Accepted Manuscript

A full competing risk analysis of hospital-acquired infections could easily be performed by a case-cohort approach

Martin Wolkewitz, Mercedes Palomar-Martinez, Pedro Olaechea-Astigarraga, Francisco Alvarez-Lerma, Martin Schumacher

Please cite this article as: Wolkewitz M, Palomar-Martinez M, Olaechea-Astigarraga P, Alvarez-Lerma F, Schumacher M, A full competing risk analysis of hospital-acquired infections could easily be performed by a case-cohort approach, *Journal of Clinical Epidemiology* (2015), doi: 10.1016/j.jclinepi.2015.11.011.

A full competing risk analysis of hospital-acquired infections could easily be performed by a case-cohort approach

Martin Wolkewitz, Mercedes Palomar-Martinez,

Pedro Olaechea-Astigarraga, Francisco Alvarez-Lerma and Martin Schumacher

Center for Medical Biometry and Medical Informatics, Institute for Medical Biometry and
Statistics, Medical Center University of Freiburg, Germany
Martin Wolkewitz
Martin Schumacher

Universitat Autonoma de Barcelona,
Hospital Vall bronzed, Barcelona, Spain
Mercedes Palomar-Martinez

Service of Intensive Care Medicine,
Hospital de Galdakao-Usansolo, Bizkaia, Spain
Pedro Olaechea-Astigarraga

Service of Intensive Care Medicine,
Parc de Salut Mar, Barcelona, Spain
Francisco Alvarez-Lerma

Correspondence to:
PD Dr. Martin Wolkewitz
Stefan-Meier-Str. 26, 79104 Freiburg, Germany
phone: +49/761/203-7702 , email: wolke@imbi.uni-freiburg.de

Abstract

Objective

We provide a case-cohort approach and show that a full competing risk analysis is feasible even in a reduced data set. Competing events for hospital-acquired infections are death or discharge from the hospital since they preclude the observation of such infections.

Study Design and Setting

Using surveillance data of 6568 patient-admissions (full cohort) from two Spanish intensive care units, we propose a case-cohort approach which uses only data from a random sample of the full cohort and all infected patients (the cases). We combine established methodology to study following measures: event-specific as well as subdistribution hazard ratios for all three events (infection, death, discharge), cumulative hazards as well as incidence functions by risk factor and also for all three events.

Results

Compared to the values from the full cohort, all measures are well approximated with the case-cohort design. For the event of interest (infection), event-specific as well as subdistribution hazards can be estimated with the full efficiency of the case-cohort design. So standard errors are only slightly increased whereas the precision of estimated hazards of the competing events are inflated according to the size of the sub-cohort.

Conclusion

The case-cohort design provides an appropriate sampling design for studying hospital-acquired infections in a reduced data set. Potential effects of risk factors on the competing events (death, discharge) can be evaluated.

2

keywords:

competing risk model, multi-state model, subdistribution, Cox proportional hazards model, Fine and Gray model, sampling designs, cumulative incidence function, nosocomial infections

---

What is new?

- The case-cohort approach allows to study the hazards as well as the cumulative risks of hospital-acquired infections and their competing events using a reduced data set.

- The occurrence of competing events (discharge, death) plays an important role in understanding potential effects of risk factors for hospital-acquired infections.

- In a hospital setting, almost all patients experience the event of interest or one of the competing events; there is almost no administrative censoring.

- The case-cohort approach provides more details about direct and indirect risk factor effects on nosocomial infections than the classical and frequently used case-control approach.

Introduction

Risk factor analyses of nosocomial, i.e. hospital-acquired infections (NIs) are complex and it is recommended to use extended survival models to account for the time-dependency of the data[1]. In a risk factor analysis, one is confronted with discharge from and death in the hospital as competing events for NI. Ignoring potential effects on these competing events can easily lead to incomplete reporting, incorrect risk estimates and potentially flawed interpretation[2,3]. The competing risks approach avoids this common pitfall but it usually needs the explicit chronological time structure which is traditionally given in a cohort data style. However, it is often required to use only a reduced data set from the full cohort if, for instance, information of potential risk factors is only feasible to collect in a *subsample*. Most applications have therefore used the classical nested case-control design in which controls were matched on time to the infected cases (incidence density sampling). This approach provides the hazard ratio of a risk factor of NI which corresponds to the event-specific hazard ratio in a competing risks setting.

