APPENDIX:
Nested case-control studies in cohorts with competing events

June 13, 2013

Characteristics of the full cohort

Suppose, one is interested in studying risk factors for nosocomial infections. Information about nosocomial infections including the time of occurrence are often collected within infection control programs. We further assume that patient data including admission and discharge date and status at time of discharge (dead or alive) are routinely collected. Due to limited resources, one decides to perform a nested case-control study. We used Spanish intensive care unit (ICU) data: two ICUs, 6567 admissions, 432 (6.58%) nosocomial infections, 762 died in ICU, 5363 discharged alive, 10 administratively censored. The data were collected within the network ENVIN-HELICS. The risk factor of interest is the APACHE (Acute Physiology And Chronic Health Evaluation) score. For illustrative purposes, we dichotomized the score at value 15, 4463 (2105) had a score$\leq$15 ($>$15). Among those patients with an APACHE score $\leq$15, 154 admissions (3.45%) acquired an infection during their stay in the ICU. In contrast, 278 admissions (13.2%) acquired an infection with an APACHE $>$15.

Since in the data used, exposure information is available for all admissions, we first estimate parameters using the full cohort, then compare the estimates obtained using the nested case control approach.

Event-specific hazards approach

Non-parametric estimation: full cohort

The standard approach to analyze survival data from such a full cohort is to study the event-specific hazards for nosocomial infection, discharge (alive) and death in ICU. A crude estimate of the infection hazard can be described in the discrete time setting as the number of individuals who experience the event divided by the number at risk at time t; a formal definition is given in[1,2]. Here, we used the Nelson-Aalen estimator[3] to calculate the cumulative event-specific hazards. The estimates for the events 'nosocomial infection', 'discharge (alive)' and 'death in ICU' are displayed in Figure 1, stratified for the two APACHE score categories. This shows that patients with a high score have an increased rate of infection (left panel). It also shows that these patients are associated with an decreased discharge and an increased death rate.

Semi-parametric estimation: full cohort

For each event, we used a proportional hazards model[3] to calculate the event-specific hazard ratios.

Results are displayed in table 1: the hazard ratio for infection is 1.56 (95% CI: 1.28-1.92), for discharge (alive) 0.35 (95% CI: 0.32-0.37) and for death 5.37 (95% CI: 4.46-6.46). If ones combines the two competing events to 'discharge (dead or alive)', the hazard ratio is 0.51 (95% CI: 0.48-0.54) since the discharge (alive) hazard is of a stronger magnitude than the death hazard (Figure 1). Given the hazard ratios for infection, one can calculate model-based cumulative hazards. They are per definition proportional and displayed in the left panel of Figure 2.

In such a event-specific analysis, competing events are technically coded as a censoring event[1]. This is, however, an informative censoring in the sense that probability estimates depend on all event-specific hazards[4].

Nested case-control approach

For the nested case-control approach we used incidence density sampling (after breaking ties). This procedure is displayed in Figure 1 of the main paper. For each infected case, controls must be disease free at the time of diagnosis of the case to which they are matched. In addition, those patients who were discharged prior to this time are not eligible as potential controls. For our data, we used an established SAS program[5]. For 432 (6.58%) admissions with nosocomial infec-

tions we matched 432 controls.

According to Lubin[6], the estimated odds ratios from the conditional logistic regression model approximate the event-specific HR for infection of the full cohort. In table 1, the hazard ratio for infection is estimated as 1.53 (95% CI: 1.16-2.03). The approximation works well, though the confidence limits are wider due to the reduced sample size.

Then, we use information from the whole cohort to calculate the cumulative hazards based on the proportional hazards model[7]. The required information is the number at risk for the event at the times of infection. These estimates are almost identical to those from the full cohort (Figure 2).

However, the competing events are ignored and hence the analysis is incomplete.

References

[1] Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol. 2009 Jul;170:244–256.

[2] Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med. 2007 May;26:2389–2430.

[3] Andersen P, Borgan O, Gill RD, Keiding N. Statistical models based on counting processes. New York: Springer Series in Statistics. Springer-Verlag Inc.; 1993.

[4] Beyersmann J, Allignol A, Schumacher M. Competing Risks and Multistate Models with R. Springer, New York; 2011.

[5] Richardson DB. An incidence density sampling program for nested case-control analyses. Occup Environ Med. 2004 Dec;61:e59.

[6] Lubin JH. Case-control methods in the presence of multiple failure times and competing risks. Biometrics. 1985 Mar;41:49–54.

[7] Langholz B. Use of Cohort Information in the Design and Analysis of Case-Control Studies. Scandinavian journal of statistics. 2007;34(1):120–136.

Table 1: Event-specific regression modeling using the full cohort and nested case-control 1:1 study for event 'infection'. Results of conditional logistic regression (averaged over 1000 runs). Hazard ratios in terms of event-specific hazards with 95% confidence intervals.

Event-specific regression modeling using the full cohort			
Apache score	Infection	Death	Discharge
'>15' versus '≤15'	1.56 (1.28-1.92)	5.37 (4.46-6.47)	0.35 (0.32-0.37)
Conditional logistic regression using the nested case-control 1:1 study (averaged over 1000 runs)			
Apache score	Infection	-	-
'>15' versus '≤15'	1.53 (1.16-2.03)	-	-

Figure 1: Cumulative hazards of the event of interest (infection) and competing events discharge/death with respect to Apache score categories. Non-parametric Nelsen-Aalen estimates using data from the full cohort.

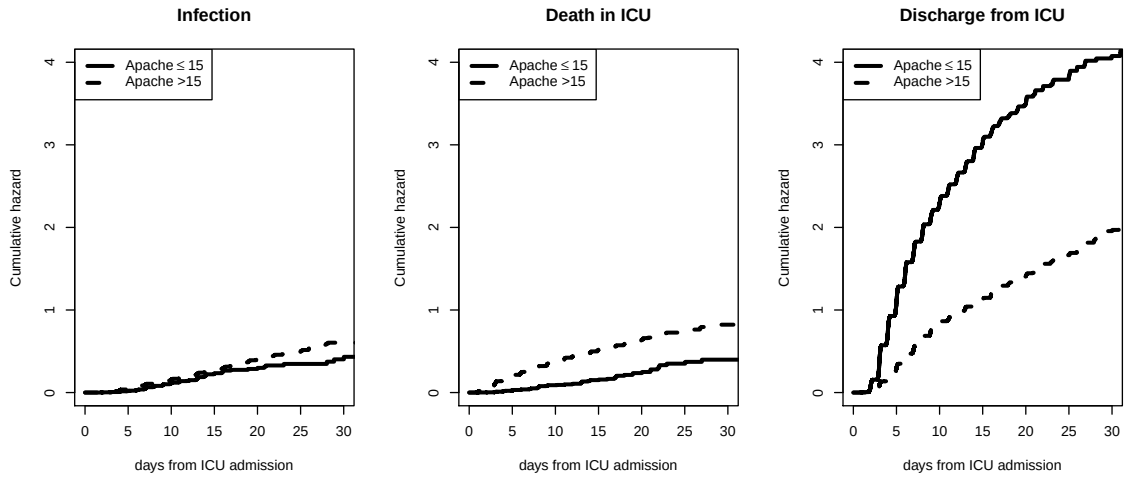

Figure 2: Cumulative hazards of the event of interest (infection) with respect to Apache score categories. Black lines: non-parametric Nelsen-Aalen estimates using data from the full cohort. Left and dark gray: model-based Cox using data from the full cohort. Left and light gray: derived from nested case-control data with using additional cohort information