



Assessing Noninferiority in Treatment Trials for Severe Infectious Diseases: an Extension to the Entire Follow-Up Period Using a Cure-Death Multistate Model

Harriet Sommer,^a Tobias Bluhmki,^b Jan Beyersmann,^b Martin Schumacher,^a on behalf of the COMBACTE-NET and COMBACTE-MAGNET consortium

^aInstitute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center-University of Freiburg, Freiburg, Germany

^bInstitute of Statistics, Ulm University, Ulm, Germany

ABSTRACT In current and former clinical trials for the development of antibacterial drugs, various primary endpoints have been used, and treatment effects are evaluated mostly in noninferiority analyses at the end of follow-up, which varies between studies. A more convincing and highly patient-relevant statement would be a noninferiority assessment over the entire follow-up period with cure and death as coprimary endpoints, while preserving the desired alpha level for statistical testing. To account for the time-dynamic pattern of cure and death, we apply a cure-death multistate model. The endpoint of interest is “get cured and stay alive over time.” Noninferiority between treatments over the entire follow-up period is studied by means of one-sided confidence bands provided by a flexible resampling technique. We illustrate the technique by applying it to a recently published study and establish noninferiority in being cured and alive over a time frame of interest for the entire population, patients with hospital-acquired pneumonia, but not for the subset of patients with ventilator-associated pneumonia. Our analysis improves the original results in the sense that our endpoint is more patient benefiting, a stronger noninferiority statement is demonstrated, and the time dependency of cure and death, competing events, and different follow-up times is captured. Multistate methodology combined with confidence bands adds a valuable statistical tool for clinical trials in the context of infection control. The framework is not restricted to the cure-death model but can be adapted to more complex multistate endpoints and equivalence or superiority analyses.

KEYWORDS endpoints, hospital-acquired infection, multistate model, noninferiority

Severe bacterial diseases arising in hospitalized patients include, for example, hospital-acquired pneumonia (HAP) and ventilator-associated pneumonia (VAP). The economic impact of HAP and VAP is an immense burden on health care resources (1). In current and former clinical trials, a variety of primary endpoints have been used (2–4)—even recommendations given by existing guidelines are not consistent. The European Medicines Agency (EMA) (5) proposes clinical cure, the clinical outcome of a test-of-cure (TOC) visit, as an acceptable primary endpoint. In contrast, the Food and Drug Administration (FDA) (6) recommends considering 28-day mortality as the primary endpoint. The drawback of the former endpoint is that cure is a sojourn state (patients may still die after cure), while the latter does not account for cure status. Below, we will account for both cure and vital status.

Especially in an intensive care unit with severely ill patients, a high mortality rate can be assumed within the first days after infection, and thus, death may preclude the TOC as a so-called “competing event” (7). Also, death following shortly after cure should be

Received 16 August 2017 **Returned for modification** 31 August 2017 **Accepted** 17 October 2017

Accepted manuscript posted online 23 October 2017

Citation Sommer H, Bluhmki T, Beyersmann J, Schumacher M, on behalf of the Combacte-Net and Combacte-Magnet consortium. 2018. Assessing noninferiority in treatment trials for severe infectious diseases: an extension to the entire follow-up period using a cure-death multistate model. *Antimicrob Agents Chemother* 62:e01691-17. <https://doi.org/10.1128/AAC.01691-17>.

Copyright © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Harriet Sommer, sommer@imbi.uni-freiburg.de.

H.S. and T.B. contributed equally to this work.

considered, since cure then does not benefit the patient. Recently, Doshi pointed to a situation where patients were considered cured but died on the same day (8). The joint analysis of nonmortal endpoints like cure and the competing endpoint, death over time, can be captured by the flexible multistate framework (9, 10), which found its way into applied infection control literature (11–18). For that purpose, it has recently been suggested to simultaneously study the two coprimary endpoints, cure and death (2, 19), in order to understand how the new treatment influences the whole etiological cure process, even beyond the TOC. This is achieved by a multistate “cure-death” model (19), which accounts for the time dependency of cure and death, the presence of competing risks, and various follow-up times. The patient-relevant outcome of interest is to be cured and alive over time. In terms of studying the efficacy of a novel treatment, such an outcome as the probability to be cured and alive is sometimes used in the context of bone marrow transplant studies (20), in a similar way, when estimating current leukemia-free survival (21) or for the estimation of being alive without relapse and immunosuppression for graft-versus-host disease in a population of patients with acute lymphoblastic leukemia (22). The purpose of such an outcome is to synthesize the different summary statistics commonly used, the proportion of patients who respond, and the average duration of response (here defined as the proportion of cured patients and the average duration of cure). Our approach can distinguish between a treatment producing a high cure rate but generally short-lived cure cases and another treatment with a low cure rate but longer cure durations. This is highly relevant from the patients’ perspective and provides a complete summary and an attractive visual display of the given data.

