MASTER: a model to improve and standardize clinical breakpoints for antimicrobial susceptibility testing using forecast probabilities

Blöchliger, Nicolas ; Keller, Peter M ; Böttger, Erik C ; Hombach, Michael

DOI: https://doi.org/10.1093/jac/dkx196

J Antimicrob Chemother 2017; **72**: 2553–2561
doi:10.1093/jac/dkx196 Advance Access publication 6 July 2017

Journal of Antimicrobial Chemotherapy

MASTER: a model to improve and standardize clinical breakpoints for antimicrobial susceptibility testing using forecast probabilities

Nicolas Blöchliger, Peter M. Keller, Erik C. Böttger and Michael Hombach*

Institute of Medical Microbiology, University of Zurich, Gloriastrasse 30/32, 8006 Zürich, Switzerland

*Corresponding author. Present address: Roche Diagnostics International, Forrenstrasse 2, 6343 Rotkreuz, Switzerland.
E-mail: michael.hombach@roche.com

Received 29 November 2016; returned 23 January 2017; revised 7 May 2017; accepted 24 May 2017

Objectives: The procedure for setting clinical breakpoints (CBPs) for antimicrobial susceptibility has been poorly standardized with respect to population data, pharmacokinetic parameters and clinical outcome. Tools to standardize CBP setting could result in improved antibiogram forecast probabilities. We propose a model to estimate probabilities for methodological categorization errors and defined zones of methodological uncertainty (ZMUs), i.e. ranges of zone diameters that cannot reliably be classified. The impact of ZMUs on methodological error rates was used for CBP optimization.

Methods: The model distinguishes theoretical true inhibition zone diameters from observed diameters, which suffer from methodological variation. True diameter distributions are described with a normal mixture model. The model was fitted to observed inhibition zone diameters of clinical *Escherichia coli* strains. Repeated measurements for a quality control strain were used to quantify methodological variation.

Results: For 9 of 13 antibiotics analysed, our model predicted error rates of <0.1% applying current EUCAST CBPs. Error rates were >0.1% for ampicillin, cefoxitin, cefuroxime and amoxicillin/clavulanic acid. Increasing the susceptible CBP (cefoxitin) and introducing ZMUs (ampicillin, cefuroxime, amoxicillin/clavulanic acid) decreased error rates to <0.1%. ZMUs contained low numbers of isolates for ampicillin and cefuroxime (3% and 6%), whereas the ZMU for amoxicillin/clavulanic acid contained 41% of all isolates and was considered not practical.

Conclusions: We demonstrate that CBPs can be improved and standardized by minimizing methodological categorization error rates. ZMUs may be introduced if an intermediate zone is not appropriate for pharmacokinetic/pharmacodynamic or drug dosing reasons. Optimized CBPs will provide a standardized antibiotic susceptibility testing interpretation at a defined level of probability.

Introduction

Antibiotic susceptibility testing (AST) results are used by clinicians as a guide to select the most appropriate drugs in the treatment of patients suffering from infectious diseases. In general, the laboratory does not provide raw data on AST, such as inhibition zone diameters, but the results are categorized using clinical breakpoints (CBPs) according to the expected clinical outcome.[1,2] Although this approach is helpful in daily clinical practice, it leads to a major loss of information.

CBP setting has traditionally been a poorly standardized procedure: susceptibility data on pathogens, pharmacodynamic/pharmacokinetic values and clinical outcome studies are taken into account, but a set of rules that would make CBP categorization reliable has not yet been established.[1,3] In contrast, in clinical chemistry raw results are transmitted to clinicians together with reference ranges for normal values for the patient's population, allowing clinicians to adequately judge the deviation of the measured value from the reference cut-off.[4]

Clinical categories in AST that are currently delineated by CBPs are 'susceptible', which is associated with clinical success, 'resistant', which predicts clinical failure, and 'intermediate', which is used in an ambiguous way. Originally, the EUCAST definition contained two aspects of the intermediate category, i.e. it 'implied that an infection due to the isolate may be appropriately treated in body sites where the drugs are physiologically concentrated or when a high dosage of drug can be used' and 'it also indicated a buffer zone that should prevent small, uncontrolled, technical factors from causing major discrepancies in interpretations.'[5] The use of a dual definition for the intermediate category resulted in confusion and hampered understanding of the primary purpose of AST reports, i.e. allowing clinicians to select the most

appropriate drug for antimicrobial therapy. EUCAST recently proposed a re-definition of the intermediate category, removing the concept of a technical buffer zone.[5] The intermediate category is now suggested to be defined as 'a level of antimicrobial activity associated with a high likelihood of therapeutic success but only when a higher dosage of the agent than normal can be used or when the agent is physiologically concentrated at the site of infection.' However, the methodological variation of AST remains and is no longer addressed by intermediate categorization, nor is it indicated on AST reports. Therefore, major erroneous forecasts of therapeutic outcome success are likely.[6–8]

