Diagnostic prediction tools for bacteraemia caused by 3rd generation cephalosporin-resistant Enterobacteriaceae in suspected bacterial infections: a nested case-control study

Wouter C. Rottier[1], Cornelis H. van Werkhoven[1], Yara R.P. Bamberg[1,a], J. Wendelien Dorigo-Zetsma[3], Ewoudt M. van de Garde[5], Babette C. van Hees[7], Jan A.J.W. Kluytmans[1,8], Emile M. Kuck[9], Paul D. van der Linden[4], Jan M. Prins[11], Steven F.T. Thijsen[10], Annelies Verbon[12], Bart J.M. Vlaminckx[6], Heidi S.M. Ammerlaan[1,b]. and Marc J.M. Bonten[1,2]

[1]Julius Center for Health Sciences and Primary Care, and [2]Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, the Netherlands; [3]Central Laboratory for Bacteriology and Serology, and [4]Department of Clinical Pharmacy, Tergooi Hospitals, Hilversum, the Netherlands; [5]Department of Clinical Pharmacy, and [6]Department of Medical Microbiology and Immunology, St. Antonius Hospital, Utrecht/Nieuwegein, The Netherlands; [7]Department of Medical Microbiology and Infection Control, Gelre Hospitals, Apeldoorn, the Netherlands; [8]Department of Medical Microbiology and Infection Control, Amphia Hospital, Breda, the Netherlands; [9]Department of Hospital Pharmacy, and [10]Department of Medical Microbiology and Immunology, Diakonessenhuis, Utrecht, the Netherlands; [11]Department of Internal Medicine, Academic Medical Center, Amsterdam, the Netherlands; [12]Department of Medical Microbiology and Infectious Diseases, Erasmus Medical Center, Rotterdam, the Netherlands.

[a] Present affiliation: Department of Microbiology, Leiden University Medical Center, Leiden, the Netherlands.

[b] Present affiliation: Department of Internal Medicine, Catharina Hospital, Eindhoven, the Netherlands.

Corresponding author

Wouter C. Rottier. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht. Huispostnummer STR 6.131, Postbus 85500, 3508 GA Utrecht, the Netherlands. Telephone: +31 88 75 576 76. Fax: +31 88 75 680 99. E-mail: w.c.rottier-2@umcutrecht.nl.

Keywords: risk factors; clinical prediction models; scoring systems; extended-spectrum beta-lactamases; empiric antibiotic therapy

Word count abstract: 248

Word count main text: 3023

Abstract

Objectives: Current antibiotic treatment guidelines on when to consider 3rd generation cephalosporin resistant Enterobacteriaceae (3GC-R EB) as a cause of infection have low specificity, thereby increasing unnecessary carbapenem use. Therefore, we aimed to develop new diagnostic scoring systems to direct initial carbapenem treatment to patients at risk of 3GC-R EB bacteraemia.

Methods: A retrospective nested case-control study was performed that included patients ≥18 years from 8 Dutch hospitals in whom blood cultures were obtained and intravenous antibiotics were initiated. Patients with 3GC-R EB bacteraemia were each matched to four control infection episodes within the same hospital, based on blood culture date and onset location (community or hospital). Starting from 32 commonly described clinical risk factors available at infection onset, selection strategies were used to derive scoring systems for the probability of community- and hospital-onset 3GC-R EB bacteraemia.

Results: Among 22,506 community-onset and 8,110 hospital-onset infections, respectively 90 (0.4%) and 82 (1.0%) were 3GC-R EB bacteraemias. As control populations, 360 community-onset and 328 hospital-onset infection episodes were included. The derived community-onset and hospital-onset scoring system consisted of 6 and 9 predictors, respectively, and both showed good discrimination with c-statistics of 0.807 and 0.842. Cutoffs for the scores could be chosen such that ~20% of patients would be eligible for empirical carbapenem treatment, which would capture ~70% of those with 3GC-R EB bacteraemia.

Conclusions: These prediction rules for 3GC-R EB bacteraemia, specifically geared towards the initiation of empiric antibiotic treatment, may improve the balance between inappropriate antibiotics and carbapenem overuse.

Introduction

As a consequence of the emergence of infections caused 3rd generation cephalosporin (3GC) resistant Enterobacteriaceae (3GC-R EB; in this manuscript used synonymously with extended-spectrum β-lactamase (ESBL) producing Enterobacteriaceae), physicians are increasingly faced with the question which patients need empiric antibiotic treatment covering these pathogens. Hence, patients and physicians might benefit from prediction rules for 3GC-R EB. For instance, some studies yielded risk factors for carriage of ESBL-producing Enterobacteriaceae at hospital admission [1–4]. Yet, although relevant for infection control purposes, that knowledge may not accurately predict etiologic causes of infection. Other studies aimed to distinguish ESBL- or carbapenemase-producing Enterobacteriaceae as a cause of infection, but only included bacteraemic patients [5–7]. Applicability of these models to choosing empiric treatment is unknown, as among patients with suspected bacterial infection, microbiologically documented Enterobacteriaceae infection only comprises a subset of all possible aetiologies.

Current Dutch empiric treatment guidelines designate patients at risk of infection caused by 3GC-R EB based on prior colonization or infection with 3GC-R EB or on prior exposure to cephalosporins or fluoroquinolones, as these were identified as risk factors in patients with bacteraemia caused by these pathogens [8]. Yet, when evaluated in a cohort of patients needing empiric antibiotic treatment, full implementation would lead to considerable carbapenem overuse, with a low sensitivity to actually detect those patients with bacteraemia caused by 3GC-R EB [9]. Therefore, we aimed to develop prediction rules to identify, among patients needing intravenous empiric antibiotic therapy, those having an infection caused by 3GC-R EB.

Methods

Setting and patients

This was a retrospective nested case-control study involving 8 hospitals, of which 3 university hospitals, in the Netherlands. Between January 1st 2008 and December 31st 2010, we included all

consecutive patients of 18 years of age or older in whom a blood culture was obtained and intravenous broad-spectrum β-lactam antibiotics (i.e. not penicillin or flucloxacillin), aminoglycosides, and/or fluoroquinolones were started on the day of the blood culture or the day after, irrespective of duration. Patients receiving any of the eligible antibiotics on the day of blood culture obtainment were excluded if these had been initiated prior to this day (see Supplementary Table 2 for illustrating examples). In addition, patients with 3GC-R EB bacteraemia in the year prior were excluded, as it was assumed that treating physicians would always provide therapy aimed at these organisms in case of renewed infection. Patients could be included more than once, if a subsequent episode complied with in- and exclusion criteria. Additional information on hospital characteristics, study periods, and databases used in each of the hospitals is provided in Supplementary Table 1.

Infection episodes were separated into two cohorts: the community-onset cohort comprised episodes in which the first blood culture was collected during the first three calendar days of hospitalization, and the hospital-onset cohort consisted of episodes in which blood cultures were obtained later during hospitalization.

The causative pathogen of each episode was based on the results of blood cultures obtained during the onset days (i.e. the first day if antibiotics were started on the same day as blood culture obtainment and the first two days if antibiotics were started the day after). Cases were defined as patients with 3GC-R EB bacteraemia (see Supplementary Table 1 for definition of 3GC resistance in each of the hospitals). We aimed to include approximately 100 cases for both cohorts, in order to be able to construct an initial logistic regression model with 10 variables, based on the 10 events per variable recommendation [10].

The control population was defined as all other infection episodes, including non-bacteraemic episodes and episodes with blood cultures yielding non-resistant Enterobacteriaceae, other bacteria or fungi. For efficiency reasons, this population was not analysed in its entirety. Instead, four controls were matched to each case, a ratio chosen because of minimal gains in statistical power with more

controls [11]. Controls were matched on hospital, being in the community or hospital-onset cohort, and being closest in time to the blood culture day of the case episode.