Further, the benefit of newly developed treatments over existing treatments in terms of efficacy is often only marginal, but these new treatments might be rather advantageous in reducing costs or having fewer adverse effects. Thus, noninferiority analyses are indispensable (23). For serious infections such as HAP and VAP, the EMA as well as the FDA recommended that it is possible to set up a reliable and consistent estimate of the efficacy of active treatment relative to a comparator, which can serve as the basis for defining a new inferiority margin for an active-controlled noninferiority trial (5, 6). The traditional noninferiority procedure for a treatment comparison is to examine the difference in cure proportions with corresponding confidence intervals at a prespecified time point, most often the end of follow-up. However, both from the patients’ and from a biological perspective, a much more convincing statement than merely demonstrating noninferiority at a single point in time is how the active treatment performs over the complete cure process (24, 25) and, simultaneously, if the treatment is also beneficial in terms of mortality. A time-simultaneous analysis also allows to put results from former trials with possibly different times of follow-up into perspective. This requires more advanced statistical methodology and has not been evaluated so far for clinical cure. To assess noninferiority over the entire follow-up period, a so-called “confidence band” is required, in which the difference in probabilities of being cured and alive for active treatment A minus control B over a relevant time interval (and not just at a single time point, possibly differing between studies and also within studies, where such time points may differ between patients), lies within a probability of, e.g., 95%. Such a confidence band generalizes the concept of a confidence interval to an entire time interval of interest. Treatment A can be deemed noninferior to B if the confidence band for the difference of the treatment-specific probability curves lies above the protocol-defined noninferiority margin over the whole time period. Finally, and to demonstrate direct implications to trial results, we applied this procedure to actual antimicrobial trial data.

In this article, the application of the cure-death model to a recently published ceftobiprole trial (26) is given in Results. In this trial, described in Materials and Methods, the new regimen of ceftobiprole was compared to the two-drug regimen of ceftazidime plus linezolid for the treatment of patients with HAP and VAP. We introduce the concept of the cure-death multistate model (Fig. 1) and describe how to assess noninferiority over an interval of interest. The application of the cure-death model to

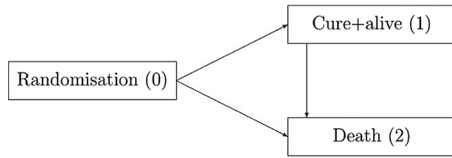


FIG 1 The cure-death model for comparing two antimicrobial therapies with an initial infection/randomization state, a cured and alive state, and a death state. The direction of arrows illustrates the potential transitions between the states determined by transition hazards. They can be interpreted as the momentary forces that pull a subject out of one state into another.

the ceftobiprole trial is shown to yield a much stronger noninferiority statement, both from the patients’ and from a biological perspective, than that published previously, without changing the noninferiority margin. We describe the benefits and limitations of the proposed methodology in Discussion. Details and an illustrative implementation example are provided in the Supplemental text.

RESULTS

The Aalen-Johansen estimator of the probability to be cured and alive over the whole time frame of interest is displayed in the upper panels of Fig. 2. For the entire sample (781 patients) and the HAP-excluding-VAP group (571 patients), the probability curves show a similar course across treatment groups. In the VAP-only group (210 patients), there is a clear distinction between treatments, favoring ceftazidime plus linezolid. The lower panels of Fig. 2 illustrate the probability of being cured and alive difference at time t [PCAD(t)] (solid black line) together with the 95% one-sided simultaneous confidence band (dashed black line) on the time interval of interest $[0, 47]$ (in days). The interval is chosen such that all observed transition times are covered. The boundary values (q) for the construction of the respective confidence band are also

