subject including: Patient identification (ID), time to event in years (TIME), status (event=1 or censoring=0) (STATUS), and the covariates: treatment group (1 or 2)(GROUP), age at enrolment (AGE), and left ventricular ejection fraction (LVEF). Ejection fraction is the percentage of blood pumped out of the left ventricle with each heartbeat. Adamson et al. (2004) have reported

ventricular ejection fraction as a prognostic factor for HF. For patients with ID numbers 35, 36, 37 and 47, the dataset would look as presented in Table 1.

Table 1: Dataset for time to first event analysis (patients with ID 35, 36, 37 and 47).

ID	TIME	STATUS	GROUP	AGE	LVEF
35	0.3723	1	2	63	15
36	2.2735	1	2	50	20
37	1.0322	0	1	52	30
47	0.0010	1	1	79	25

Despite the convenient simplicity of this analysis, there are two important drawbacks. The first one is that information is being wasted-only the first event is considered; the second one is the identical treatment of hospitalization and death, which clinically should be treated as different events.

A more appropriate alternative is an analysis of time to first event, where event can be either hospitalization or death. Since the baseline hazards for the two event types are not expected to be the same, the analysis is stratified by type of event. The dataset for the combined analysis must allow accommodation for competing risks. In this case, the dataset contains one stratum for each outcome type, with each patient appearing in each stratum. The rows in the dataset for patients with ID numbers 35, 36, 37 and 47 are presented in Table 2.

Table 2: Dataset for time to first event analysis with competing risks.

ID	TIME	STATUS	OUTCOME	GROUP	AGE	LVEF
35	0.3723	0	Death	2	63	15
35	0.3723	1	Hospitalization	2	63	15
36	2.2735	0	Death	2	50	20
36	2.2735	1	Hospitalization	2	50	20
37	1.0322	0	Death	1	52	30
37	1.0322	0	Hospitalization	1	52	30
47	0.0010	1	Death	1	79	25
47	0.0010	0	Hospitalization	1	79	25

Note that the first event for patients with ID numbers 35 and 36 is hospitalization, which removes the individuals from being at risk for the first event, Death (therefore, for the first event, Death, Status appears as censor). The lost to follow up of patient with ID=37 makes the status for this individual as censored for both Death and Hospitalization. Finally, the patient with ID=47 has the event Death, therefore removing the patient from being at risk for the first event: Hospitalization.

Several models can be analyzed. However, the interest has been to evaluate the treatment effect adjusted by AGE and LVEF. In the above two model approaches (first one ignoring type of event and second one with competing risks), the estimated treatment effect is not significant (p-value$>$0.6).

3.2. Ordered multiple events

The first alternative to account for multiple events is the AG model. This model treats the two events, hospitalization and death, as if they were the same. Subjects can re-enter the same state multiple times (Figure 3).

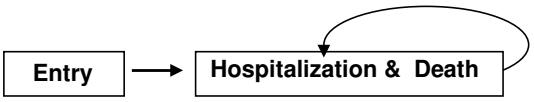

FIGURE 3: Schematic for the AG model approach.

We have the following history of events for patients with ID numbers from 35 to 39 (All patients start at time=0, Table 3). The first event for patient 35 is Hospitalization at 0.372 years after entry; after this the patient has the event Death at time 0.565 years. The first event for the patient with ID=36 is Hospitalization at 2.273 years; after this the patient is loss to follow up at 2.387 years. Patients with ID=37 and 38 do not have any event, they present censoring at times 1.032 and 1.516 years, respectively. Finally, for the patient with ID=39, the first event is hospitalization at 1.117 years after entry, the second event is hospitalization at 1.188 years, and later censoring at time 1.202. Table 3 shows the dataset for the AG model.

TABLE 3: Dataset for multiple events analysis with the AG model (columns: ID, Time1, Time2, Status, Group, Age and LVEF). To adjust a PWP model it is only necessary to add the last column (EventNumber).