It is known that reporting this event-specific hazard ratio alone is not enough to fully understand the effect of a potential risk factor[2,3] since the effect on the competing events of NI remains unknown. Furthermore, it is well known that in the presence of competing events, the event-specific hazard ratio does not compare cumulative risks[2,3]. In contrast to the event-specific hazard, the so-called subdistribution hazard is directly linked to the cumulative incidence function (CIF) which, in particular, estimates the risk of NIs. Even though the subdistribution hazard has itself no meaningful interpretation, it can be used to calculate the CIF due to this one-to-one relationship. Therefore, the event-specific hazard ratios for discharge or death without NI as well as the subdistribution hazard ratios of NI are also needed to provide a complete risk factor analysis[4].

In a previous paper[5], we studied a case-control approach to estimate subdistribution hazard ratios and CIFs. There we noted that the choice of controls is linked to the hazard of primary interest (event-specific or subdistribution hazard). Thus, before the control selection one needs to decide which hazard one aims to study. The re-use of controls seems to be a promising approach for studying other event-specific hazards[6,7] but there is so far no established methodology for studying the subdistribution hazard.

In contrast to the case-control design, the case-cohort design is known to be more flexible since it is possible to study other events and the time-dependent cohort data structure is still conserved after suitable weighting adaptations[8,9]. Pintilie et al.[10] applied the case-cohort design on a competing risks setting to estimate the risk of delayed cardiac toxicity among Hodgkin Lymphoma survivors. We will develop a similar approach for a hospital

4

infection setting where we will make use of the fact that we have almost no administrative censoring meaning that all patients experience one of the three events: NI, death or discharge without NI.

The aim of this paper is to provide the methodology to study an established competing risks model for hospital-acquired infections by using a reduced data set from a case-cohort design. This approach includes an evaluation of event-specific as well as subdistribution hazards; the latter are necessary to compare CIFs. Common pitfalls such as censoring bias are avoided and adequately addressed. We demonstrate that the methodology works well by means of a real cohort study in which the risk factor information is available from *all* cohort subjects. This full information makes it possible to compare all measures of interest from the case-cohort approach with the ones from the full cohort.

Methods

Competing risks models for nosocomial infections

The occurrence of NI highly depends on the time spent in the hospital/intensive care unit (ICU). Therefore, the time from admission to NI needs to be accounted for in the model. Most patients, however, were discharged from or die in the ICU without acquiring a NI. Thus, the occurrence of discharge/death usually precludes the observation of a NI. Competing risks models address this issue. An appropriate way to evaluate risk factors is to perform standard event-specific Cox regression models for each event (NI, death, discharge) separately by treating the other events technically as censored. This approach evaluates the way how risk factors are associated on each outcome event (NI, discharge and death) in terms of hazard ratios and can be considered as an etiological exploration. It provides information about factors which are directly associated on the daily risk of the event of interest (here: acquiring a NI). The problem with this etiological approach is that it is not interpretable in terms of cumulative risks since the cumulative risk of NI also depends on the discharge/death hazard[2] (a mathematical description is given in the supplementary material of this paper). Therefore, the hazard ratios cannot be interpreted as effects on the CIF (as it would be if there were no competing events).

One way to estimate the CIF is to use plug-in estimates from the cause-specific hazards (see supplement). However, with this method it is not possible to directly assess the effect of risk factors on the CIF. Therefore, Fine&Gray[11] proposed a subdistribution approach to study factors associated with the CIF using an adapted Cox regression model. Latouche et

5

al[12] emphasize that both approaches (event-specific as well as subdistribution hazard) are required to fully understand how factors are associated with NI. As we will see later in the results section, reporting only the hazard ratio for NI would be incomplete and potentially misleading.