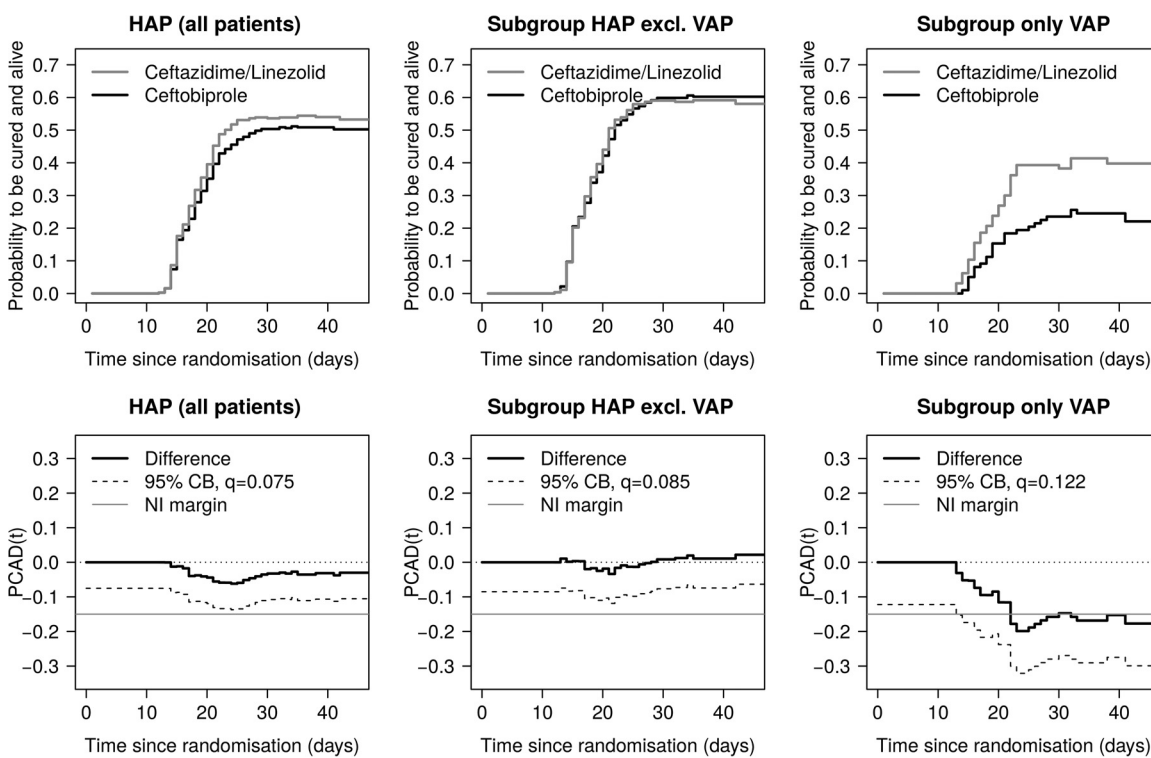


FIG 2 Transition probabilities and differences in probabilities derived from the Aalen-Johansen estimator for subgroups in the ceftobiprole trial (26). (Top) Probability to be cured and alive. (Bottom) Estimated difference of probabilities, PCAD, with the 95% one-sided simultaneous confidence band (CB), the corresponding boundary value, q , and the protocol-defined noninferiority (NI) margin of -15% on the time interval $[0, 47]$.

displayed [see Materials and Methods for details on $PCAD(t)$ and q]. It can be seen that for both the entire sample and the HAP-excluding-VAP group, the confidence band lies above the noninferiority margin of -15% (solid gray line) for the entire interval $[0, 47]$, but this is not true for the group of VAP patients. Hence, for the entire sample and the HAP-excluding-VAP group, noninferiority concerning “cured and alive” over the time period of interest is shown. All results do support the original analysis (26), which showed noninferiority of overall cure proportions for both the entire sample (cure risk difference ceftobiprole minus ceftazidime/linezolid of -2.9 [$-10.0, 4.1$]) and the HAP-excluding-VAP group (cure risk difference ceftobiprole minus ceftazidime/linezolid of 0.8 [$-7.3, 8.8$]) at the end of follow-up.