ID	TIME1	TIME2	STATUS	GROUP	AGE	LVEF	EventNumber
35	0	0.3723	1	2	63	15	1
35	0.3723	0.5651	1	2	63	15	2
36	0	2.2735	1	2	50	20	1
36	2.2735	2.3874	0	2	50	20	2
37	0	1.0322	0	1	52	30	1
38	0	1.5168	0	1	70	20	1
39	0	1.1170	1	1	80	15	1
39	1.1170	1.1882	1	1	80	15	2
39	1.1882	1.2019	0	1	80	15	3

Note that "type of event" is ignored as one of the assumptions in the AG model. Table 4 shows the parameter estimates for the AG model formulation. As expected in marginal models, robust standard errors (S.E.) are generally larger than model based S.E.

TABLE 4: Parameter estimates for the AG model formulation.

Effect	Parameter estimate	Model based S.E.	Robust S.E.	p-value
GROUP	−0.2787	0.1784	0.2405	0.250
AGE	0.0048	0.0075	0.0089	0.590
LVEF	−0.0562	0.0120	0.0169	0.001

The AG model assumes that all events are identical, which may be too strong an assumption. Figure 4 presents the cumulative hazards for the consecutive events. It clearly suggests that the risk of a new event does not remain constant; on the contrary, it shows that the risk of an event depends on previous events.

Table 5: Parameter estimates for the conditional PWP model formulation.

Effect	Parameter estimate	Model based S.E.	Robust S.E.	p-value
GROUP	−0.3720	0.1809	0.1843	0.044
AGE	0.0028	0.0073	0.0061	0.640
LVEF	−0.0412	0.0121	0.0129	0.001

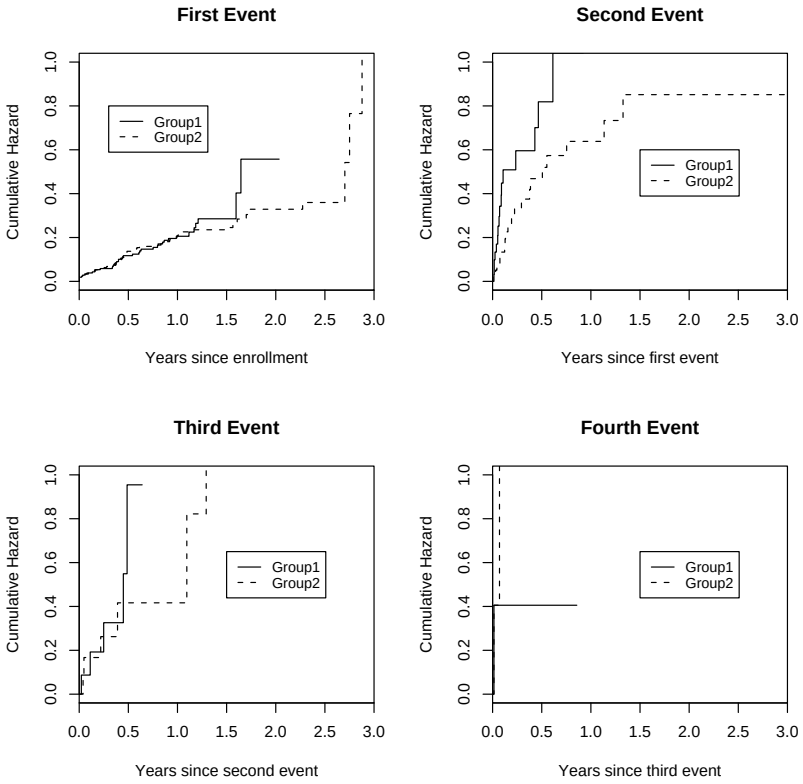

Figure 4: Cumulative hazard for consecutive events.

There is a pre-specified order in the events in HF-patients. A model that encompasses this feature is the PWP model. This model was fitted to incorporate the use of time-depending strata, which means that the underlying hazard function may vary from event to event. The dataset needed to fit a PWP model is the same one used for the AG model, the only difference being that it includes an extra variable, the number of the event (column: EventNumber in Table 3), which is

used for stratification. Table 5 shows the parameter estimates for the conditional PWP model formulation.