Due to its retrospective nature, the Dutch Medical Research Involving Subjects Act did not apply to this study. In each of the participating hospitals, applicable local guidelines for non-interventional studies were followed. In accordance with Dutch regulations, informed consent was waived for the study. Reporting of this study was in accordance with the TRIPOD Statement [12].

Data collection

All selected cases and controls were subjected to chart review to obtain information that was considered available at the moment that the initial antibiotics were prescribed (referred to as infection onset). Blinding for the outcome during chart review was not considered feasible. Please refer to Supplementary Table 3 for an overview of all collected variables.

Statistical analysis

Two separate prediction models were constructed, one for community-onset and one for hospital-onset infections. Data analyses were performed in R (version 3.2.2) [13], including packages *mice* 2.25 [14], *rms* 4.5-0 [15], *pROC* 1.8 [16], and *xtable* 1.8-2 [17]. Descriptive analyses of predictors were based on non-missing data only. Some variables were aggregated because of high correlation, low prevalence, and/or similar associations with the outcome (indicated in Table 2). Additionally, the number of categories for suspected sources was reduced to four by combining categories with low frequencies into a single remaining group (original categories in Supplementary Table 3), and categories for antibiotic use were created based on prevalence and assumed predictive power for 3GC-R EB infection. Twenty imputed datasets were created to deal with missing values (see Supplementary Material for exact procedures).

For both community-onset and hospital-onset infections, the models in Table 1 were constructed. Continuous predictors were initially introduced into models with restricted cubic spline functions

with three knots to allow for non-linear associations. In the *final models*, we evaluated if simplification to a linear predictor was possible. Performances of the *final* (*simplified*) *models* and sensitivity analyses were compared by means of the Akaike's Information Criterion (AIC). The *final simplified models* were additionally compared to two basic models, the *prior identification model* and *two-predictor model* (Table 1).

Regression coefficients of the *final simplified models* were pooled over imputation datasets by means of Rubin's rules and shrunk according to model optimism (see description further on). Also, intercepts of the models were adjusted for the sampling fraction of the controls, and controls were weighted by the inverse of the sampling fraction, as previously described [18]. Calibration of the predicted and observed probabilities was visually inspected for separate imputation datasets. Discrimination was assessed by areas under the curves for receiver operating characteristic curves (referred to as C-statistic), averaged over the imputation datasets. Sensitivity, specificity and positive and negative predictive values, and prevalence (i.e. fraction of the population classified as at risk of 3GC-R EB bacteraemia) were calculated for different cutoffs of the predicted risk, again averaged over the imputation datasets. A simplified score was created by multiplying the regression coefficients with a constant, followed by rounding to easy-to-use values. Performance of this score was determined similarly.

Estimation of model optimism

For the *final simplified models*, optimism was estimated by creating 2000 bootstrap samples, creating a new prediction model for each of these samples, and comparing the model's performance in the original and bootstrapped data. Optimism results from the fact that models are developed on a population sample and suffer from overfitting, which jeopardizes generalizability to the total population [19]. By means of bootstrapping, the expected effects when applying the model within the entire population are mimicked (see Supplementary Material for further details). Optimism was estimated for model coefficients, derived odds ratios and C-statistics. During the same procedure, it

was also evaluated how much lower sensitivity and higher prevalence would be due to optimism, when applying a cutoff probability above which patients are classified as at risk. For this evaluation cutoff probabilities were selected for which either sensitivity or prevalence corresponded to the basic *two-predictor model*.

Results

Probabilities of 3GC-R EB bacteraemia were 0.4% (n = 90) for the community-onset infection cohort (22,506 episodes) and 1.0% (n = 82) for the hospital-onset infection cohort (8,110 episodes). These case populations were matched to 360 community-onset control episodes and 328 hospital-onset control episodes (Table 2). Multiple selection of individual patients, albeit with different episodes, as case and/or control were allowed and occurred 8 times within the community-onset, and 9 times within the hospital-onset dataset. Isolated pathogens from blood cultures and initial antibiotic therapy are presented in Supplementary Tables 4 and 5.

Community-onset infection

The *final simplified model* for predicting 3GC-R EB bacteraemia in community-onset infection (Table 3) showed adequate discrimination (C-statistic = 0.808 (95% CI 0.756-0.855), optimism-corrected 0.775 (95% CI 0.705-0.839)) and calibration (Supplementary Figure 1). In sensitivity analyses, a model variant that included both prior use of cephalosporins or fluoroquinolones, and prior use of a residual category of antibiotics (instead of any antibiotic use in the *final simplified model*) was deemed most parsimonious (Supplementary Tables 6 and 7).

The derived scoring system (Table 4) had a performance similar to the original model (Figure 1a; C-statistic 0.807 (95% CI 0.756-0.855), not corrected for optimism). Figure 2a depicts the trade-off between sensitivity and prevalence at different cutoffs for being at risk of 3GC-R EB bacteraemia, compared to using the *prior identification model* (sensitivity 24.4% and prevalence 2.8%), and the *two-predictor model* (sensitivity 53.9% and prevalence 21.5%).

For instance, patients with a score of 120 or higher would have a probability of 1.7% (positive predictive value) of having 3GC-R EB bacteraemia, but when using 120 as a cutoff 45.7% of all patients with 3GC-R EB bacteraemia would be missed (1 – sensitivity). This sensitivity (or proportion missed) is comparable to the simpler *two-predictor model*; however, the scoring system reduces eligibility for carbapenem use (prevalence) by approximately 40%, from 21.5% to 12.8%.

Bootstrapping of the underlying model indicated that when applying this cutoff in a future patient population some performance loss should be expected due to model optimism. The optimism-corrected sensitivity for future populations was 6.2 percentage points lower, whereas a change in prevalence was hardly noticeably (Table 6). Similarly, when basing the probability cutoff on the prevalence of the *two-predictor model* within the study sample, prevalence is expected to be robust in future populations, but a decrease in sensitivity of approximately 4.0 percentage points should be anticipated.

Hospital-onset infection

Discrimination and calibration also appeared adequate for the *final simplified model* predicting 3GC-R EB bacteraemia in hospital-onset infection (Table 5; C-statistic = 0.842 (95% CI 0.793-0.886), optimism-corrected 0.811 (95% CI 0.742-0.873); Supplementary Figure 2). Sensitivity analyses revealed better performance when including more variables, especially those related to antibiotic use and the suspected source of infection, and a model including 13 instead of 9 parameters had the lowest AIC (Supplementary Tables 10 and 11).

The derived simplified scoring system (Table 4) performed very similar to the original model (Figure 1b; C-statistic 0.842 (95% CI 0.794-0.887), not corrected for optimism). In Figure 2b, sensitivity and prevalence at different scoring cutoffs are compared to using the *prior identification model* (sensitivity 35.4% and prevalence 5.2%), and the *two-predictor model* (sensitivity 79.3% and prevalence 52.8%).

Patients with scores of 110 or higher would have a positive predictive value for 3GC-R EB bacteraemia of 3.1%, and with this cutoff 18.5% of all patients with 3GC-R EB bacteraemias would be missed, similarly to the *two-predictor model*. Yet, carbapenem eligibility would be reduced with almost 50% (27.0% vs. 52.8%).

In this scenario, bootstrapping indicated that sensitivity in future patient populations should again be expected to be somewhat lower (-5.3%; Table 6). If a cutoff were however based on prevalence of the *two-predictor model*, substantially compromised performance in future patients is not likely to occur.