DISCUSSION

Our analysis not only confirms the noninferiority of ceftobiprole as found by Awad et al. (26) but also provides a stronger and more patient-benefiting noninferiority result. First, the endpoint “get cured and stay alive over time” is a highly relevant outcome in the context of antimicrobial trial data, because patients benefit from cure only when staying alive for a certain time. Second, we demonstrate noninferiority of being cured and alive over the complete treatment process and not only at the end of follow-up as intended in the original analysis. To include all time points simultaneously avoids having to decide for a single time point, especially when there is no agreement of the adequate time point for the TOC (4) or when a patient’s status may be highly time dependent on a daily basis. Third, in contrast to the traditional proportion comparisons, the comprehensive cure-death multistate model captures the complex timing of cure and death and competing events and also allows for different follow-up times due to death or right-censoring.

We emphasize that a multistate analysis may offer a more detailed insight into how treatment influences the cure and death process in other situations. For instance, even though noninferior cure and (30-day) all-cause mortality proportions for the treatment group may be observed at some prespecified time point, the probability of being cured and alive as a function in time may be substantially higher for the control group during some stages of follow-up, i.e., noninferiority cannot be established for some points in time. However, such time-dynamic patterns are important from a patients’ perspective, may have direct impact on patient care (see also references 27 and 28), and may provide additional biological insight. Depending on the disease and the medical indication, these additional clinical data can point out, e.g., that additional clinical action may be required. But this is not the case in the present study example; thus, continual noninferiority is demonstrated. Further, in a recent article (19), we examined several simulation scenarios in which a treatment might be beneficial by the end of follow-up but not over the entire treatment process. Such time-varying treatment effects might be detected by the approach suggested in the present paper but not by a conventional analysis.

In the aforementioned article (19), we used innovative regression methods for a risk ratio of being cured and alive and a restricted log rank-based test for a formal treatment comparison. However, the approach in reference 19 does not immediately lend itself to comparing absolute differences between the outcome probabilities and a subsequent noninferiority analysis.

The applied cure-death multistate model requires a time-inhomogeneous Markov assumption (9), which is tantamount to assuming that the $1 \rightarrow 2$ hazard depends only on the time since randomization and not on the time since cure. According to the approach described by Beyersmann et al. (10), we assumed the Markov assumption to be reasonable in the present situation. If the Markov assumption is violated, a time-simultaneous analysis is still feasible, but the technical details become more involved. To begin, the Aalen-Johansen estimation of state occupation probabilities as considered in this work will still apply (29). The technical requirements on the censoring mechanism will be stricter but likely be met in a situation such as ours. Confidence bands may again be constructed using resampling techniques, either sampling with replacement from the individual patients or using an approach as in the present paper,

adapted to the method seen in reference 29. We also note that modeling is done within treatment strata.

Another point worth mentioning is that while this is a time-simultaneous approach, we do not extrapolate beyond follow-up. In other words, also the time-simultaneous approach is limited to the duration of follow-up within a study.

The proposed resampling procedure is not restricted to the cure-death model but can be applied to arbitrary multistate models with finite state spaces, which makes it attractive for various applications not restricted to the context of antimicrobial trial data. For instance, an illness-death model with recovery can be used to quantify alternative efficacy measures in clinical trials such as the probability to be alive and not under mechanical ventilation in patients with acute respiratory distress (30), the probability to be alive and not under immunosuppressive therapy in leukemia patients after stem cell transplantation (22, 31), or the probability to have a sustained cure (to be cured and free of recurrent infection) in patients with *Clostridium difficile* infection (27). Furthermore, it can easily be adapted to investigate the equivalence or superiority of a treatment A compared to a treatment B concerning an outcome transition probability over a time frame of interest, or it can be applied to more complex functionals incorporating the matrix of transition probabilities that quantify, e.g., length of stay (32).

A different question that can be posed is what if cure were assessed as a continuous variable and artificial dichotomization should be avoided. Then, one option would be a joint model (33) for the longitudinal continuous outcome, informatively censored by death. One might still envisage a comparison of cure curves in a such a case; however, how to define an endpoint that accounts for both vital and cure status is less clear.