Tools that optimize and harmonize CBPs by reducing error rates seem beneficial for antibiogram reliability and may provide the basis for standardizing the process of CBP setting.[9] If acceptable categorization error rates cannot be achieved by adapting CBPs, the implementation of zones of methodological uncertainty (ZMUs) may be helpful to indicate less reliable measurements, avoiding the term 'intermediate' and the associated confusion.[8]

In this study we present a mathematical tool to calculate the rates of misclassifications as far as they are due to methodological variation. This tool, named MASTER (Model for AST Error Rates), takes into account technical variation, as well as biological and epidemiological factors. It is therefore suitable for studying the effect of CBP changes and/or introduction of ZMUs in individual epidemiological settings and in aggregated datasets. For proof of principle we applied MASTER to systematically evaluate the current CBPs issued by EUCAST using a broad set of inhibition zone diameters of non-duplicate, non-outbreak clinical *Escherichia coli* strains originating from our clinical laboratory.

Methods

Model for true and observed inhibition zone diameters

We developed MASTER to determine the methodological error for categorization of Kirby–Bauer inhibition zone diameters. MASTER distinguishes between the experimentally observed inhibition zone diameter Y, which suffers from methodological variation, and the theoretical true diameter X, which is not observed. The relationship between X and Y is given by $Y = X + E$, where E models the methodological variation.

We modelled the distribution of the true diameter X as a mixture of $n \geq 2$ components with weights w_i satisfying $w_i \geq 0$ and $\sum_{i=1}^{n} w_i = 1$. The true diameter is 6 mm for the first component, which therefore captures all the strains whose growth is not affected by the antibiotic. For the other components the true diameter is normally distributed with mean μ_i and variance σ_i^2. The cumulative distribution function of X thus has the form:

$$F_X(x) = w_1 H(x - 6\,\text{mm}) + \sum_{i=2}^{n} w_i \Phi\left(\frac{x - \mu_i}{\sigma_i}\right),$$

where H is the Heaviside step function and Φ is the cumulative distribution function of the standard normal distribution.

The true diameter X is unknown, and we observe $Y = X + E$ instead. Our model for the methodological variation E depends on the component. The variation associated with the first component is assumed to be zero. For the other components we assume E to be normally distributed (Figures S1 and S2, available as Supplementary data at *JAC* Online) and independent of X with zero mean and variance σ_E^2 (Figure S3 and Table S1).

Under these assumptions regarding X and E the cumulative distribution function of Y is given by:

$$F_Y(y) = w_1 H(y - 6\,\text{mm}) + \sum_{i=2}^{n} w_i \Phi\left(\frac{y - \mu_i}{\sqrt{\sigma_i^2 + \sigma_E^2}}\right).$$

Note that we do not account for the fact that the observed diameters are typically rounded to integer values.

Probabilities of methodological misclassification errors

Misclassification errors are traditionally split according to their therapeutic implications. Erroneous categorizations of true-susceptible isolates as resistant are referred to as major errors (MEs) and lead to unnecessary restriction of therapeutic options. The most serious clinical implications result from very major errors (VMEs), i.e. categorizations of true-resistant isolates as susceptible, as there is a high likelihood of therapeutic failure. Minor errors (mEs), i.e. misclassifications of true-intermediate isolates as resistant or as susceptible or of true-resistant or true-susceptible isolates as intermediate, have limited therapeutic consequences.[10] Based on a pair of CBPs $z_S \geq z_R > 6$ mm, strains are defined to be resistant if the true diameter satisfies $x < z_R$, intermediate if $z_R \leq x < z_S$ and susceptible if $z_S \leq x$. However, the prediction is based on the observed diameter Y, which due to the methodological variation E inevitably results in misclassification errors. The probability that a very major methodological misclassification error occurs is, based on MASTER:

$$p_{VME} = p(X < z_R \text{ and } Y \geq z_S) = \int_{-\infty}^{z_R} \int_{z_S}^{\infty} g(x,y)\,\mathrm{d}y\,\mathrm{d}x,$$

where:

$$\begin{aligned} g(x,\ y) &= p(X = x \text{ and } Y = y) \\ &= \left(\sum_{i=2}^{n} w_i \varphi\left(x; \mu_i, \sigma_i^2\right)\right)\ \varphi\left(y - x; 0, \sigma_E^2\right) \end{aligned}$$

and $\varphi(\cdot; \mu, \sigma^2)$ is the density function of a normally distributed random variable with mean μ and variance σ^2. Similarly, the probability of a major methodological misclassification error is:

$$p_{ME} = p(X \geq z_S \text{ and } Y < z_R) = \int_{z_S}^{\infty} \int_{-\infty}^{z_R} g(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

and the probability of a minor methodological misclassification error is:

$$\begin{aligned} p_{mE} &= p(X < z_R \text{ and } z_R \leq Y < z_S) + p(z_R \leq X < z_S \text{ and } Y < z_R) \\ &+ p(z_R \leq X < z_S \text{ and } z_S \leq Y) + p(\ z_S \leq X \text{ and } z_R \leq Y < z_S) \\ &= \int_{-\infty}^{z_R} \int_{z_R}^{z_S} g(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_{z_R}^{z_S} \int_{-\infty}^{z_R} g(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_{z_R}^{z_S} \int_{z_S}^{\infty} g(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_{z_S}^{\infty} \int_{z_R}^{z_S} g(x,y)\,\mathrm{d}y\,\mathrm{d}x. \end{aligned}$$

In the present study we used the CBPs issued by EUCAST throughout.[11]

Forecast probabilities

Given an observed diameter y, what is the probability that no ME or VME occurred? This probability, which is a function of y, is the quantity of

principal interest in the present work. Under the model developed above, the probability of a very major methodological misclassification error given y is:

$$p_{VME}(y) = \begin{cases} 0, \text{ if } y < z_S, \\ p(X < z_R | Y = y) = \dfrac{\int_{-\infty}^{z_R} g(x,y)\mathrm{d}x}{w_1 I(y - 6\,\text{mm}) + \sum_{j=2}^{n} w_j \varphi\left(y; \mu_j, \sigma_j^2 + \sigma_E^2\right)}, \text{ else,} \end{cases}$$

where:

$$I(y) = \begin{cases} 1, \text{ if } y = 0, \\ 0, \text{ else.} \end{cases}$$

Similarly, the probability of a major methodological misclassification error given y is:

$$p_{ME}(y) = \begin{cases} p(X \geq z_S | Y = y) = \dfrac{\int_{z_S}^{\infty} g(x,y)\mathrm{d}x}{w_1 I(y - 6\,\text{mm}) + \sum_{j=2}^{n} w_j \varphi\left(y; \mu_j, \sigma_j^2 + \sigma_E^2\right)}, \text{ if } y < z_R, \\ 0, \text{ else,} \end{cases}$$

The forecast probability, i.e. the probability given y that no major or very major misclassification error occurs due to methodological variation, is then simply:

$$p_f(y) = 1 - p_{VME}(y) - p_{ME}(y).$$

ZMU

We can identify inhibition zone diameters for which the AST categorization is ambiguous due to methodological variation based on the forecast probabilities $p_f(y)$. In particular, we define the ZMU to encompass the observed diameters that satisfy $p_f(y) < 0.99$ or $z_R \leq y < z_S$. The ZMU contains all the observed diameters for which the risk of an ME or VME is >1% or which are in the intermediate zone.

Clinical isolates

Antimicrobial susceptibility data on non-duplicate, non-outbreak clinical *E. coli* strains were used in this study. Clinical isolates included were collected over a 5 year period from 2010 to 2014 in the clinical microbiology laboratory of the Institute of Medical Microbiology, University of Zurich. The majority of the isolates were isolated from specimens of the University Hospital of Zurich, an 850 bed tertiary care hospital.

AST

AST was performed by the disc diffusion method following EUCAST standard procedures[12] on Mueller–Hinton II agar (Becton-Dickinson, Franklin Lakes, NJ, USA), and with antibiotic discs from i2a (Montpellier, France). Inhibition zone diameters were automatically recorded using the Sirscan/Sirweb system (i2a).[13] The automated zone reader had been previously calibrated using EUCAST quality control (QC) strains and tables, and all technical settings had been parameterized to meet EUCAST QC target and range requirements. Furthermore, calibration to EUCAST QC requirements was assured by running weekly QC strains. In addition, isolates with inhibition zone sizes in the left tail area (i.e. a range of 4 mm starting at the left end) of the distribution curves were verified by visual inspection of plate images. Adjustments, if necessary, were done on-screen, and the adjusted result was used in the analysis. Isolates were eliminated from the data set if errors were found [e.g. contaminated plates, false identification, technically failed measurements of inhibition zones (i.e. the software was clearly not measuring the zone of inhibition as the digital calliper was visually distant from the zone edge), incompatibility with WT definition within an antibiotic class (i.e. if the expected susceptibility pattern for a given WT was inconsistent, e.g. norfloxacin susceptible, levofloxacin resistant)]. Isolates were considered as duplicates and excluded from the analysis if they originated from the same patient and showed at most one major and two minor differences in AST interpretation.[14] Table 1 lists the antibiotics for which AST was performed and CBPs were available. Our data set comprised inhibition zone diameters from 9766 strains for β-lactams, from 9761 strains for quinolones, from 3521 strains for aminoglycosides and from 1089 strains for tigecycline.

