Annals of Epidemiology xxx (2018) 1–6

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

Methodological challenges in using point-prevalence versus cohort data in risk factor analyses of nosocomial infections

Martin Wolkewitz, PhD [a,*], Micha Mandel, PhD [b], Mercedes Palomar-Martinez, PhD, MD [c], Francisco Alvarez-Lerma, PhD, MD [d], Pedro Olaechea-Astigarraga, MD [e], Martin Schumacher, PhD [a]

[a] Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, Germany
[b] Department of Statistics, The Hebrew University of Jerusalem, Jerusalem, Israel
[c] Hospital Universitari Arnau de Vilanova, Universitat Autonoma de Barcelona, Lleida, Spain
[d] Service of Intensive Care Medicine, Parc de Salut Mar, Barcelona, Spain
[e] Service of Intensive Care Medicine, Hospital de Galdakao-Usansolo, Bizkaia, Spain

ARTICLE INFO

Article history:
Received 30 October 2017
Accepted 26 March 2018
Available online xxx

Keywords:
Competing events
Cohort study
Multi-state models
Study design

ABSTRACT

Purpose: To explore the impact of length-biased sampling on the evaluation of risk factors of nosocomial infections (NIs) in point-prevalence studies.
Methods: We used cohort data with full information including the exact date of the NI and mimicked an artificial 1-day prevalence study by picking a sample from this cohort study. Based on the cohort data, we studied the underlying multistate model which accounts for NI as an intermediate and discharge/death as competing events. Simple formulas are derived to display relationships between risk, hazard, and prevalence odds ratios.
Results: Due to length-biased sampling, long stay and thus sicker patients are more likely to be sampled. In addition, patients with NIs usually stay longer in hospital. We explored mechanisms that are—due to the design—hidden in prevalence data. In our example, we showed that prevalence odds ratios were usually less pronounced than risk odds ratios but more pronounced than hazard ratios.
Conclusions: Thus, to avoid misinterpretation, knowledge of the mechanisms from the underlying multistate model is essential for the interpretation of risk factors derived from point-prevalence data.

Introduction

Point-prevalence studies about hospital-acquired or nosocomial infections (NIs) are popular as they are much easier to perform than cohort studies. In such studies, the prevalence of NI is often interpreted as the risk, and prevalence data were often used for risk factor analyses of NI [1–7]. But, due to the design, one has to keep in mind that long-stay patients are more likely to be sampled in a prevalence study because a patient who stays 10 days has the double chance to get sampled than a patient who stays 5 days (length-biased sampling) [8]. Thus, long stay and therefore sicker patients are over-represented in prevalence studies due to this sampling selection. It is well known that long-stay patients are at higher risk for NI than short stayers as length of hospital stay is a risk factor for NI, especially for NI caused by resistant pathogens. Furthermore, patients infected with NI require additional care and therefore stay longer in hospital [9].

Already in the 80s, Freeman and Hutchison [10] as well as Rhame and Sudderth [11] described the relationships between prevalence, incidence, and duration of NIs. We show that their formula are still relevant but can be rewritten in a modern multistate frame work that facilitates interpretation and improves understanding of the relations between important epidemiological measures.

Different association measures are used in risk factor analyses: risk-based measures such as risk or subdistribution hazard ratios (HRs) and rate-based measures such as hazard or rate ratios. Because of competing events (discharge and death), these two

Conflicts of interest: None.
Authors' contributions: All authors have participated in (a) conception and design or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

* Corresponding author: Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center–University of Freiburg, Stefan-Meier-Str. 26, 79104 Freiburg, Germany. Tel.: +49-761-203-7702; fax: +49-761-203-7700.
E-mail address: wolke@imbi.uni-freiburg.de (M. Wolkewitz).

https://doi.org/10.1016/j.annepidem.2018.03.017
1047-2797/

measures often differ [12–15]. For a complete interpretation of risk factor association with infection, it is necessary to understand how the prevalence odds (PO) ratio as a further association measure is related to rate- or risk-based measures.

In this article, we study and discuss the question how length-biased sampling as well as the prolonged stay of NI patients do influence prevalence studies. In particular, we investigate the impact on the prevalence and PO ratio and discuss the relationships between three association measures (ratio of prevalences, risks, or hazards). To do this, we make use of real cohort-style data of Spanish intensive care units' patients with full information on the exact date of NI, and we mimic a 1-day prevalence study by picking a sample from that cohort study. Thus, we make use of the full knowledge of the underlying cohort, which is usually unknown in prevalence studies. In addition, we show in a simplified theoretical setting that Rhame-Sudderth's formula can be considered within a multistate framework. Using the multistate approach, we derived simple formulas to display the relationships between prevalence, risk, and HR of NI.

Methods

We used a multicenter data base from the Spanish surveillance network hospitals in Europe link for infection control through surveillance (HELICS)-ENVIN (http://hws.vhebron.net/envin-helics/), embedded in the HELICS project [16]. This research project was approved by the Ethics committee of University Medical Center Freiburg, Germany.