In conclusion, we propose an approach that addresses three significant issues often arising in noninferiority clinical trials in the context of infection control, namely, the arbitrary nature of the timing of endpoint assessment within and not across the time interval of interest, the erosion of the 5% significance level over the entire endpoint interval due to multiple testing, and the lack of accounting for vital status after cure. Our approach optimizes use of the available data by assessing the endpoint “get cured and stay alive” over the entire time period of interest while calculating a confidence band over that period preserving the 5% confidence level. Most importantly, our proposal develops both a patient-centric and a biologically relevant approach and adds a valuable statistical tool for such analyses.

MATERIALS AND METHODS

Example study data. Treatment with ceftobiprole was established to combat a wide range of Gram-positive bacteria, and this drug belongs to the class of β -lactam antibiotics. The noninferiority trial described here (26) compared the new regimen of ceftobiprole to the two-drug regimen of ceftazidime plus linezolid. It was a double-blind, randomized, multicenter comparison involving 781 patients with HAP, among them 210 with VAP.

Clinical cure diagnosed at the TOC visit, mostly held within a time frame of 7 to 14 days after the end of treatment, served as the primary endpoint, and all-cause mortality was the secondary endpoint. With the protocol-defined noninferiority margin of -15% , risk differences of proportions of cured patients at the TOC showed that ceftobiprole is noninferior to ceftazidime plus linezolid for the entire study population of patients with HAP (-2.9 [$-10.0, 4.1$]) and for the population of HAP patients excluding the VAP patients (0.8 [$-7.3, 8.8$]). But noninferiority was not demonstrated in VAP patients (-13.7 [$-26.0, -1.5$]). These results were given by Awad et al. (26) and refer to ceftobiprole versus ceftazidime/linezolid.

The cure-death model. In order to suitably account for the time-dynamic pattern of cure and death after randomization, we focus on the “illness-death model without recovery” embedded in the flexible and powerful multistate model framework (9, 10). Since the context here is hospital-acquired infection, we call this model “cure-death model” (19) (Fig. 1), with state 0 defined as initial, state 1 as cured and alive, and state 2 as absorbing death. According to the study protocol, all patients start in state 0, which is randomization to treatment, immediately after infection. The time scale of interest is “time since randomization” in days. The outcome most relevant for patients is “being cured and alive” (state 1); however, all patients, whether cured or not, are permanently at risk for death (state 2) during the entire follow-up. If a patient is still alive at the end of follow-up, (s)he is right-censored. Note that we summarize the competing events “death from the condition treated” and “death from any cause” that can be analyzed separately if information is available; such a separation does not influence the present outcome of interest but may be interesting for a broader treatment comparison. The arrows in Fig. 1 illustrate the possible transitions within our multistate model determined by the so-called “transition hazards.” They can be interpreted as momentary forces that pull a subject out of one state into another.

However, a hazard-based analysis applying, for instance, the well-known Cox proportional hazards model does not allow direct statements regarding the probability of interest because the latter involves all hazards. Instead, the Aalen-Johansen estimator (9) is employed for nonparametric estimation of the probability of interest, which generalizes the well-known Kaplan-Meier estimator to multistate settings. Note that in the absence of right-censoring, the Aalen-Johansen estimator is identical to the relative proportion of patients being in a certain state at time t .

Assessing noninferiority over the entire follow-up period via confidence bands. The target aim of our analysis is to establish noninferiority for the difference in probabilities of being cured and alive for patients of two treatment groups, A and B, on a relevant time interval of interest. In practice, the maximum of this interval can be, e.g., the largest observed cure or death time. Pointwise 95% confidence intervals can be constructed for the difference of these probabilities. Similar to risk differences, we consider the difference of the treatment-specific Aalen-Johansen estimates, \hat{P} (cured and alive at t |treatment group) function (PCAD, which stands for the “probability of being cured and alive difference”), calculated as follows: $PCAD(t) = \hat{P}$ (cured and alive at t |A) – \hat{P} (cured and alive at t |B), in the comparison of treatments A and B.