Discussion

We developed prediction models to more accurately identify patients with bacteraemia caused by 3GC-R EB bacteraemia among those in whom empiric intravenous antibiotic therapy aimed at Gram-negatives is initiated. The use of the derived scoring systems could improve appropriateness of empiric antibiotic therapy and reduce unnecessary use of broad-spectrum therapy. Compared to a basic model incorporating only prior 3GC-R EB identification and exposure to cephalosporins and/or fluoroquinolones, eligibility for empiric carbapenem use could be reduced by approximately 40% while maintaining a similar risk of missing patients with 3GC-R EB bacteraemia.

With a global emergence of antibiotic resistance, physicians must assess the risks of missing resistant causative pathogens when starting empiric antibiotic treatment [20]. Risk avoidance, albeit imaginable in many situations, is one of the driving forces for broad-spectrum antibiotic use, fuelling the global pandemic of antimicrobial resistance. Better prediction rules for infections caused by antibiotic-resistant pathogens are therefore needed. Prediction scores have been developed for Gram-negative bacteraemia in septic patients [21], carriage of or infection with ESBL-producing Enterobacteriaceae at hospital admission [1,22,23], and distinguishing bacteraemia with ESBL- or carbapenemase-producing pathogens from bacteraemia with susceptible Enterobacteriaceae [5–7]. Yet, guidance on incorporating the risk of 3GC-R EB in selecting empiric antibiotics is currently

lacking. For clarity, 3GC-R EB bacteraemia is a subset of Enterobacteriaceae bacteraemia, which is a subset of all bacteraemia episodes. Risk factors for any of these overarching categories may alter the probability of bacteraemia caused by antibiotic-resistant Enterobacteriaceae. This emphasizes the need of selecting a clinically meaningful patient population when deriving a prediction rule. We therefore focused on patients receiving initial antibiotic therapy aimed at Enterobacteriaceae, rather than selecting patients that had, in retrospect, bacteraemia. The differences in predictors for community-onset and hospital-onset bacteraemia underscore the relevance of distinguishing both entities in clinical prediction of infections caused by resistant Enterobacteriaceae.

As expected, prior identification of 3GC-R EB was the strongest predictor in both models. Identification was mostly based on previous clinical cultures, rather than on screening for carriage. The latter is in the Netherlands only practiced in intensive care units and for highly selected risk groups. Naturally, more screening will further increase the sensitivity of this predictor for bacteraemia with 3GC-R EB. Yet, as infection rates among colonized patients are low [24,25], it is unsure whether positive predictive values of models will improve. In fact, if low-risk carriers would be identified by screening more frequently than high-risk carriers, positive predictive values might even decline.

We applied sophisticated modelling techniques, including multiple imputation and internal validation by means of bootstrapping. The latter resulted in optimism-adjusted odds ratios and C-statistics, giving insight in values expected when applying models to an external cohort. Expected performance loss when selecting specific probability cutoffs for clinical use has also been calculated (Table 4). Nevertheless, we still recommend prospective external validation of the models before clinical implementation, for several reasons. First, even after shrinkage, optimism may still be present, as some steps could not be replicated in the bootstrap procedure, such as aggregation after observing similar associations with the outcome, simplification of continuous variables to linear predictors, and derivation of a scoring system. Second, explorative model variants indicated that predictive

performance could be enhanced, albeit at the cost of including more determinants. The optimism of these variants has not been calculated, and may well offset potential increases in performance observed in this study sample in external validation. Second, the current study relied on data available in medical charts. We used pragmatic in- and exclusion criteria, which might not fully reflect intended clinical use, and as data collection was not blinded for outcome, information bias is not excluded. Moreover, potentially relevant predictors, especially for community-onset infection, such as international travel, animal contact, known colonization in household members, and dietary preferences could not be collected. The same holds for determination of colonization pressure, which might be a relevant predictor for hospital-onset infections. Third, although prevalence of ESBL in Enterobacteriaceae infection in the Netherlands, especially in the community setting, does not differ considerably from other Western European countries [26–29], it is worthwhile to evaluate performance of risk factors in settings with varying ESBL prevalence.

A study limitation is that the outcome was restricted to bacteraemic episodes, not including non-bacteraemic infections caused by 3GC-R EB, and physicians should be aware of this when applying the score in clinical practice. Non-bacteraemic 3GC-R EB infection are more common than bacteraemic infections in patients being empirically treated [9], but with an overall prevalence of <5%, they will not have had a substantial impact on the composition of control groups. Future studies may consider classifying these infections as outcomes. However, due to the more benign course, initial treatment with carbapenems may not have a high priority in non-bacteraemic infections. Ultimately, using such broad-spectrum antibiotics as initial therapy is futile if patients would not actually benefit from them.

Another limitation of our study is that empiric coverage of 3GC-R EB is just one aspects of selection of appropriate empiric therapy. Other potential pathogens and resistance mechanisms, such as *Pseudomonas aeruginosa*, might justify alterations in empiric treatment even in the absence of risk factors for 3GC-R EB. In some countries, high incidences of infections with carbapenemase-producing

Enterobacteriaceae may limit usefulness of our models. On the other hand, escape therapy for 3GC-R EB might not necessarily involve carbapenems, due to underlying resistance mechanisms other than ESBL, or favourable patterns of co-resistance. Ideally, frameworks for selecting empiric therapy should evaluate the probability of success of many different antibiotic agents. An example of such an approach is TREAT, an automated system for recommending antibiotic treatment based on, amongst others, patient and infection characteristics and local epidemiology [30]. TREAT can predict the presence of Gram-negative causative pathogens in infection with some accuracy [31], but performance with regard to resistant variants remains unknown. However, TREAT has not been widely adopted, and simple prediction rules may be easier to incorporate into clinical practice.

In conclusion, identification of only those patients with an infection caused by 3GC-R EB amongst all patients that need empiric antibiotic therapy remains a daunting task. This is reflected by the fact that an acceptable level of empiric carbapenem use is likely to be associated with a considerable fraction of 3GC-R EB bacteraemias being missed. Yet, the prediction rules developed here are the first to truly quantify this trade-off for the appropriate patient domain. In addition, they offer considerable improvement in detecting such patients as compared to guidelines currently in place, and as such, they provide useful starting points for optimizing empiric antibiotic strategies.

Transparency declaration

W.C.R. has received research funding from Merck Sharp & Dohme (MSD). All other authors declare no conflicts of interest. This work was supported by a research grant from the Netherlands Organisation for Health Research and Development (project number 205200007) to H. S. M. A.

We would like to thank Evelien Stevenson and Tim Deelen for assistance with the data collection, and Maaike van Mourik and Loes de Bruin for assistance with the statistical analysis.

Results from this study were presented at the 26th European Congress of Clinical Microbiology and Infectious Diseases, Amsterdam, the Netherlands (9-12 April 2016, E-poster eP0083).

References

[1] Tumbarello M, Trecarichi EM, Bassetti M, De Rosa FG, Spanu T, Di Meco E, et al. Identifying patients harboring extended-spectrum-beta-lactamase-producing Enterobacteriaceae on hospital admission: derivation and validation of a scoring system. Antimicrob Agents Chemother 2011;55:3485–90. doi:10.1128/AAC.00009-11.

[2] Platteel TN, Leverstein-van Hall M a., Cohen Stuart JW, Thijsen SFT, Mascini EM, van Hees BC, et al. Predicting carriage with extended-spectrum beta-lactamase-producing bacteria at hospital admission: a cross-sectional study. Clin Microbiol Infect 2015;21:141–6. doi:10.1016/j.cmi.2014.09.014.

[3] Pasricha J, Koessler T, Harbarth S, Schrenzel J, Camus V, Cohen G, et al. Carriage of extended-spectrum beta-lactamase-producing Enterobacteriacae among internal medicine patients in Switzerland. Antimicrob Resist Infect Control 2013;2:20. doi:10.1186/2047-2994-2-20.