The following R code could be used to fit the AG and PWP models.

```
> library(survival)
# AG model #
> coxph(Surv(time1, time2, status) ~ factor(group) + age + lvef
 + cluster(id) , data=data1)
# PWP model #
> coxph(Surv(time1, time2, status) ~ factor(group) + age + lvef
 + cluster(id) + strata(EventNumber), data=data1)
```

3.3. Multi-state models

Two of the characteristics of the HF-dataset are captured by the conditional PWP model presented above, which is able to detect a significant difference between the two treatment groups after controlling for AGE and LVEF (Table 5). First, a patient is not at risk for the kth event until he/she has experienced event $k-1$st. Second, the underlying hazard function may vary from event to event. Unfortunately, the conditional PWP model does not capture the distinction between the two types of events. A very important feature is that Hospitalization and Death, in practice (clinically), cannot be considered equal due to their nature and severity.

The maximum number of hospitalizations for a patient in the HF-dataset is 3. Follow-up after the third hospitalization stops either due to Death or Censoring. Other patients have one or two hospitalizations and afterwards die. A multi-state model where patients transit from one state to another, was proposed to include re-hospitalization and to distinguish the event Death from the event Hospitalization.

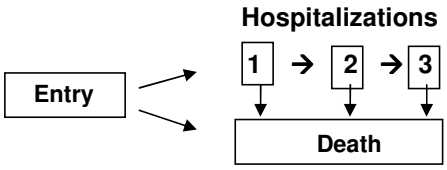

FIGURE 5: Schematic for the full multi-state model.

The multi-state model displayed in Figure 5 does not fit neatly into any of the marginal models described above. The first event can be either Death or Hospitalization (competing risks), a patient is not at risk for a second Hospitalization until he/she experiences Hospitalization as a first event (ordered events), and the underlying risk for the transition from first Hospitalization to Death cannot be assumed to be the same for the transition from the first Hospitalization to a second

Hospitalization. A similar argument holds for the different risks for patients with two or three hospitalizations. In the computer model of the data, there will be seven possible strata-one for each possible transition (Table 6).

Table 6: Strata for the full multi-state model.

Transition	Stratum	Representation
1	Entry - Hospitalization1	E→H1
2	Hospitalization1 - Hospitalization2	H1→H2
3	Hospitalization2 - Hospitalization3	H2→H3
4	Entry - Death	E→D
5	Hospitalization1 - Death	H1→D
6	Hospitalization2 - Death	H2→D
7	Hospitalization3 - Death	H3→D

In setting up the dataset for the multi-state model one must consider the possible transitions that a patient could have at a particular state. For example, the patient with ID=26 has no events, only a censoring at time 1.454 years. This patient has two possible transitions, that is, from Entry to Hospitalization1 (E→H1), or from Entry to Death (E→D); the status in both cases is censoring (0). For the patient with ID=27 the first event is Hospitalization at 0.621 years after entry, his second event is another Hospitalization at time 0.644 years, and finally the patient presents the event Death at 0.672 years after entry. The dataset structure for these two patients is presented in Table 7.

Table 7: Dataset for the full multi-state model.

ID	TIME1	TIME2	TRANSITION	STATUS	GROUP	AGE	LVEF
26	0	1.4543	E→ H1	0	1	73	25
26	0	1.4543	E→D	0	1	73	25
27	0	0.6215	E→H1	1	2	50	20
27	0	0.6215	E→D	0	2	50	20
27	0.6215	0.6439	H1→H2	1	2	50	20
27	0.6215	0.6439	H1→D	0	2	50	20
27	0.6439	0.6720	H2→H3	0	2	50	20
27	0.6439	0.6720	H2→D	1	2	50	20

Note that after the first event (Hospitalization1), the patient with ID=27 has two possible transitions: from Hospitalization1 to Hospitalization2 (H1→H2), or from Hospitalization1 to Death (H1→D). Once the second event is Hospitalization2, the patient is no longer at risk to present transition (H1→D), therefore, the status for this transition is a censoring (0). The same mechanism is applied to the final event.

