Clinical Nutrition xxx (2015) 1–6

Contents lists available at ScienceDirect

Clinical Nutrition

journal homepage: http://www.elsevier.com/locate/clnu

Original article

Length bias correction in one-day cross-sectional assessments – The nutritionDay study

Sophie Frantal [a], Elisabeth Pernicka [a], Michael Hiesmayr [b], Karin Schindler [c], Peter Bauer [a,*]

[a] Section for Medical Statistics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University Vienna, A-1090 Vienna, Austria
[b] Division Cardiac-, Thoracic-, Vascular Anaesthesia and Intensive Care, Medical University Vienna, A-1090 Vienna, Austria
[c] Medical Clinic III, Division Endocrinology, Medical University Vienna, A-1090 Vienna, Austria

ARTICLE INFO

Article history:
Received 11 March 2014
Accepted 30 March 2015

Keywords:
Cross-sectional study
Length bias
Weighting
Validation
NutritionDay

SUMMARY

Background & aims: A major problem occurring in cross-sectional studies is sampling bias. Length of hospital stay (LOS) differs strongly between patients and causes a length bias as patients with longer LOS are more likely to be included and are therefore overrepresented in this type of study. To adjust for the length bias higher weights are allocated to patients with shorter LOS. We determined the effect of length-bias adjustment in two independent populations.
Methods: Length-bias correction is applied to the data of the nutritionDay project, a one-day multinational cross-sectional audit capturing data on disease and nutrition of patients admitted to hospital wards with right-censoring after 30 days follow-up. We applied the weighting method for estimating the distribution function of patient baseline variables based on the method of non-parametric maximum likelihood. Results are validated using data from all patients admitted to the General Hospital of Vienna between 2005 and 2009, where the distribution of LOS can be assumed to be known. Additionally, a simplified calculation scheme for estimating the adjusted distribution function of LOS is demonstrated on a small patient example.
Results and conclusion: The crude median (lower quartile; upper quartile) LOS in the cross-sectional sample was 14 (8; 24) and decreased to 7 (4; 12) when adjusted. Hence, adjustment for length bias in cross-sectional studies is essential to get appropriate estimates.

1. Introduction – the problem of length bias

Cross-sectional data are frequently used to determine prevalence of diseases, risk factors and attitudes [1–3] in large samples. Cross-sectional data acquisition has also been used in hospitalised patients for such diverse conditions as malnutrition [4,5] or nosocomial infections [6,7]. Often the aim of these studies is to determine the burden of disease or risk factors on outcomes such as mortality, length of stay and costs for the hospital, health care system or community.

If every patient who is in hospital at the time of a cross-sectional study is included, patients with longer LOS have a higher probability to be included in the study. This leads to a selection bias called "length-bias" [8] with an over-representation of patients more severely ill, older or malnourished, having more comorbidities, necessitating more complex repeated interventions or interventions that limit mobility and autonomy such as surgical interventions. Hence, the sampled patients may not well represent the population of hospitalized patients admitted to a hospital over a given period of time. If the patients are only followed up by a predefined time period τ after the cross-sectional sampling also the issue of right censoring of LOS has to be taken into account [9–12].

This effect can easily be seen: In the left panel of Fig. 1 there is a general Lexis diagram of all patients staying in a hospital during a certain time period, while the Lexis diagram in the right panel of Fig. 1 only shows the 5 patients that would actually be included in a cross-sectional study. These are the patients where the vertical line at day 0 crosses their trajectory. Predominantly patients with short LOS do not enter the cross-sectional sample.

* Corresponding author. Medical University of Vienna, Institute for Medical Statistics, Spitalgasse 23, A-1090 Vienna, Austria. Tel.: +43 664 2835553; fax: +43 0 40400 74770.
E-mail address: peter.bauer@meduniwien.ac.at (P. Bauer).