But assessing noninferiority only pointwise separately for each time point t does not have an overall 5% significance level due to the problem of multiple testing (34). Instead, our aim is to construct a so-called (one-sided) confidence band over the time interval of interest. A confidence band is a region in which the entire true curve (here defined as the difference in probabilities over time) falls within a probability of 95%. Statistical inference is commonly based on approximate normality of the estimators. For the Aalen-Johansen estimator, this enables the construction of 95% confidence intervals via the usual “ $\pm 1.96 \times$ standard error” rule. However, a generalization to confidence bands requires some additional effort because analytical solutions for the calculation of such bands are available only for the standard survival setting and not for more complex multistate models as in the present data situation. As a solution, we suggest the use of a convenient and computationally attractive resampling procedure known as “wild bootstrap,” which was originally established for competing risks settings (35, 36).

The idea is to introduce computer-generated standard normal variates that are plugged into a transformation of the Aalen-Johansen estimator. Repeating this approach, say 1,000 times, allows deriving a quantile q . The one-sided 95% confidence band for the difference in probabilities is then given by its lower boundary, $PCAD(t) - q$.

Adapting the principles of confidence interval inclusion as discussed in reference 37 to simultaneous confidence bands, treatment A can be deemed noninferior to B if the 95% confidence band, i.e., $PCAD(t) - q$, lies above the protocol-defined noninferiority margin on the entire time interval of interest. This approach corresponds to a formal statistical noninferiority test at the 5% significance level.

The practical construction of the quantile q and an exemplary implementation of the procedure in the statistical software R (38) are given in the Supplemental text. Fast computation of the Aalen-Johansen estimators is provided by the `etm` function (39).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AAC.01691-17>.

SUPPLEMENTAL FILE 1, PDF file, 0.2 MB.

ACKNOWLEDGMENTS

We thank Marc Engelhardt from Basilea Pharmaceutica International Ltd. for providing the data from the ceftobiprole trial and appreciate the help of Frangiscos Sifakis for reviewing the manuscript prior to submission.

Harriet Sommer was supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115523 COMBACTE-NET and 115737-2 COMBACTE-MAGNET, resources of which are composed of financial contributions from the European Union’s Seventh Framework Programme (FP7/2007-2013) and EFPIA companies. Tobias Bluhmki was supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement n [115737-2 – COMBACTE-MAGNET], resources of which are composed of financial contributions from the European Union’s Seventh Framework Programme (FP7/2007-2013) and EFPIA companies. Jan Beyersmann was partially supported by DFG grant BE 4500/1-2.

We declare that we have no conflicts of interest relevant to this article.

REFERENCES

1. Arthur L, Kizor R, Selim A, van Driel ML, Seoane L. 2016. Antibiotics for ventilator-associated pneumonia. The Cochrane Library, John Wiley and Sons, Hoboken, NJ.
2. Timsit JF, de Kraker MEA, Sommer H, Weiss E, Bettiol E, Wolkewitz M, Wilson D, Harbarth S. 2017. Appropriate endpoints for evaluation of new antibiotic therapies for severe infections: a white paper from the COMBACTE network. *Intensive Care Med* 43:1002–1012. <https://doi.org/10.1007/s00134-017-4802-4>.
3. Harhay M, Wagner J, Ratcliffe S, Bronheim RS, Gopal A, Green S, Cooney E, Mikkelsen ME, Kerlin MP, Small DS, Halpern SD. 2014. Outcomes and