Data

In this analysis, we used 47,564 individual intensive-care-unit (ICU) patients collected in April-May-June in years 2006–2010. Each admission, during April-May-June each year, was followed until discharge or death. This is our cohort style database that we will consider as the underlying population. To mimic a point-prevalence study, we took five random prevalence dates in May for each year. We sampled 4302 patients for our artificial prevalence study that is the one merged from the prevalence days in the years 2006–2010. We considered only covariates measured on admission to ensure that risk factor precedes NI. We defined NI as infections that occurred during ICU stay and 48 hours after admission. In line with previous considerations [11], the prevalent cases are those patients with an active or cured NI at the prevalence study day.

Underlying multistate model

Figure 1 shows the multistate model that is usually hidden in a prevalence study. We assume the steady state condition that is defined by the following: (i) patients enter the hospitals at times governed by a stationary homogeneous Poisson point processes with admission intensity β; (ii) length of hospitalization and time to infection are independent of the date of entry to hospital; and (iii) patients' length of hospitalization and infection are independent of those of other patients and are independent of the Poisson process and of the sampling date. Each arrow represents a potential transition between the states 0 (admission [free of NI]), 1 (NI), and 2 (discharge/death). The hazards λ_{01}, λ_{02}, and λ_{12} can be estimated and studied with the time-dependent data from the full underlying cohort. Note that, the constant hazards assumption is here just for simplification purposes and is not required for applying multistate models on cohort data.

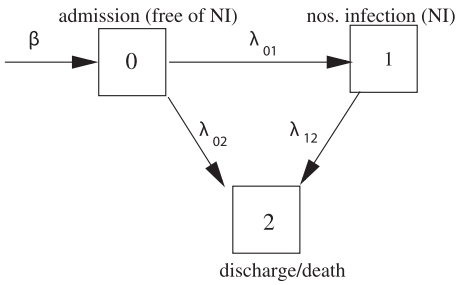

Fig. 1. Underlying multistate model that is usually hidden in a prevalence study. It is fed by a stationary homogeneous Poisson process with admission intensity β. The hazards λ_{ij} from state i–j are assumed to be time constant.

Statistical analysis

We discuss the following quantities for NIs: first, the constant hazard rate of NI (incidence density), which is defined as the daily probability to acquire a NI during hospital stay; second, the risk of NI (incidence proportion), which is the cumulative probability to acquire a NI during hospital stay; third, the risk odds of NI, which are the odds to acquire a NI during hospital stay; and fourth, the PO of NI, which are the odds to have a NI at a specific point in calendar time. The following risk factors were considered: Acute Physiologic Assessment and Chronic Health Evaluation (APACHE) II score (categorized in 0–10 [reference], 11–20, 21–30, and >31), diagnosis (cardiovascular [reference], respiratory, gastrointestinal, central nervous system, and other), antibiotic treatment 48 hours before and/or after ICU admission (yes vs. no), trauma (yes vs. no), and calendar year (2006 [reference], 2007, 2008, 2009, and 2010). In this article, we used general risk factors. However, for specific NIs, it is necessary to include specific risk factors, in particular the corresponding use of device such as mechanical ventilation for pneumonia, central venous catheter for blood stream infections, or indwelling catheter for urinary tract infections. To calculate HRs, Cox proportional hazards regression model were performed separately for the following transitions of the multistate model (Fig. 1): admission → NI, admission → discharge/death, NI → discharge/death. A Cox model was fitted with NI as a time-dependent covariate for outcome “discharge/death”. All Cox regression models were stratified by ICU. Generalized linear mixed (logistic) regression models were applied to calculate prevalence and risk odds ratios, respectively; ICUs entered as random effects in the model.

Results

Theoretical considerations

Rhame and Sudderth [11] derived a formula to display the relationship between Prevalence (P) and Incidence proportion (I) or risk of NI: $P = I \times LN - INT/LA$ where LA—average length of stay of all patients, LN—average length of stay of patients who acquire one or more NIs, and INT—average interval between admission and onset of the first NI for those patients who acquire one or more NIs.

These quantities can be written in terms of constant hazards from the multistate model: $I = \lambda_{01}/\lambda_{01} + \lambda_{02}$, $LN - INT = 1/\lambda_{12}$, and $LA = 1/\lambda_{01} + \lambda_{02} + \lambda_{01}/\lambda_{01} + \lambda_{02} \times 1/\lambda_{12}$.

More generally and not restricted to NIs, Keiding [17] proposed following relationship for the PO in a multistate frame work (Fig. 1): $PO = \lambda_{01}/\lambda_{12}$. Both approaches are mathematically

Please cite this article in press as: Wolkewitz M, et al., Methodological challenges in using point-prevalence versus cohort data in risk factor analyses of nosocomial infections, Annals of Epidemiology (2018), https://doi.org/10.1016/j.annepidem.2018.03.017