[4] Shitrit P, Reisfeld S, Paitan Y, Gottesman B-S, Katzir M, Paul M, et al. Extended-spectrum beta-lactamase-producing Enterobacteriaceae carriage upon hospital admission: prevalence and risk factors. J Hosp Infect 2013;85:230–2. doi:10.1016/j.jhin.2013.07.014.

[5] Goodman KE, Lessler J, Cosgrove SE, Harris AD, Lautenbach E, Han JH, et al. A clinical decision tree to predict whether a bacteremic patient is infected with an extended-spectrum β-lactamase-producing organism. Clin Infect Dis 2016;63:896–903. doi:10.1093/cid/ciw425.

[6] Martin ET, Tansek R, Collins V, Hayakawa K, Abreu-Lanfranco O, Chopra T, et al. The carbapenem-resistant Enterobacteriaceae score: A bedside score to rule out infection with carbapenem-resistant Enterobacteriaceae among hospitalized patients. Am J Infect Control 2013;41:180–2. doi:10.1016/j.ajic.2012.02.036.

[7] Leibman V, Martin ET, Tal-Jasper R, Grin L, Hayakawa K, Shefler C, et al. Simple bedside score to optimize the time and the decision to initiate appropriate therapy for carbapenem-resistant Enterobacteriaceae. Ann Clin Microbiol Antimicrob 2015;14:1–5. doi:10.1186/s12941-015-0088-y.

[8] Stichting Werkgroep Antibioticabeleid. SWAB guidelines for antibacterial therapy of adult patients with sepsis. Amsterdam, the Netherlands: 2010.

[9] Rottier WC, Bamberg YRP, Dorigo-Zetsma JW, van der Linden PD, Ammerlaan HSM, Bonten MJM. Predictive value of prior colonization and antibiotic use for third-generation cephalosporin-resistant Enterobacteriaceae bacteremia in patients with sepsis. Clin Infect Dis 2015;60:1622–30. doi:10.1093/cid/civ121.

[10] Pavlou M, Ambler G, Seaman SR, Guttmann O, Elliott P, King M, et al. How to develop a more accurate risk prediction model when there are few events. BMJ 2015;351:h3868. doi:10.1136/bmj.h3868.

[11] Rothman KJ, Greenland S, Lash TL. Case-Control Studies. In: Rothman KJ, Greenland S, Lash TL, editors. Mod. Epidemiol. 3rd ed., Philadelphia, PA, USA: Lippincott Williams & Wilkins; 2008, p. 111–27.

[12] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Ann Intern Med 2015;162:55. doi:10.7326/M14-0697.

[13] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: 2015.

[14] Buuren S van, Groothuis-Oudshoorn K. mice : Multivariate Imputation by Chained Equations in R. J Stat Softw 2011;45. doi:10.18637/jss.v045.i03.

[15] Harrell Jr FE. rms: Regression Modeling Strategies. 2016.

[16] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77. doi:10.1186/1471-2105-12-77.

[17] Dahl DB. xtable: Export Tables to LaTeX or HTML. 2016.

[18] Huang Y, Pepe MS. Assessing risk prediction models in case-control studies using semiparametric and nonparametric methods. Stat Med 2010;29:1391–410. doi:10.1002/sim.3876.

[19] Steyerberg EW. Chapter 5: Overfitting and optimism in prediction models. Clin. Predict. Model. A Pract. approach to Dev. validation, Updat., New York, NY: Springer Science+Business Media; 2009, p. 83–100.

[20] Pogue JM, Kaye KS, Cohen DA, Marchaim D. Appropriate antimicrobial therapy in the era of multidrug-resistant human pathogens. Clin Microbiol Infect 2015;21:302–12. doi:10.1016/j.cmi.2014.12.025.

[21] Bates DW, Sands K, Miller E, Lanken PN, Hibberd PL, Graman PS, et al. Predicting bacteremia in patients with sepsis syndrome. J Infect Dis 1997;176:1538–51. doi:10.1086/514153.

[22] Slekovec C, Bertrand X, Leroy J, Faller J-P, Talon D, Hocquet D. Identifying patients harboring extended-spectrum-β-lactamase-producing Enterobacteriaceae on hospital admission is not that simple. Antimicrob Agents Chemother 2012;56:2218–9; author reply 2220. doi:10.1128/AAC.06376-11.

[23] Johnson SW, Anderson DJ, May DB, Drew RH. Utility of a clinical risk factor scoring model in predicting infection with extended-spectrum β-lactamase-producing Enterobacteriaceae on hospital admission. Infect Control Hosp Epidemiol 2013;34:385–92. doi:10.1086/669858.

[24] Reddy P, Malczynski M, Obias A, Reiner S, Jin N, Huang J, et al. Reply to Blot et al. Clin Infect Dis 2008;46:482–482. doi:10.1086/526350.

[25] Ruppé E, Pitsch a, Tubach F, de Lastours V, Chau F, Pasquet B, et al. Clinical predictive values of extended-spectrum beta-lactamase carriage in patients admitted to medical wards. Eur J Clin Microbiol Infect Dis 2012;31:319–25. doi:10.1007/s10096-011-1313-z.

[26] Huijbers PMC, de Kraker M, Graat E a M, van Hoek a H a M, van Santen MG, de Jong MCM, et al. Prevalence of extended-spectrum β-lactamase-producing Enterobacteriaceae in humans living in municipalities with high and low broiler density. Clin Microbiol Infect 2013;19:E256–9. doi:10.1111/1469-0691.12150.

[27] Reuland E a., Al Naiemi N, Kaiser a. M, Heck M, Kluytmans J a JW, Savelkoul PHM, et al. Prevalence and risk factors for carriage of ESBL-producing Enterobacteriaceae in Amsterdam.

J Antimicrob Chemother 2016;71:1076–82. doi:10.1093/jac/dkv441.

[28] Valenza G, Nickel S, Pfeifer Y, Eller C, Krupa E, Lehner-Reindl V, et al. Extended-spectrum-β-lactamase-producing Escherichia coli as intestinal colonizers in the German community. Antimicrob Agents Chemother 2014;58:1228–30. doi:10.1128/AAC.01993-13.

[29] Nicolas-Chanoine M-H, Gruson C, Bialek-Davenet S, Bertrand X, Thomas-Jean F, Bert F, et al. 10-Fold increase (2006-11) in the rate of healthy subjects with extended-spectrum β-lactamase-producing Escherichia coli faecal carriage in a Parisian check-up centre. J Antimicrob Chemother 2013;68:562–8. doi:10.1093/jac/dks429.

[30] Leibovici L, Paul M, Nielsen AD, Tacconelli E, Andreassen S. The TREAT project: decision support and prediction using causal probabilistic networks. Int J Antimicrob Agents 2007;30 Suppl 1:S93–102. doi:10.1016/j.ijantimicag.2007.06.035.

[31] Paul M, Nielsen AD, Goldberg E, Andreassen S, Tacconelli E, Almanasreh N, et al. Prediction of specific pathogens in patients with sepsis: evaluation of TREAT, a computerized decision support system. J Antimicrob Chemother 2007;59:1204–7. doi:10.1093/jac/dkm107.

Table 1 Specification of models[a]

Main models	
Final model	1. Selection of 10 relevant predictors based on eyeballing associations of predictors with outcomes, and considerations related to coverage of the entire spectrum of known risk factors for 3GC-R EB, and ease-of-use of any resulting model. 2. Backward stepwise logistic regression analysis until all remaining predictors had p-values < 0.2[b].
Simplified final model	Multivariable logistic regression analysis with predictors retained in the final model, but all continuous variables modelled linearly[c].
Sensitivity analyses	
Variant 1	The set of 10 predictors that were selected in step 1 of the final model all forced into the multivariable logistic regression analysis; no further selection performed.
Variant 2	Forward stepwise multivariable logistic regression analysis with all potential predictors indicated in Table 2, with p value for inclusion < 0.2[b].
Variant 3A	1. Within the set of 10 predictors that were selected in step 1 of the final model, the antibiotic use predictor was replaced by *use of cephalosporins or fluoroquinolones, use of other beta-lactams*, and in case of CO, *use of aminoglycosides, macrolides or other antibiotics*, and in case of HO, *use of carbapenems*, *use of aminoglycosides*, *use of macrolides*, and *use of other antibiotics*. 2. Backward stepwise logistic regression analysis until all remaining predictors had p-values < 0.2[b].