- statistical power in adult critical care randomized trials. *Am J Resp Crit Care* 189:1469–1478. <https://doi.org/10.1164/rccm.201401-0056CP>.
4. Weiss E, Essaied W, Adrie C, Zahar JR, Timsit JF. 2017. Treatment of severe hospital-acquired and ventilator-associated pneumonia: a systematic review of inclusion and judgment criteria used in randomized controlled trials. *Crit Care* 21:162. <https://doi.org/10.1186/s13054-017-1755-5>.
 5. European Medicines Agency. 2013. Addendum to the guideline on the evaluation of medicinal products indicated for treatment of bacterial infections. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/11/WC500153953.pdf. Accessed 15 August 2017.
 6. Food and Drug Administration. 2014. Guidance for industry: hospital-acquired bacterial pneumonia and ventilator-associated bacterial pneumonia: developing drugs for treatment. <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm234907.pdf>. Accessed 15 August 2017.
 7. Wolkewitz M, Cooper B, Bonten M, Barnett AG, Schumacher M. 2014. Interpreting and comparing risks in the presence of competing events. *Br Med J* 349:g5060. <https://doi.org/10.1136/bmj.g5060>.
 8. Doshi P. 2015. Speeding new antibiotics to market: a fake fix? *Br Med J* 350:h1453. <https://doi.org/10.1136/bmj.h1453>.
 9. Aalen O, Borgan Ø, Gjessing H. 2008. Survival and event history analysis: a process point of view. Springer Science & Business Media, New York, NY.
 10. Beyersmann J, Allignol A, Schumacher M. 2011. Competing risks and multistate models with R. Springer Science & Business Media, New York, NY.
 11. Wolkewitz M, von Cube M, Schumacher M. 2017. Multistate modeling to analyze nosocomial infection data: an introduction and demonstration. *Infect Control Hosp Epidemiol* 38:953–959. <https://doi.org/10.1017/ice.2017.107>.
 12. Munoz-Price L, Frencken J, Tarima S, Bonten M. 2016. Handling time dependent variables: antibiotics and antibiotic resistance. *Clin Infect Dis* 62:1558–1563. <https://doi.org/10.1093/cid/ciw191>.
 13. Harbarth S. 2013. What can we learn from each other in infection control? Experience in Europe compared with the USA. *J Hosp Infect* 83:173–184. <https://doi.org/10.1016/j.jhin.2012.12.003>.
 14. Schumacher M, Allignol A, Beyersmann J, Binder N, Wolkewitz M. 2013. Hospital-acquired infections—appropriate statistical treatment is urgently needed. *Int J Epidemiol* 42:1502–1508. <https://doi.org/10.1093/ije/dyt111>.
 15. De Angelis G, Allignol A, Murthy A, Wolkewitz M, Beyersmann J, Safran E, Schrenzel J, Pittet D, Harbarth S. 2011. Multistate modelling to estimate the excess length of stay associated with methicillin-resistant *Staphylococcus aureus* colonisation and infection in surgical patients. *J Hosp Infect* 78:86–91. <https://doi.org/10.1016/j.jhin.2011.02.003>.
 16. de Kraker MEA, Wolkewitz M, Davey PG, Grundmann H. 2011. Clinical impact of antimicrobial resistance in European hospitals: excess mortality and length of hospital stay related to methicillin-resistant *Staphylococcus aureus* bloodstream infections. *Antimicrob Agents Chemother* 55:1598–1605. <https://doi.org/10.1128/AAC.01157-10>.
 17. Beyersmann J, Gastmeier P, Grundmann H, Bärwolff S, Geffers C, Behnke M, Rüden H, Schumacher M. 2006. Use of multistate models to assess prolongation of intensive care unit stay due to nosocomial infection. *Infect Control Hosp Epidemiol* 27:493–499. <https://doi.org/10.1086/503375>.
 18. Samore M, Harbarth S. 2004. A methodologically focused review of the literature in hospital epidemiology and infection control, p 1645–1657. *In* Mayhall CG (ed), *Infection control and hospital epidemiology*, 3rd ed. Lippincott, Williams and Wilkins, Philadelphia, PA.
 19. Sommer H, Wolkewitz M, Schumacher M. 2017. The time-dependent cure-death model investigating two equally important endpoints simultaneously in trials treating high-risk patients with resistant pathogens. *Pharm Stat* 16:267–279. <https://doi.org/10.1002/pst.1809>.
 20. Klein P, Shu Y. 2002. Multi-state models for bone marrow transplantation studies. *Stat Methods Med Res* 11:117–139. <https://doi.org/10.1191/0962280202sm277ra>.