We aim to approximate the following measures by solely using case-cohort data: event-specific as well as subdistribution hazard ratios for all three events, cumulative hazards as well as incidence functions by risk factor and all three events. This provides a complete picture of potential effects of a risk factor.

Hospital data

Full cohort
Following data are often routinely collected from all patients in a hospital: patient identi-fier, date of admission, date of discharge, vital status at the time of discharge. In addition, most infection surveillance systems provide information about the date and type of infec-tion and the causative pathogen. Once merged, the combination of these two data sets builds the skeleton of what we will consider as the full cohort. The time from admis-sion to potential censoring (needed to study subdistribution hazards) can be calculated by subtracting the calendar date of admission from the calendar date of study end.

In our practical example, we used Spanish intensive care unit (ICU) data: two ICUs, 6568 admissions, 432 (6.6%) nosocomial infections, 762 (11.7%) deaths in ICU without NI, 5363 (81.7%) discharges without NI, 10 administratively censored. The data were collected within the network ENVIN-HELICS and have previously be used for studying a nested case-control approach[5]. The risk factor of interest is the APACHE (Acute Phys-iology And Chronic Health Evaluation) score. For illustrative purposes, we dichotomized the score at value 15, 4463 (2105) had a score≤15 (>15). Among those patients with an APACHE score ≤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.

Case-cohort sampling
The case-cohort design contains of a random sub-cohort and all cases. In our setting, we selected a random sample of 500 patient-admissions (and 1000 patient-admissions in a second scenario) and combined them with the 432 infected patients. To account for the sampling variability, we performed 100 runs following the procedure above and averaged the coefficients and confidence intervals from the corresponding regression models. The cumulative hazards and incidence functions were calculated using ten runs of a random

sample of 500 patient-admissions, respectively.

Case-cohort approach for event-specific hazards

We define S - subcohort, $\bar{S}$ - non-subcohort, C - cases (nosocomial infections), $\bar{C}$ - non-cases, N - number of admissions in full cohort, n- number of admissions in subcohort, t - follow-up time, i - ith admission.

Case-cohort data for NI hazard
Due to the case-cohort design, infected patients are overrepresented and it is necessary to perform a weighted Cox model for event NI. According to Barlow et al[9], we used the inverse of the sampling fraction to weight the individuals of the case-cohort.

We define following weightings for $\lambda_{01}(t|X)$

$$w_i(t) = \begin{cases} 0 & \text{if } i \; \epsilon \; \bar{S} \cap C \text{ and } t \leq \text{time of infection} - \epsilon \\ 1 & \text{if } i \; \epsilon \; C \text{ and time of infection} - \epsilon \leq t \leq \text{time of infection} \\ N/n & \text{if } i \; \epsilon \; S \cap C \text{ and } t \leq \text{time of infection} - \epsilon \\ N/n & \text{if } i \; \epsilon \; S \cap \bar{C} \text{ and } t \leq \text{time of discharge / death / censoring} \end{cases}$$

The first line ($w_i(t) = 0$) ensures that the admission-days before infection for infected cases outside the subcohort do not contribute to the risk set. As already emphasized by Barlow et al[9], a case outside the subcohort is considered to be not at risk until just before failure and is not included in earlier risk sets. The second line ($w_i(t) = 1$) ensures that all infected cases contribute to the model with weight equals one; ϵ chosen to be very small. The third ($w_i(t) = N/n$) represents the weights according to the inverse of the sampling fraction for the observations from the subcohort cases up to the time of infection (here: 6568/500). The same weights are used for the uninfected subcohort observations (fourth line). This approach is in line with the definition by Barlow et al[9]. As in the event-specific analysis of the full data, the competing events are technically treated as censored and modeled separately. The robust variance was used[9]. Based on the weighted Cox model, we calculated the cumulative hazards of NI for both APACHE score groups.