Variant 3B	Same as Variant 3A, but *use of cephalosporins or fluoroquinolones* separated into *use of cephalosporins* and *use of fluoroquinolones*.
Variant 4	1. To the set of 10 predictors that were selected in step 1 of the final model, *suspected source of infection: intra-abdominal infection*, and in case of HO, suspected *source of infection: urinary tract infection* were added. 2. Backward stepwise logistic regression analysis until all remaining predictors had p-values < 0.2[b].
Robust variant	Predictors retained in the final model and all model variants forced into the multivariable logistic regression analysis; no further selection performed.
Cross-validation	Predictors retained in the final simplified model for the other setting (HO vs. CO; except *length of hospital stay prior to infection*, as this was deemed irrelevant for CO infection) forced into the multivariable logistic regression analysis; no further selection performed.
Basic models	
Prior identification model	Population classified as at risk of 3GC-R EB bacteraemia consists of: • *Prior identification of 3GC-R EB (prior one year)*
Two-predictor model (modification of Dutch sepsis guidelines) [8]	Population classified as at risk of 3GC-R EB bacteraemia consists of: • *Prior identification of 3GC-R EB (prior one year)* (guideline restricts to ESBL-producing, but we assumed all forms of 3GC resistance to be relevant for selecting empiric antibiotics) • *Use of cephalosporins or fluoroquinolones (prior two months)* (guideline uses a one-month period, but data on antibiotics were collected in such a

	manner that separation by month was impossible)

Abbreviations: CO, community-onset (model); HO, hospital-onset (model).

[a] Predictors retained in main models and sensitivity analyses are indicated in Supplementary Tables 6 (CO) and 10 (HO).

[b] P-value as calculated by Wald test, pooled from 20 imputation sets by means of Rubin's rules.

[c] In all other multivariable logistic regression models, continuous variables (*age* and *length of hospital stay prior to infection*) were modelled as restricted cubic splines with 3 knots (*rcs* function from *rms* package version 4.5-0 for *R*).

Table 2 Clinical characteristics of cases and controls from both the community-onset and hospital-onset cohort

Predictor[a]	Community-onset infection			Hospital-onset infection		
	Cases (N = 90)[b], n/N with data (%)	Controls (N = 360)[c], n/N with data (%)	OR (95% CI)[d]	Cases (N = 82)[b], n/N with data (%)	Controls (N = 328)[c], n/N with data (%)	OR (95% CI)[d]
Female gender	39/90 (43)	158/360 (44)	0.98 (0.61-1.56)	32/82 (39)	129/328 (39)	0.99 (0.60-1.62)
Age in years, median (IQR)	69 (61-76)[e]	63 (50-76)[e]	1.02 (1.00-1.03)	64 (55-73)	64 (52-75)	1.00 (0.99-1.02)
Hospital ward (at infection onset)						
Emergency room	58/90 (64)[f]	216/360 (60)[f]	1.21 (0.75-1.96)	0/82 (0)[f]	1/328 (0)[f]	
Internal medicine	18/90 (20)[f]	78/360 (22)[f]	0.90 (0.51-1.61)	31/82 (38)[f]	193/328 (59)[f]	0.42 (0.26-0.69)
Surgery	11/90 (12)[f]	40/360 (11)[f]	1.11 (0.55-2.27)	33/82 (40)[f]	82/328 (25)[f]	2.01 (1.21-3.34)

Intensive care unit	3/90 (3)[f]	26/360 (7)[f]	0.44 (0.13-1.50)	18/82 (22)	52/328 (16)	1.49 (0.82-2.73)
Healthcare-associated infection	50/90 (56)[e]	141/353 (40)[e]	1.81 (1.13-2.89)	[g]	[g]	
Admission from long-term care facility	9/90 (10)	16/353 (4)	2.09 (0.89-4.95)	[g]	[g]	
Hospital admission (prior one year)	60/87 (69)	186/353 (53)	1.97 (1.20-3.23)	45/81 (56)	129/318 (41)	1.85 (1.13-3.02)
Length of hospital stay prior to infection in days, median (IQR)	[g]	[g]		20 (10-48)[e]	11 (6-19)[e]	1.03 (1.02-1.04)
Chronic pulmonary disease	8/90 (9)	68/358 (19)	0.42 (0.19-0.91)	10/81 (12)	39/328 (12)	1.09 (0.52-2.29)
Diabetes mellitus	28/90 (31)[e]	83/358 (23)[e]	1.48 (0.89-2.46)	16/81 (20)	62/328 (19)	1.10 (0.60-2.03)
Liver disease	2/90 (2)	5/358 (1)	1.42 (0.27-7.37)	4/81 (5)	4/328 (1)	4.62 (1.14-18.78)
Biliary tract disease	2/90 (2)	4/358 (1)	1.76 (0.32-9.83)	1/81 (1)	4/328 (1)	1.33 (0.15-11.43)
Any solid malignancy[h]	16/90 (18)	60/358 (17)	1.07 (0.58-1.97)	25/81 (31)[e]	70/328 (21)[e]	1.67 (0.97-2.87)

Without metastases	9/90 (10)[i]	34/358 (10)[i]	1.06 (0.49-2.30)	17/81 (21)[i]	45/328 (14)[i]	1.71 (0.92-3.18)
Metastasized	7/90 (8)[i]	26/358 (7)[i]	1.07 (0.45-2.55)	9/81 (11)[i]	25/328 (8)[i]	1.56 (0.70-3.49)
Haematological malignancy	11/90 (12)	28/358 (8)	1.62 (0.77-3.40)	9/81 (11)	44/328 (13)	0.85 (0.40-1.82)
Renal disease	13/90 (14)[e]	21/358 (6)[e]	2.54 (1.22-5.27)	14/81 (17)[e]	17/328 (5)[e]	3.98 (1.87-8.45)
Haemodialysis	1/90 (1)	5/353 (1)	0.55 (0.06-4.76)	[g]	[g]	
Immunocompromised[j]	27/87 (31)[e]	62/356 (17)[e]	2.03 (1.19-3.46)	16/80 (20)	76/323 (24)	0.85 (0.47-1.56)
Immunosuppressant use	23/90 (26)[i]	59/358 (16)[i]	1.71 (0.98-2.96)	16/81 (20)[i]	74/328 (23)[i]	0.89 (0.49-1.62)
Neutropenia (at infection onset)	7/87 (8)[i]	14/357 (4)[i]	2.09 (0.81-5.40)	5/81 (6)[i]	35/323 (11)[i]	0.53 (0.20-1.42)
Any transplant[h]	14/90 (16)[k]	22/358 (6)[k]	2.67 (1.31-5.45)	15/81 (18)[e]	23/327 (7)[e]	3.10 (1.54-6.23)
Solid organ transplant	11/90 (12)[i]	12/358 (3)[i]	3.71 (1.58-8.70)	9/81 (11)[i]	14/327 (4)[i]	2.93 (1.23-6.99)
Stem cell transplant	3/90 (3)	10/358 (3)	1.13 (0.30-4.21)	7/81 (9)[i]	9/327 (3)[i]	3.50 (1.26-9.68)