Table 8 shows parameter estimates for the multi-state model. Due to the assumption of unequal risk for the different transitions, the analysis is stratified by transition.

This multi-state model fully describes the characteristics of the multiple events of the HF-dataset, and is able to detect a significant difference between the two treatment groups after controlling for AGE and LVEF. Subjects in treatment group 2 have a "rate" of state change that is approximately 67% of the rate of those in

Table 8: Parameter estimates for the full multi-state model.

Effect	Parameter estimate	Model based S.E.	Robust S.E.	p-value
GROUP	−0.3880	0.1815	0.1858	0.037
AGE	0.0036	0.0074	0.0061	0.550
LVEF	−0.0414	0.0122	0.0130	0.001

treatment group 1. The assumption for proportional hazards is not rejected (p-value=0.2308). As proposed by Grambsch & Therneau (1994), this proportional hazards test can be seen as a test for assessing the correlation of a scatter plot of the scaled Schoenfeld residuals versus a function of time. This test is implemented and available in R.

The following R code could be used to fit the multi-state model and test for proportional hazards.

```
> library(survival)
> fit1 <- coxph(Surv(time1, time2, status) ~ factor(group) + age
+ lvef + cluster(id) + strata(transition) , data=multistate)
> print(cox.zph(fit1))
```

We can also explore whether treatment affects some transitions more than others by looking at the GROUP by TRANSITION interaction in Table 9. The group effect is strongest with respect to transitions after the first hospitalization (group effect in the H3→D transition is not estimated due to lack of cases in the data). The group by transition interaction is certainly interesting and has to be rigorously proven in order to make conclusions.

Table 9: GROUP by TRANSITION interaction in the full multi-state model.

Effect	Parameter estimate	Exp of parameter estimate
GROUP,E→H1	−0.11310	0.893
GROUP,E→D	−0.18208	0.834
GROUP,H1→H2	−1.03639	0.355
GROUP,H1→D	−0.06922	0.933
GROUP,H2→H3	−1.51823	0.219
GROUP,H2→D	−1.03254	0.356

A simplification of the multi-state model for the HF-dataset is done by assuming the baseline hazards of the H1→H2 and H2→H3 transitions to be proportional, i.e. to have only one Hospitalization state that can be visited more than once. More generally, we may assume a simplification of the multi-state model as shown in Figure 6. Furthermore, this simplification is assuming that baseline hazards for transitions H1→D, H2→D and H3→D are proportional.

To assess whether the assumption of proportional hazards for the transitions is reasonably fulfilled, we explore the estimated cumulative baseline hazards for the multi-state model described in Figure 5. The cumulative baseline hazards for transitions H1→D, H2→D and H3→D are shown in Figure 7.

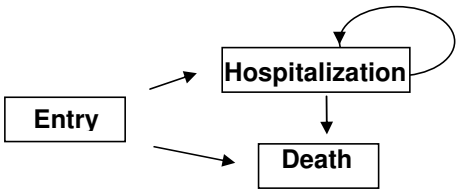

FIGURE 6: Schematic for the simplified multi-state model.

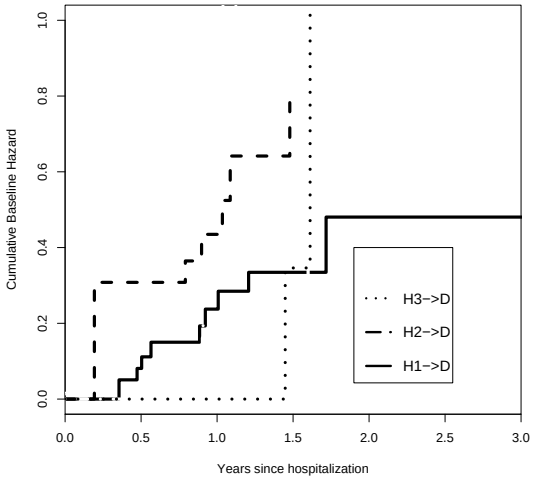

FIGURE 7: Cumulative baseline hazard for transitions H1→D, H2→D and H3→D in the full multi-state model.

