Commentary
Competing risks models and time-dependent covariates

Adrian Barnett and Nick Graves

Institute of Health and Biomedical Innovation, Queensland University of Technology, 60 Musk Avenue, Kelvin Grove Urban Village, Kelvin Grove, Queensland 4059, Australia

Corresponding author: Adrian Barnett, a.barnett@qut.edu.au

Published: 11 April 2008
This article is online at http://ccforum.com/content/12/2/134

Critical Care 2008, **12:**134 (doi:10.1186/cc6840)

See related research by Wolkewitz *et al.*, http://ccforum.com/content/12/2/R44

Abstract

New statistical models for analysing survival data in an intensive care unit context have recently been developed. Two models that offer significant advantages over standard survival analyses are competing risks models and multistate models. Wolkewitz and colleagues used a competing risks model to examine survival times for nosocomial pneumonia and mortality. Their model was able to incorporate time-dependent covariates and so examine how risk factors that changed with time affected the chances of infection or death. We briefly explain how an alternative modelling technique (using logistic regression) can more fully exploit time-dependent covariates for this type of data.

In the present issue of *Critical Care* Wolkewitz and colleagues use competing risks models to examine risk factors for nosocomial pneumonia and mortality in an intensive care unit [1]. Competing risks models offer significant advantages over standard survival analysis [2]. In a standard survival analysis there is one event (for example, death) and one time (for example, days until death). Often we have a set of covariates and want to know which are most predictive of the event. In competing risks models the number of events can be greater than one. In the study by Wolkewitz and colleagues there were three competing risks: nosocomial pneumonia, death and discharge. Covariates can depend on the competing risk. A good example from the study by Wolkewitz and colleagues is elective surgery before admission, which increased the risk of nosocomial pneumonia but decreased the risk of death and discharge.

Competing risks models can incorporate time-dependent covariates using a Cox proportional hazards model. A time-dependent covariate is one that changes during the study period; for example, ventilation (yes/no). A time-independent covariate does not change; for example, sex. Time-dependent covariates can be richer than time-independent covariates because they offer the chance to examine the order of exposure and outcome [3]. The study by Wolkewitz and colleagues involved 10 binary time-dependent covariates, including nosocomial pneumonia.

To use time-dependent covariates the data need to be arranged in a nonstandard format [2,4]. A new row needs to be added each time a covariate changes. As an example of a competing risks model, consider the subject presented in Table 1. When this subject entered the intensive care unit their start time was set to 0. After 2 days they contracted a nosocomial infection, and so their period with this pattern of covariates was censored. On day 3 the subject was ventilated, and so a new row is added to include this covariate pattern. After being ventilated for 1 day the subject died. *Ventilated* and *nosocomial infection* are binary time-dependent covariates, whereas *sex* is a time-independent covariate. Data in this format can be analysed using a Cox proportional hazards model.

An alternative method to the competing risks model is a multistate model [2]. Using this multistate method, subjects move over time between a set of states. A three-state model using the present example is shown in Figure 1.

A survival analysis is then run for every transition (that is, every arrow in the diagram). For this multistate model the example data would be arranged as presented in Table 2. There is therefore one row per subject per transition. The big disadvantage of this arrangement is that time-dependent covariates can only be updated when a subject first enters a state. For the example data, therefore, the change in ventilation on day 3 has been lost.

An alternative arrangement is to exploit the equally spaced nature of the data and create a row for each day. Such an arrangement is presented in Table 3 for the example data. Using this format we can analyse the data using logistic regression (with *status* as the dependent variable) [4]. This type of logistic regression has been shown to be equivalent to time-dependent analysis using Cox models [5], but has some important advantages.