Urological patient[h]	25/90 (28)[e]	40/357 (11)[e]	2.96 (1.68-5.22)	5/81 (6)[k]	21/323 (6)[k]	1.05 (0.39-2.83)
Recurrent urinary tract infection	16/90 (18)[i]	25/358 (7)[i]	2.81 (1.43-5.53)	2/81 (2)	8/324 (2)	0.96 (0.20-4.63)
Obstructive urinary disease	5/90 (6)[i]	9/358 (2)[i]	2.13 (0.70-6.52)	0/81 (0)	6/328 (2)	Not available
Urological procedure (prior 30 days)	7/90 (8)[i]	7/357 (2)[i]	4.01 (1.36-11.79)	3/82 (4)	7/326 (2)	1.71 (0.43-6.77)
Surgical procedure (prior 30 days)	4/90 (4)	34/357 (10)	0.43 (0.15-1.24)	37/82 (45)[e]	116/327 (36)[e]	1.50 (0.92-2.46)
Endoscopic procedure (prior two days)	1/90 (1)	4/358 (1)	0.84 (0.09-7.60)	6/82 (7)	9/326 (3)	2.65 (0.92-7.66)
Central vascular catheter (at infection onset)	5/89 (6)	20/344 (6)	0.93 (0.34-2.55)	46/75 (61)[e]	106/299 (36)[e]	2.72 (1.62-4.57)
Urinary catheter (at infection onset)	22/88 (25)	61/342 (18)	1.47 (0.84-2.56)	38/71 (54)	142/291 (49)	1.21 (0.73-2.00)
Other catheter/drain (at infection onset)	4/90 (4)	15/347 (4)	0.89 (0.29-2.73)	17/74 (23)	72/300 (24)	0.99 (0.54-1.80)

Signs of hypoperfusion (at infection onset)	12/86 (14)	35/340 (10)	1.46 (0.73-2.93)	25/77 (32)[e]	38/296 (13)[e]	2.82 (1.57-5.06)
Suspected source of infection (at infection onset)						
Urinary tract infection or intra-abdominal infection[h]	55/90 (61)[k]	94/359 (26)[k]	4.44 (2.73-7.22)	26/80 (32)	46/325 (14)	3.00 (1.71-5.26)
Urinary tract infection	41/90 (46)[e]	48/359 (13)[e]	5.44 (3.25-9.11)	12/80 (15)[i]	20/325 (6)[i]	2.85 (1.35-6.04)
Intra-abdominal infection	14/90 (16)	46/359 (13)	1.26 (0.66-2.41)	14/80 (18)[i]	26/325 (8)[i]	2.42 (1.20-4.89)
Lower respiratory tract infection	8/90 (9)[e]	111/359 (31)[e]	0.22 (0.10-0.46)	4/80 (5)[e]	86/325 (26)[e]	0.14 (0.05-0.40)
Other infection	5/90 (6)	42/359 (12)	0.45 (0.17-1.16)	11/80 (14)	35/325 (11)	1.37 (0.66-2.85)
Unknown	22/90 (24)[l]	112/359 (31)[l]	0.71 (0.42-1.21)	39/80 (49)[l]	159/325 (49)[l]	0.98 (0.60-1.60)
Prior identification of 3GC-R EB (prior one year)	22/90 (24)[e]	9/359 (2)[e]	11.82 (5.25-26.63)	29/82 (35)[e]	16/328 (5)[e]	10.67 (5.41-21.03)

Any antibiotic use (prior two months)[h]	51/85 (60)[e]	140/346 (40)[e]	2.22 (1.37-3.60)	68/82 (83)	228/324 (70)	2.02 (1.08-3.77)
Cephalosporins or fluoroquinolones[h]	28/85 (33)[i]	66/346 (19)[i]	2.12 (1.26-3.55)	58/82 (71)	165/323 (51)	2.27 (1.34-3.84)
Cephalosporins	14/86 (16)[i]	33/351 (9)[i]	1.91 (0.99-3.68)	49/82 (60)[e]	114/322 (35)[e]	2.67 (1.62-4.39)
Fluoroquinolones	17/85 (20)[i]	44/346 (13)[i]	1.81 (0.98-3.35)	25/82 (30)	81/322 (25)	1.28 (0.75-2.18)
Carbapenems	4/86 (5)[i]	2/351 (1)[i]	4.95 (1.02-24.02)	12/82 (15)	29/321 (9)	1.66 (0.81-3.42)
Other beta-lactams	25/85 (29)[i]	72/345 (21)[i]	1.65 (0.97-2.80)	29/82 (35)	110/320 (34)	1.04 (0.62-1.72)
Aminoglycosides, macrolides or other antibiotics[h]	33/85 (39)[i]	73/345 (21)[i]	2.31 (1.39-3.84)	56/82 (68)[k]	131/323 (41)[k]	3.11 (1.85-5.21)
Aminoglycosides	4/86 (5)[i]	13/351 (4)[i]	1.21 (0.40-3.67)	13/81 (16)	35/319 (11)	1.49 (0.75-2.98)
Macrolides	3/86 (4)[i]	18/347 (5)[i]	0.75 (0.23-2.44)	17/81 (21)	37/320 (12)	2.01 (1.06-3.82)
Other antibiotics	29/85 (34)[i]	57/345 (16)[i]	2.57 (1.51-4.39)	49/82 (60)	98/323 (30)	3.38 (2.04-5.58)
Selective digestive/oropharyngeal	1/86 (1)[k]	2/351 (1)[k]	1.63 (0.24-11.12)	10/82 (12)	26/325 (8)	1.56 (0.72-3.40)

decontamination (prior two months)						
At risk of 3GC-R EB bacteraemia according to *two-predictor model*[m]	46/86 (54)[n]	71/347 (20)[n]	4.32 (2.63-7.09)	65/82 (79)[n]	168/323 (52)[n]	3.46 (1.94-6.17)

Abbreviations: OR, odds ratio; CI: confidence interval, IQR, interquartile range.

[a] See Supplementary Table 3 for definitions used.

[b] Patients with 3GC-R EB bacteraemia.

[c] Sample of patients with non-resistant Enterobacteriaceae bacteraemia, bacteraemia with other causative pathogens, or negative blood cultures.

[d] OR from 20 imputed datasets, combined by means of Rubin's rules.

[e] Predictor selected by means of eyeballing association with outcome.

[f] Predictor not considered for model construction purposes (and hence not used in forward stepwise regression analysis for model variant 2 and univariable preselection during the bootstrapping procedure) because of expected problems in generalization to other settings.

[g] Predictor not recorded for this setting.

[h] Aggregated variable combining indented variables below.

[i] Predictor not considered for model construction purposes (see f for implications) because of aggregation.

[j] Aggregated variable combining *immunosuppressant use*, *neutropenia (at infection onset)*, and *solid organ transplant*.

[k] Predictor only shown for comparison with other cohort and not considered for model construction purposes (see f for implications).

[l] Predictor not considered for model construction purposes (see f for implications) because it was used as reference category.

[m] Aggregated variable combining *use of cephalosporins or fluoroquinolones (prior two months)*, and *prior identification of 3GC-R EB (prior one year)*.

[n] Predictor only shown to evaluate performance of *two-predictor model* and not considered for model construction purposes (see f for implications).