Sub-cohort data for death/discharge hazard
To study the event-specific hazards of the competing events, we solely use the sub-cohort data and fit two standard (unweighted) Cox models for each outcome (death without NI, discharge without NI); infected cases within the sub-cohort were censored at the time

of infection. In contrast to the model-based cumulative NI hazards, we can now use the non-parametrical Nelson-Aalen estimator for calculating the cumulative hazards of death/discharge without NI, separately for both APACHE groups[13].

Case-cohort approach for cumulative incidence functions

Case-cohort data for cumulative incidence function of NI
The procedure to study the CIF in a case-cohort design is exactly the same as for the event-specific hazards except that we replaced the times of the corresponding competing events. Hence, for studying the subdistribution of NI, the weightings are the similar as the one for the event-specific approach by replacing the time of discharge / death with the time of potential censoring. Then, based on the weighted subdistribution model, we calculated the model-based cumulative hazards of NI for both APACHE score groups.

$$w_i(t) = \begin{cases} 0 & \text{if } i \, \epsilon \, \bar{S} \cap C \text{ and } t \leq \text{time of infection} - \epsilon \\ 1 & \text{if } i \, \epsilon \, C \text{ and time of infection} - \epsilon \leq t \leq \text{time of infection} \\ N/n & \text{if } i \, \epsilon \, S \cap C \text{ and } t \leq \text{time of infection} - \epsilon \\ N/n & \text{if } i \, \epsilon \, S \cap \bar{C} \text{ and } t \leq \text{time of potential censoring / censoring} \end{cases}$$

Sub-cohort data for cumulative incidence function of death/discharge
Finally, for the events death/discharge without NI, we use solely the sub-cohort data and fit two standard Fine&Gray models to get subdistribution hazard ratios for each outcome; thus, the corresponding competing event times are replaced by the time to potential censoring. In addition, we also use solely the sub-cohort data to calculate the non-parametric Aalen-Johansen estimator for the CIFs[13].

Multiple risk factors and control for confounding and interactions

In presence of multiple risk factors of interest, we recommend to plot cumulative hazards and incidence functions for each (categorical) risk factor univariately in a first step. Then, multiple factors can be included in the corresponding regression models (standard/weighted Cox models with or without modified censoring times) for a multivariate analysis and controlling for confounding and interactions. Adjusted (subdistribution) hazards ratio can be reported beside the unadjusted ones (see[4] as an example).

8

Results

In table 1, the event-specific hazard ratios from the case-cohort approach are compared to the 'gold-standard' values from the full cohort. The values from the full cohort are interpreted as follows. Patients with an APACHE score ≥15 are associated with a higher hazard to acquire a NI (HR=1.56) but also with an increased hazard to die (HR=5.37), furthermore those patients require longer care and stay longer at ICU (HR=0.35). These effects are also displayed in the cumulative hazards plots (left panels in figure). Since the discharge hazard is the strongest hazard, the latter effect has a greater impact on the CIFs of NI: patients with APACHE score ≥15 stay longer at ICU and have therefore an additional risk of acquiring NIs. In other words, in addition to the increased daily risk of acquiring a NI (HR=1.56), the extended length of stay make patients with higher APACHE score acquire more NIs. Both effects are summarized in the CIFs (right panels in figure). The effect with regard to the CIF of NI (and death) is therefore even more pronounced for patients with APACHE score ≥15 whereas less for discharge. The subdistribution hazard ratios (sHR) describe the effects which can be seen in the CIFs and are sHR=4.02 (NI), 10.6 (death) and 0.3 (discharge).

The estimated hazard ratios from the case-cohort design approximate these values quite well even though for the death outcome the approximation is not optimal (table 1 and 2). This is due to the fact that (from the statistical point) only 11.6% die at ICU. The corresponding cumulative hazards and incidence functions from the case-cohort and sub-cohort are very close to those of the full cohort (figure).