http://dx.doi.org/10.1016/j.clnu.2015.03.019
0261-5614/

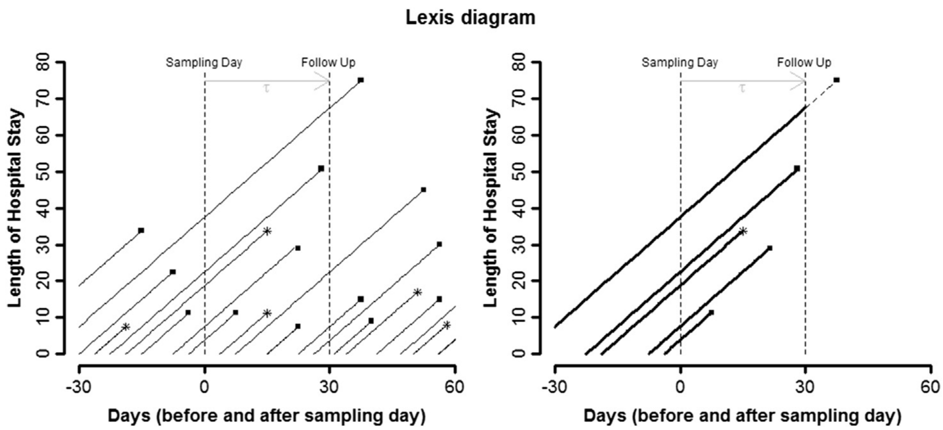

Fig. 1. shows two Lexis diagrams of a small random sample of patients (n = 20) around the day of the cross-sectional investigation. While the left panel shows all patients staying in hospital, the right panel only show those patients who are then actually included in the cross-sectional study. Note that in this example the fix follow-up period ends and hence patients are censored at 30 days after sampling day. X-axis: calendar time around sampling day (x = 0). Y-axis: length of hospital stay Y_i.

The aim of the present study is to evaluate the effect on estimates of median length of stay in hospital when adjusting for length-bias and right censoring [4,9,11,12]. To that end individual weights depending on the length of stay are given to each patient. The method of weighting is applied to the cross-sectional nutritionDay database of hospitalised patients [4] for estimating the distribution of length of stay. We validated the method of adjustment by creating artificial cross-sectional samples from a complete outcome data set of a large university hospital. The estimated distribution functions for LOS and age as well as the gender prevalence were compared to the crude data.

2. Methods – data samples

2.1. The cross-sectional nutritionDay sample

The nutritionDay is an international project with the aim to improve knowledge and awareness of malnutrition in hospitals. It is a one-day annual cross-sectional study, repeated 8 times between 2006 and 2012 (2010 two nutritionDays were done) and hence, shows a "snap-shot" of the actual situation in self-selected hospitals all over the world.

The study has been designed to assess nutritional and clinical risk factors of patients in hospital as well as their outcome within the next month. Online questionnaires in different languages were used to collect ward specific and patient specific variables. The latter are self-reported by the patients giving information on food intake, disease related factors and mobility, as well as by caregivers giving information on hospital admission and demographics. Patients received help when needed.

Outcome evaluation is done after a fix follow-up period of 30 days after nutritionDay. The number of days already in hospital at the nutritionDay was asked as well as the date of outcome and type of outcome (still in hospital, transferred, rehabilitation, discharged, readmitted or deceased).

2.2. Validation samples of hospitalised patients with complete follow-up

To check whether the observed results were reliable and consistent over time, they were validated based on an available sample comprising patients admitted to units of the General Hospital of Vienna in the years 2005, 2007 to 2009 (until august). Only the first hospital stay in this time period of patients with complete follow-up was considered. Patients with missing LOS status and children are not included. This results in a total sample of about 145000 patients. As outcome is recorded for all patients there is no censoring. As a full sample is considered no length bias occurs. Hence, due to the large sample size we have a very precise estimate of the true LOS distribution (assumed to be stable over the years).

Out of this overall sample 15 cross-sectional samples were drawn. Within each quarter (each 3-month-period) of the observed years one day was randomly chosen. Hence, the achieved days cover different seasons and week days. Data are then artificially censored after 30 days.

3. Results

3.1. Length bias in the nutritionDay study

In the past seven years slightly more than 100.000 patients from about 50 countries, 1000 hospitals and 3500 wards have been part of the nutritionDay study. Patients with missing LOS status and children (18 years or younger) are not included in the following results.

In the unadjusted crude distribution of the length of stay when patients still in hospital 30 days after the day of survey are excluded (n = 71093), the median (lower quartile; upper quartile) LOS was 14 (8; 24) days. When in the total sample (n = 80011) using the censored LOS values after 30 days as observed LOS values, the corresponding values are 16 (8; 31), creating even a larger bias. In the sample, adjusted for the length bias and censored at 30 days after the sampling day as done in the nutritionDay project [11,12], the LOS was 7 (4; 12) days (compare Fig. 2). Point–wise asymptotic 95% confidence intervals for the estimated cumulative distribution function as given in [12] have been calculated. Due to the very large sample size the variability of the estimates is very small (at the median LOS the width of the confidence interval is below 0.002) and hence, confidence intervals are not shown.

Adjusting for length bias in the distribution function by the non-parametric maximum likelihood estimation procedure (see Appendix) reduced the median LOS by a half. Adjustment can be interpreted as a change in the point of view: while the crude data belong to the view of the caregiver describing the population lying in a hospital at a specific calendar day, the adjusted data belong to the patient population admitted to the hospital.

In order to check the assumption of stationary length biasing (assuming that the truncation time follows a uniform distribution)

Please cite this article in press as: Frantal S, et al., Length bias correction in one-day cross-sectional assessments – The nutritionDay study, Clinical Nutrition (2015), http://dx.doi.org/10.1016/j.clnu.2015.03.019