Table 3 Simplified final model for community-onset infection

Predictor	Original model		Optimism-corrected model[a]	
	β coefficient	OR (95% CI)	β coefficient	OR (95% CI)
Intercept	-7.632		-7.248	
Prior identification of 3GC-R EB (prior one year)	2.355	10.53 (4.26-26.08)	1.963	7.12 (2.88-17.62)
Suspected source of infection: Urinary tract infection	1.297	3.66 (2.04-6.57)	1.081	2.95 (1.64-5.29)
Immunocompromised	0.590	1.80 (0.96-3.39)	0.491	1.63 (0.87-3.08)
Any use of antibiotics (prior two months)	0.377	1.46 (0.83-2.55)	0.314	1.37 (0.78-2.39)
Age (per year increase)	0.022	1.02 (1.01-1.04)	0.018	1.02 (1.00-1.04)
Suspected source of infection: Lower respiratory tract infection	-1.075	0.34 (0.15-0.78)	-0.896	0.41 (0.18-0.94)

The optimism-corrected predicted probability of 3GC-R EB bacteraemia can be calculated with the following formula: 1/(1 + exp(-(-7.248 + 1.963 x *prior identification of 3GC-R EB (prior one year)* + 1.081 x *suspected source of infection: urinary tract infection* + 0.491 x *immunocompromised* + 0.314 x *any use of*

antibiotics (prior two months) + 0.018 x *age in years* - 0.896 x *suspected source of infection: lower respiratory tract infection*))). For categorical predictors, fill in 1 if present, and 0 if absent.

Abbreviations: OR, odds ratio; CI, confidence interval.

[a] Derived by multiplication with a shrinkage factor (0.834) obtained by bootstrapping described in Supplementary Material, followed by re-estimation of the intercept and correction for the sampling fraction of controls to match overall predicted incidence by the model with observed incidence.

Table 4 Derived scoring systems

Predictor	Score in community-onset infection	Score in hospital-onset infection
Renal disease		120
Prior identification of 3GC-R EB (prior one year)	100	120
Any solid malignancy		80
Suspected source of infection: Urinary tract infection	50	
Signs of hypoperfusion (at infection onset)		40
Surgical procedure (prior 30 days)		40
Central vascular catheter (at infection onset)		40
Use of cephalosporins (prior two months)		40
Immunocompromised	25	
Any use of antibiotics (prior two months)	25	
Age (per year)	1	

Length of hospital stay prior to infection (per day)		1
Suspected source of infection: Lower respiratory tract infection	-50	-160

For example, a 75-year-old immunocompromised patient who has community-onset infection, suspected to be pneumonia, and whose previous cultures do not show any 3GC-R EB, scores 75 for age, -50 for the suspected pneumonia, and 25 for being immunocompromised, i.e. a total of 50, in the community-onset scoring system.

Table 5 Simplified final model for hospital-onset infection

Predictor	Original model		Optimism-corrected model[a]	
	β coefficient	OR (95% CI)	β coefficient	OR (95% CI)
Intercept	-6.210		-5.807	
Renal disease	1.743	5.71 (2.24-14.55)	1.372	3.94 (1.55-10.05)
Prior identification of 3GC-R EB (prior one year)	1.718	5.57 (2.41-12.89)	1.353	3.87 (1.67-8.95)
Any solid malignancy	0.917	2.50 (1.29-4.87)	0.722	2.06 (1.06-4.01)
Signs of hypoperfusion (at infection onset)	0.646	1.91 (0.91-4.01)	0.509	1.66 (0.79-3.49)
Surgical procedure (prior 30 days)	0.564	1.76 (0.94-3.28)	0.444	1.56 (0.84-2.91)
Central vascular catheter (at infection onset)	0.533	1.70 (0.88-3.31)	0.420	1.52 (0.78-2.95)
Use of cephalosporins (prior two months)	0.527	1.69 (0.90-3.17)	0.415	1.51 (0.81-2.83)

Length of hospital stay prior to infection (per day increase)	0.014	1.01 (1.00-1.03)	0.011	1.01 (1.00-1.03)
Suspected source of infection: Lower respiratory tract infection	-2.196	0.11 (0.04-0.35)	-1.729	0.18 (0.06-0.56)

The optimism-corrected predicted probability of 3GC-R EB bacteraemia can be calculated with the following formula: 1/(1 + exp(-(-5.807 + 1.372 x *renal disease* + 1.353 x *prior identification of 3GC-R EB (prior one year)* + 0.722 x *any solid malignancy* + 0.509 x *signs of hypoperfusion (at infection onset)* + 0.444 x *surgical procedure (prior 30 days)* + 0.420 x *central vascular catheter (at infection onset)* + 0.415 x *use of cephalosporins (prior two months)* + 0.011 x *length of hospital stay prior to infection in days* - 1.729 x *suspected source of infection: lower respiratory tract infection*))). For categorical predictors, fill in 1 if present, and 0 if absent.

Abbreviations: OR, odds ratio; CI: confidence interval.

[a] Derived by multiplication with a shrinkage factor (0.788) obtained by bootstrapping described in Supplementary Material, followed by re-estimation of the intercept and correction for the sampling fraction of controls to match overall predicted incidence by the model with observed incidence.

Table 6 Expected optimism when selecting probability cutoffs based on performance of *two-predictor model*

	Community-onset infection			Hospital-onset infection		
	Two-predictor model	Cutoff 1 for final simplified model[a]	Cutoff 2 for final simplified model[b]	Two-predictor model	Cutoff 1 for final simplified model[a]	Cutoff 2 for final simplified model[b]
Apparent performance in study sample						
Sensitivity[c] (95% CI)	53.9% (44.2-63.9%)	55.2% (43.7-63.7%)	68.3% (58.2-78.2%)	79.3% (70.7-87.8%)	80.6% (71.8-88.8%)	91.6% (86.6-97.6%)
Prevalence[d] (95% CI)	21.5% (17.3-25.8%)	12.8% (9.8-16.7%)	21.0% (16.9-25.4%)	52.8% (47.3-57.9%)	27.6% (22.6-32.1%)	52.3% (47.3-57.6%)
Optimism-corrected performance[e]						

Sensitivity[c] (95% CI)	53.9%[f] (44.2-63.9%)	49.0% (32.3-62.2%)	64.3% (50.3-77.8%)	79.3%[e] (70.7-87.8%)	75.3% (61.8-86.6%)	89.8% (82.0-98.9%)
Prevalence[d] (95% CI)	21.5%[f] (17.3-25.8%)	13.2% (6.9-18.6%)	22.5% (16.5-29.2%)	52.8%[e] (47.3-57.9%)	28.2% (18.7-35.0%)	53.8% (46.6-62.8%)

Abbreviations: CI, confidence interval.

[a] Cutoff 1 (above which patients are classified as 'at risk of 3GC-R EB bacteraemia') is chosen such that the resulting sensitivity is as close as possible to the sensitivity of the *two-predictor model*. In community-onset infection, this cutoff (mean value of 20 imputed datasets) was 0.67%, and, in hospital-onset infection, 0.86%.

[b] Cutoff 2 (above which patients are classified as 'at risk of 3GC-R EB bacteraemia') is chosen such that the resulting prevalence is as close as possible to the prevalence of the *two-predictor model*. In community-onset infection, this cutoff (mean value of 20 imputed datasets) was 0.42%, and, in hospital-onset infection, 0.40%.

[c] Proportion of patients with 3GC-R EB bacteraemias categorized as 'at risk of 3GC-R EB bacteraemia'; mean value of 20 imputed datasets

[d] Proportion of total population categorized as 'at risk of 3GC-R EB bacteraemia'; mean value of 20 imputed datasets

[e] As obtained by bootstrapping described in Supplementary Material.

[f] Not affected by optimism due to pre-specification of models.