Discussion

In this article, we proposed a case-cohort approach for studying the occurrence of hospital-acquired infections. We demonstrated that a case-cohort sampling is appropriate to study two complex competing risks models: the event-specific as well as the subdistribution hazard approach. It is a special feature with hospital data that one usually has almost complete data meaning almost no administratively censoring. Thus, the data solely from the sub-cohort might have enough events (discharges, deaths) to fit robust models.

In our example, the number of death events were sufficient for the statistical analysis. But we note that a distinction between death and discharge without NI is actually not required for estimating the cumulative risk of NI. In other words, a combination of these

9

two competing events into one event 'discharge (dead or alive)' would neither effect the cumulative hazard nor the CIF of NI. The hazard of the combined competing event is just the sum of both hazards and the formula for the CIF does not change (see supplement). However, it is often of medical interest to distinguish these two opposite clinical outcomes. If the number of deaths in the sub-cohort is too small for a multivariate Cox analysis (less than 10 deaths per risk factor), one could extend the approach to a 'case-case-cohort' by including also all deaths from the full cohort. The analysis is straightforward.

The question, whether case-cohort or nested case-control designs are more efficient in terms of variance estimates, depends mainly on the magnitude of censoring and staggered entries[14]. Since we use exactly the same data for a nested case-control design[5], we are able to compare both designs for this specific cohort. We can see that the case-cohort is less efficient than the classical nested case-control approach when calculating the event-specific hazard ratio for NI (the hazard ratio for a 1:1 nested case-control approach was 1.53 (95%-CI: 1.16-2.03); see[5]). The reason is the strong discharge hazard early after admission which are technically handled as censored when calculating the NI hazard. Interestingly, when comparing the subdistribution hazard ratios of both approaches[5] (the subdistribution hazard ratio for a 1:1 nested case-control approach was 4.03 (95%-CI: 2.94-5.54); see[5]), the case-cohort design produce smaller confidence intervals since patients who experience a competing event remain technically in the risk set, thus we have only the low number of administratively censored observations. This comparison is very crude since the number of distinct patients of both approaches are not the same and we refer to Ohneberg et al. for a deeper comparison between case-control and case-cohort design using these data[15].

The proposed case-cohort approach can be extended to study a multi-state model if one includes outcome data after NI (time from NI until death or discharge) of all infected patients. Then, one is able to study the impact of NI on clinical outcomes such as mortality or length of hospital stay[16].

Funding

References

[1] Schumacher M, Allignol A, Beyersmann J, Binder N, Wolkewitz M. Hospital-acquired infections–appropriate statistical treatment is urgently needed! Int J Epidemiol. 2013 Oct;42(5):1502–1508.

[2] Wolkewitz M, Cooper BS, Bonten MJ, Barnett AG, Schumacher M. Interpreting and comparing risks in the presence of competing events. BMJ. 2014;349:g5060.

[3] Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? Stat Med. 2012 May;31(11-12):1089–1097.

[4] Wolkewitz M, Cooper B, Palomar-Martinez M, Alvarez-Lerma F, Olaechea-Astigarraga P, Barnett A, et al. Multilevel competing risk models to evaluate the risk of nosocomial infection. Crit Care. 2014 Apr;18(2):R64.

[5] Wolkewitz M, Cooper BS, Palomar-Martinez M, Olaechea-Astigarraga P, Alvarez-Lerma F, Schumacher M. Nested case-control studies in cohorts with competing events. Epidemiology. 2014 Jan;25(1):122–125.

[6] Samuelsen SO. A pseudolikelihood approach to analysis of nested case-control studies. Biometrika. 1997;84(2):379–394.

[7] Støer NC, Samuelsen SO. Comparison of estimators in nested case-control studies with multiple outcomes. Lifetime Data Anal. 2012 Jul;18(3):261–283.

[8] Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika. 1986;73(1):1.

[9] Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case-cohort designs. J Clin Epidemiol. 1999 Dec;52:1165–1172.