The first advantage is that lagged covariates can be added [3]. For example, the effect of a nosocomial infection on the

Table 1

Example of data for a competing risks model

Subject number	Sex	Ventilated	Nosocomial infection	Start (days)	Stop (days)	Status
1	Female	No	No	0	2	Censored
1	Female	No	Yes	2	3	Censored
1	Female	Yes	Yes	3	4	Dead

Table 2

Example of the data for a multistate model

Subject number	Sex	From	To	Ventilated	Start (days)	Stop (days)	Status
1	Female	ICU entry	Nosocomial infection	No	0	2	Uncensored
1	Female	ICU entry	Discharge/death	No	0	2	Censored
1	Female	Infected	Death	No	2	4	Uncensored

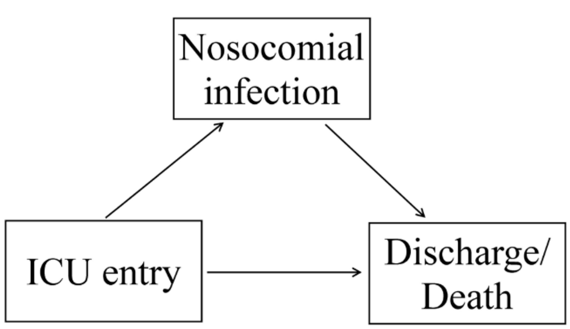
ICU, intensive care unit.

Table 3

Example of the data for a logistic regression model

Subject number	Day	Sex	Ventilated	Nosocomial infection	Days since infection	Status
1	0	Female	No	No	0	Alive
1	1	Female	No	No	0	Alive
1	2	Female	No	Yes	1	Alive
1	3	Female	Yes	Yes	2	Alive
1	4	Female	Yes	Yes	3	Dead

Figure 1



Example of a three-state model. ICU, intensive care unit.

likelihood of being discharged might be strongest on the actual day of infection, and then wane over time. In the present example, this could be modelled using *days since infection* as a covariate, rather than the binary indicator *nosocomial infection*.

Another advantage is that the proportional hazards assumption can be broken, either by allowing the effect of the covariate to change with time (as per the lagged covariate) or by allowing the intercept to vary with time (which alters the overall baseline risk).

The third advantage is that, using a mixed effects logistic regression, random effects can be added to help explain differences between subjects [3].

The main disadvantage of such logistic regression models is that they rely on equally spaced data (for example, days, hours), and are not applicable to continuous time results. In intensive care unit studies, however, minutes are rarely important and the data could safely be rounded to hours.

Time-dependent covariates are critically important in studies concerning lengths of stay in hospital. Covariates such as *nosocomial infection* occur at varying times, and a key question is estimating how much longer a patient can expect to stay in hospital if they become infected. Wolkewitz and colleagues have made a big step toward properly incorporating time-dependent covariates in this context using competing risks models [1]. The model we have suggested more fully exploits these important covariates and relaxes the assumptions concerning proportional hazards.

Competing interests

The authors declare that they have no competing interests.

References

1. Wolkewitz M, Vonberg R-P, Grundmann H, Beyersmann J, Gastmeier P, Baerwolff S, Geffers C, Behnke M, Rueden H, Schumacher M: **Risk factors for the development of nosocomial pneumonia and mortality on intensive care units: application of competing risks models**. *Critical Care* 2008, **12:**R44.
2. Putter H, Fiocco M, Geskus RB: **Tutorial in biostatistics: competing risks and multi-state models**. *Stat Med* 2007, **26:**2389-2430.
3. Diggle P, Heagerty P, Liang K-Y, Zeger S: *Analysis of Longitudinal Data*. 2nd edition. New York: Oxford University Press; 2002.
4. Hedeker D, Gibbons RD: *Longitudinal Data Analysis*. Hoboken, NJ: Wiley; 2006.
5. D'Agostino RB, Lee M-L, Belanger AJ, Cupples LA, Anderson K, Kannel WB: **Relation of pooled logistic regression to time dependent cox regression analysis: the Framingham heart study**. *Stat Med* 1990, **9:**1501-1515.