The graphical exploration indicates that the baseline hazards are not proportional. This can be checked more rigorously by testing the significance of an interaction between time, and an indicator distinguishing between the three transitions; although this should be considered with caution due to the small sample size. For this case the interaction was significant (p-value <0.01), supporting the graphical check of disproportionate hazards.

The lack of proportionality between the specific transitions suggests not simplifying the multi-state model. The preferred model for the analysis of the HF-dataset is the full multi-state model fitted in Table 8.

4. Conclusion and discussion

Survival analysis has been a common and well-accepted strategy to study treatment effect in a population of patients. During the last few years, there has been an increasing interest in assessing therapy effect not only by using time to Death, but also time to surrogate events; a good example of which is time to hospitalizations. The combined endpoint of time to Death and time to disease-related Hospitaliza-

tions is often analyzed with a time-to-first-event analysis, which has the drawback of waste of information and indistinct handling of two clinically different events.

The analysis of multiple events per subject cannot be approached by a standard Cox model, where the assumption of independence of observations is not valid. In order to account for intra-subject correlation, we have presented the use of marginal and multi-state models using a counting process approach for the joint analysis of survival and time to disease-related Hospitalizations. The characteristics and limitations of the WLW, AG, and PWP models have been illustrated in the modelling of time to HF-related Hospitalizations and Death. All of these models allow for population average estimates of treatment effect and are easily approached using standard statistical software such as SAS, R, and S-Plus. The AG model assumes that all events are identical and can be revisited, and the WLW model only accommodates unordered competing risk. Both models make strong assumptions that are not suitable in the analysis of HF-related Hospitalizations and Death. The PWP model accounts for pre-specified order in the events and competing risks, but has the drawback of assuming Hospitalization and Death as the same type of events, which, given their nature and severity, is clinically unacceptable.

The most appropriate model should be chosen based on the nature of the data. For the HF-dataset we recommend a multi-state model as it allows the incorporation of important features in the analysis of HF-related hospitalizations and death, such as multiple ordered events per subject, event history data, accommodation of competing risks, and the distinction between two different clinical events: death and hospitalization. The proposed simple and flexible multi-state model extends relevant characteristics of the WLW, AG, and PWP models, while capturing important features of time to disease-related Hospitalizations and Death in the HF-dataset, and allows for a more precise understanding of the disease-control process in this particular group of patients.

Acknowledgments

The authors wish to thank the referees for their valuable comments.

[Recibido: enero de 2009 — Aceptado: marzo de 2010]

References

Adamson, P. B., Smith, A. L., Abraham, W. T., Kleckner, K. J., Stadler, R. W., Shih, A. & Rhodes, M. M. (2004), 'Continuous Autonomic Assessment in Patients with Symptomatic Heart Failure. Prognostic Value of Heart Rate Variability Measured by an Implanted Cardiac Resynchronization Device', *Circulation* **110**, 2389–2394.

Andersen, P. K., Borgan, O., Gill, R. D. & Keiding, N. (1993), *Statistical Models Based on Counting Processes*, Springer, New York, United States.

Andersen, P. K. & Gill, R. D. (1982), 'Cox's Regression Model for Counting Pprocesses: A large Sample Study', *Annals of Statistics* **10**, 1100–1120.

Andersen, P. K. & Keiding, N. (2002), 'Multi-State Models for Event History Analysis', *Statistical Methods in Medical Research* **11**, 91–115.

Chung, E. S., Leon, A. R., Tavazzi, L., Sun, J. P., Nihoyannopoulos, P., Merlino, J., Abraham, W. T., Ghio, S., Leclercq, C., Bax, J. J., Yu, C. M., Gorcsan, J., Sutton, S. M., De Sutter, J. & Murillo, J. (2008), 'Results of the Predictors of Response to CRT (PROSPECT) Trial', *Circulation* **117**, 2608–2616.