[10] Pintilie M, Bai Y, Yun L, Hodgson DC. The analysis of case cohort design in the presence of competing risks with application to estimate the risk of delayed cardiac toxicity among Hodgkin lymphoma survivors. Stat Med. 2010 Nov;29(27):2802–2810.

[11] Fine J, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc. 1999;94(446):496–509.

[12] 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 Jun;66(6):648–653.

11

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

[14] Langholz B, Thomas DC. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. Am J Epidemiol. 1990 Jan;131(1):169–176.

[15] Ohneberg K, Wolkewitz M, Beyersmann J, Palomar-Martinez P, Olaechea-Astigarraga P, Alvarez-Lerma F, et al. Analysis of Clinical Cohort Data Using Nested Case-control and Case-cohort Sampling Designs. Methods of Information in Medicine. 2015;.

[16] Wolkewitz M, Vonberg RP, Grundmann H, et al. Risk factors for the development of nosocomial pneumonia and mortality on intensive care units: application of competing risks models. Crit Care. 2008 Apr;12:R44.

Figure 1: Cumulative hazards (left figures) and cumulative incidence functions (right figures) for all events, separately for APACHE score groups. The black curves display the Nelson-Aalen (cumulative hazards) and the Aalen-Johansen estimates (cumulative incidence functions) using the full cohort. The gray curves display the corresponding estimates from the case-cohort/sub-cohort using ten sub-cohort samples of size n=1000. Note that the 'case-cohort' estimates are derived from the adapted Fine&Gray model whereas the 'sub-cohort' are non-parametric estimates (Nelson-Aalen / Aalen-Johansen).

Table 1: Event-specific hazard ratios with 95% confidence intervals in brackets. When using the case-cohort / sub-cohort designs, we considered two sub-cohort sizes (n=500 and n=1000); we performed 100 runs and displayed the averaged estimates.

Event-specific regression modeling using the full cohort			
Apache score	Infection	Death w/o infection	Discharge w/o infection
'>15' versus '≤15'	1.56 (1.28-1.92)	5.37 (4.46-6.47)	0.35 (0.32-0.37)
Event-specific regression modeling using the case-cohort / sub-cohort (size of sub-cohort=500)			
Apache score	Infection	Death w/o infection	Discharge w/o infection
	(via case-cohort)	(via sub-cohort)	(via sub-cohort)
'>15' versus '≤15'	1.57 (1.06-2.31)	5.58 (2.79-11.14)	0.34 (0.26-0.43)
Event-specific regression modeling using the case-cohort / sub-cohort (size of sub-cohort=1000)			
Apache score	Infection	Death w/o infection	Discharge w/o infection
	(via case-cohort)	(via sub-cohort)	(via sub-cohort)
'>15' versus '≤15'	1.57 (1.16-2.14)	5.58 (3.43-9.09)	0.35 (0.29-0.41)

14

Table 2: Subdistribution hazard ratios with 95% confidence intervals in brackets. When using the case-cohort / sub-cohort designs, we considered two sub-cohort sizes (n=500 and n=1000); we performed 100 runs and displayed the averaged estimates.

Fine&Gray regression modeling using the full cohort			
Apache score	Infection	Death w/o infection	Discharge w/o infection
'>15' versus '≤15'	4.02 (3.30-4.89)	10.62 (8.85-12.74)	0.31 (0.29-0.33)
Fine&Gray regression modeling using the case-cohort / sub-cohort (size of sub-cohort=500)			
Apache score	Infection	Death w/o infection	Discharge w/o infection
	(via case-cohort)	(via sub-cohort)	(via sub-cohort)
'>15' versus '≤15'	4.00 (3.06-5.24)	11.03 (5.6-21.74)	0.30 (0.24-0.39)
Fine&Gray regression modeling using the case-cohort / sub-cohort (size of sub-cohort=1000)			
Apache score	Infection	Death w/o infection	Discharge w/o infection
	(via case-cohort)	(via sub-cohort)	(via sub-cohort)
'>15' versus '≤15'	3.97 (3.15-5.01)	10.92 (6.77-17.63)	0.31 (0.26-0.37)