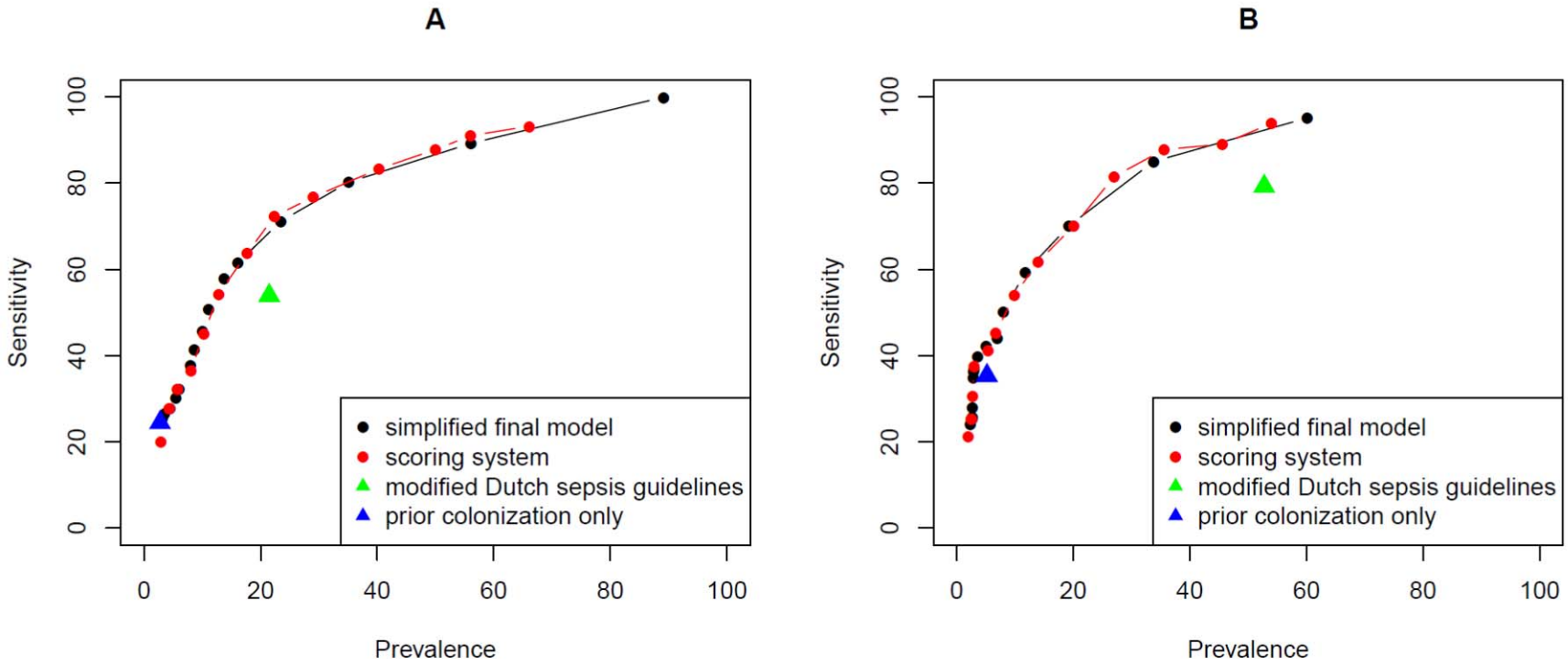

Figure 1 Comparison of performance at different cutoffs of simplified final community-onset (A) and hospital-onset (B) models and derived scoring systems

Figures show association between sensitivity (proportion of patients with 3GC-R EB bacteraemias categorized as 'at risk') and prevalence (proportion of total population categorized as 'at risk') when moving cutoffs above which patients are categorized as 'at risk of 3GC-R EB bacteraemia', both for probabilities resulting from the *simplified final models*, and for the derived scoring systems. Performance is compared to two basic models (*two-predictor model* and *prior*

identification model; see Table 1 for definition). As prevalence approximates 1 - specificity in this population, figures are identical to receiver operating characteristic (ROC) curves. Results are mean values of 20 imputed datasets. See Supplementary Tables 8, 9, 12 and 13 for exact values.

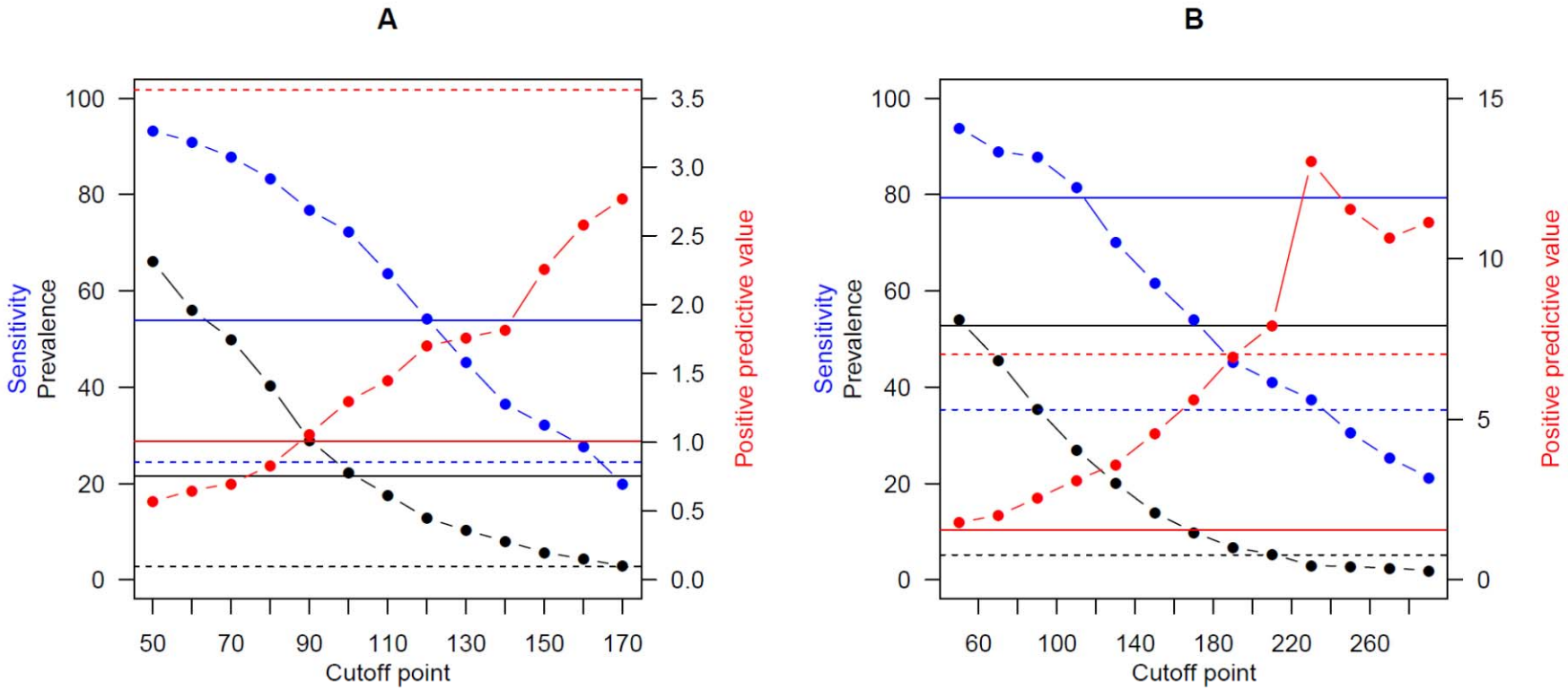

Figure 2 Performance of community-onset (A) and hospital-onset (B) scoring systems at different cutoff values

Figures show sensitivity (proportion of patients with 3GC-R EB bacteraemias categorized as 'at risk'; in blue), prevalence (proportion of total population categorized as 'at risk'; in black), and positive predictive value (proportion of patients with 3GC-R EB bacteraemias among those categorized as 'at risk') at different cutoffs for derived scoring systems above which patients are categorized as 'at risk of 3GC-R EB bacteraemia'. These are compared to the

(constant) sensitivities, prevalences, and positive predictive values for the basic *two-predictor model* (solid lines) and *prior identification model* (dashed lines) (see Table 1 for definition). Results are mean values of 20 imputed datasets. See Supplementary Tables 9 and 13 for exact values.