Commenges, D. (2002), 'Inference for Multi-State Models From Interval-Censored Data', *Statistical Methods in Medical Research* **11**, 167–182.

Cook, R. J. & Lawless, J. F. (2002), 'Analysis of Repeated Events', *Statistical Methods in Medical Research* **11**, 141–166.

Cox, R. D. (1972), 'Regression Models and Life-Tables (with Discussion)', *Journal on the Royal Statistical Society. Series B* .

Fleming, T. R. & Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, John Wiley & Sons, Inc.

Gheorghiade, M., Zannad, F., Sopko, G., Klein, L., Piña, I. L., Konstam, M. A., Massie, B. M., Roland, E., Targum, S., Collins, S. P., Filippatos, G. & Tavazzi, L. (2005), 'Acute Heart Failure Syndromes-Current State and Framework for Future Research', *Circulation* **112**, 3958–3968.

Grambsch, P. M. & Therneau, T. M. (1994), 'Proportional Hazards Tests and Diagnostics Based on Weighted Residuals', *Biometrika* **81**, 515–526.

Johnson, C. J., Boyce, M. S., Schwartz, C. C. & Haroldson, M. A. (2004), 'Modeling Survival: Application of the Andersen-Gill Model to Yellowstone Grizzly Bears', *The Journal of Wildlife Management* **68**, 966–978.

Kelly, P. J. & Lim, L. L. (2000), 'Survival Analysis for Recurrent Event Data: an Application to Childhood Infectious Diseases', *Statistics in Medicine* **19**, 13–33.

Klein, J. P. & Shu, Y. (2002), 'Multi-State Models for Bonemarrow Transplantation Studies', *Statistical Methods in Medical Research* **11**, 117–139.

Liang, K. Y. & Zeger, S. L. (1986), 'Longitudinal Data Analysis Using Generalized Linear Models', *Biometrika* **73**(1), 13–22.

Lipsitz, S. R., Laird, N. M. & Harrington, D. P. (1990), 'Using the Jackknife to Estimate the Variance of Regression Estimators from Repeated Measures Studies', *Communication in Statistics. Theory and Methods* **19**, 821–845.

Metcalfe, C. & Thompson, S. G. (2007), 'Wei, Lin and Weissfeld's Marginal Analysis of Multivariate Failure Time Data', *Statistical Methods in Medical Research* **16**, 103–122.

Prentice, R. L., Williams, B. J. & Peterson, A. V. (1981), 'On the Regression Analysis of Multivariate Failure Time Data', *Biometrika* **68**, 373–379.

Putter, H., Fiocco, M. & Geskus, R. B. (2007), 'Tutorial in Biostatistics: Competing Risks and Multi-State Models', *Statistics in Medicine* **26**, 2389–2430.

Solomon, S. D., Dobson, J., Pocock, S., Skali, H., McMurray, J. J., Granger, C. B., Tusuf, S., Swedberg, K., Young, J. B., Michelson, E. L. & Pfeffer, M. A. (2007), 'Influence of Nonfatal Hospitalization for Heart Failure on Subsequent Mortality in Patients with Chronic Heart Failure', *Circulation* **116**, 1482–1487.

Therneau, T. M. & Grambsch, P. M. (2000), *Modelling Survival Data: Extending the Cox Model*, Springer, New York, United States.

Wei, L. J. & Glidden, D. V. (1997), 'An Overview of Statistical Methods for Multiple Failure Time Data in Clinical Trials', *Statistics in Medicine* **16**, 833–839.

Wei, L. J., Lin, D. Y. & Weissfeld, L. (1989), 'Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions', *Journal of the American Statistical Association* **84**, 1065–1073.

Yu, C. M., Abraham, W. T., Bax, J. J., Chung, E. S., Fedewa, M., Ghio, S., Leclercq, C., Leon, A. R., Merlino, J., Nihoyannopoulos, P., Notabartolo, D., Sun, J. & Tavazzi, L. (2005), 'Predictors of Response to Cardiac Resynchronization Therapy (PROSPECT): study design', *American Heart Journal* **149**, 600–605.