 21. Klein JP, Keiding N, Shu Y, Szydlo RM, Goldman JM. 2000. Summary curves for patients transplanted for chronic myeloid leukaemia salvaged by a donor lymphocyte infusion: the current leukaemia-free survival curve. *Br J Haematol* 109:148–152. <https://doi.org/10.1046/j.1365-2141.2000.01982.x>.
 22. Eefting M, de Wreede LC, Halkes CJM, von dem Borne PA, Kersting S, Marijt EWA, Veelken H, Putter H, Schetelig J, Falkenburg JHF. 2016. Multi-state analysis illustrates treatment success after stem cell transplantation for acute myeloid leukemia followed by donor lymphocyte infusion. *Haematology* 101:506–514. <https://doi.org/10.3324/haematol.2015.136846>.
 23. Fay MP, Follmann D. 2016. Non-inferiority tests for anti-infective drugs using control group quantiles. *Clin Trials* 13:632–640. <https://doi.org/10.1177/1740774516654861>.
 24. Powers J, Howard K, Saretsky T, Clifford S, Hoffmann S, Llorens L, Talbot G. 2016. Patient-reported outcome assessments as endpoints in studies in infectious diseases. *Clin Infect Dis* 63:S52–S56. <https://doi.org/10.1093/cid/ciw317>.
 25. Muscedere J, Day A, Heyland D. 2010. Mortality, attributable mortality, and clinical events as end points for clinical trials of ventilator-associated pneumonia and hospital-acquired pneumonia. *Clin Infect Dis* 51: S120–S125. <https://doi.org/10.1086/653060>.
 26. Awad S, Rodriguez A, Chuang Y, Marjanek Z, Pareigis AJ, Reis G, Scheeren TWL, Sánchez AS, Zhou X, Saulay M, Engelhardt M. 2014. A phase 3 randomized double-blind comparison of ceftobiprole medocaril versus ceftazidime plus linezolid for the treatment of hospital-acquired pneumonia. *Clin Infect Dis* 59:51–61. <https://doi.org/10.1093/cid/ciu219>.
 27. Sommer H, Timsit JF, Wolkewitz M. 2017. Bezlotoxumab and recurrent *Clostridium difficile* infection. *N Engl J Med* 376:1593–1596. <https://doi.org/10.1056/NEJMc1702531>.
 28. Bluhmki T, Bramlage P, Volk M, Kaltheuner M, Danne T, Rathmann W, Beyersmann J. 2017. Time-to-event methodology improved statistical evaluation in register-based health services research. *J Clin Epidemiol* 82:103–111. <https://doi.org/10.1016/j.jclinepi.2016.11.001>.
 29. Datta S, Satten GA. 2001. Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov model. *Stat Probabil Lett* 55:403–411. [https://doi.org/10.1016/S0167-7152\(01\)00155-9](https://doi.org/10.1016/S0167-7152(01)00155-9).
 30. Schoenfeld D, Bernard G. 2002. Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials of treatments for acute respiratory distress syndrome. *Crit Care Med* 30:1772–1777. <https://doi.org/10.1097/00003246-200208000-00016>.
 31. Schmoor C, Schumacher M, Finke J, Beyersmann J. 2013. Competing risks and multistate models. *Clin Cancer Res* 19:12–21. <https://doi.org/10.1158/1078-0432.CCR-12-1619>.
 32. Allignol A, Schumacher M, Beyersmann J. 2011. Estimating summary functionals in multistate models with an application to hospital infection data. *Comp Stat* 26:181–197. <https://doi.org/10.1007/s00180-010-0200-x>.
 33. Rizopoulos D. 2012. Joint models for longitudinal and time-to-event data: with applications in R. CRC Press Taylor & Francis Group, Boca Raton, FL.
 34. Dmitrienko A, Tamhane A, Bretz F. 2009. Multiple testing problems in pharmaceutical statistics. Chapman and Hall/CRC, Boca Raton, FL.
 35. Lin D. 1997. Non-parametric inference for cumulative incidence functions in competing risks studies. *Stat Med* 16:901–910. [https://doi.org/10.1002/\(SICI\)1097-0258\(19970430\)16:8<901::AID-SIM543>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1097-0258(19970430)16:8<901::AID-SIM543>3.0.CO;2-M).
 36. Beyersmann J, Di Termini S, Pauly M. 2013. Weak convergence of the wild bootstrap for the Aalen-Johansen estimator of the cumulative incidence function of a competing risk. *Scand J Stat* 40:387–402. <https://doi.org/10.1111/j.1467-9469.2012.00817.x>.
 37. Wellek S. 2010. Testing statistical hypotheses of equivalence and non-inferiority, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL.
 38. R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
 39. Allignol A, Schumacher M, Beyersmann J. 2011. Empirical transition matrix of multi-state models: the etm package. *J Stat Soft* 38:1–15. <https://doi.org/10.18637/jss.v038.i04>.