Model fitting

We fitted our model to measured inhibition zone diameters as follows. First, we estimated σ_E^2 (the variance of the methodological variation) based on 55 independent repeated measurements of inhibition zone diameters for the QC strain *E. coli* ATCC 25922 using different lots of antibiotic discs and different lots of Mueller–Hinton agar plates from the same manufacturer. Experiments were performed and read by different experienced persons on different days to ensure that the variation closely resembled routine conditions in the clinical laboratory.

Second, we used AST data to fit the parameters describing the distribution of the observed diameter Y. In particular, we estimated w_1 as the fraction of data points in the sample that are equal to 6 mm. The parameters of the normally distributed components of Y, i.e. w_i, μ_i and $\sigma_i^2 + \sigma_E^2$, were estimated by fitting a normal mixture model of $n - 1$ components to the data with observed diameters >6 mm using the expectation-maximization algorithm.[15] Models were generated for $n = 2, \ldots, 10$, and the best one was selected based on the Bayesian information criterion[16,17] after excluding models that did not satisfy $\sigma_i > 0$, i.e. models that had a component with variance below σ_E^2. Our data set did not contain any strains with diameters below the CBP defining resistance for imipenem and tigecycline. We thus refrained from applying our model to these two antibiotics.

We used quantile–quantile (Q–Q) plots to compare the empirical distribution of inhibition zone diameters with the fit from our model.[18] Briefly, Q–Q plots compare the quantiles of two distributions and are thus a graphical means of assessing similarity between them. If two distributions agree perfectly, their quantiles coincide and the plotted points lie on the identity line.

Software

All computations were performed with the free software R, version 3.2.3.[19] Normal mixture models were fitted using the R package mclust, version 5.2.[20]

Results

Model fitting

Repeated measurements of inhibition zone diameters for the QC strain *E. coli* ATCC 25922 were approximately normally distributed (Figures S1 and S2) and were in agreement with the target ranges published by EUCAST.[21] Standard deviations for *E. coli* ATCC 25922 ranged from 1.2 to 2.0 mm (Table 1). Comparison with clinical strains suggested modelling the methodological variation as approximately constant for diameters >6 mm (Figure S3 and Table S1).

Figure 1 visualizes our model. The Q–Q plots showed that MASTER described the empirical distributions well. The model captured accumulation of diameters of 6 mm and was sufficiently complex to capture non-normal distributions for WT strains (e.g. cefotaxime, Figure 1f). Differences between fitted model and

Table 1. Estimated standard deviation of the methodological variation derived from repeated measurements for the QC strain *E. coli* ATCC 25922, number of components in the mixture model, Kolmogorov–Smirnov (K–S) distance between observed distribution of inhibition zone diameters and fitted distribution, official CBPs (EUCAST)[11] and ZMUs derived from MASTER

				CBP (mm)		ZMU[a]			
Antibiotic	σ_E (mm)	Components (*n*)	K–S distance	R	S	min (mm)	max (mm)	width (mm)	weight (%)
Ampicillin	1.4	6	0.03	14	14	11	16	5	3
Cefoxitin	1.4	3	0.08	19	19	15	21	6	6
Amoxicillin/clavulanic acid	1.5	3	0.06	19	19	15	22	7	41
Piperacillin/tazobactam	1.3	4	0.07	17	20	16	20	4	5
Cefuroxime	1.4	3	0.07	18	18	14	20	6	4
Cefotaxime	1.5	5	0.06	17	20	17	21	4	1
Ceftazidime	1.4	4	0.06	19	22	18	22	4	2
Ceftriaxone	1.8	3	0.07	20	23	19	23	4	1
Cefepime	1.5	5	0.06	21	24	20	24	4	2
Ertapenem	1.2	6	0.05	22	25	22	25	3	1
Imipenem[b]	1.3	–	–	16	22	–	–	–	–
Meropenem	1.5	4	0.06	16	22	15	22	7	0
Gentamicin	1.2	3	0.14	14	17	14	17	3	1
Tobramycin	1.4	3	0.11	14	17	14	17	3	2
Norfloxacin	1.9	4	0.04	19	22	17	22	5	1
Ciprofloxacin	2	3	0.07	19	22	18	23	5	1
Levofloxacin	1.9	5	0.03	19	22	17	22	5	2
Tigecycline[b]	1.2	–	–	15	18	–	–	–	–

R, resistant (i.e. a high likelihood of therapeutic failure); S, susceptible (i.e. a high likelihood of therapeutic success).
[a]We set min and max such that the ZMU is $\{y : \min \leq y < \max\}$. If the ZMU coincides with the intermediate zone, min and max are equal to the CBPs. The weight of the ZMU is defined as the proportion of strains with observed diameters in the ZMU.
[b]The data set does not contain any strains with diameters below the CBP defining resistance for imipenem and tigecycline. We thus refrained from applying MASTER to these two antibiotics.

experimental data were most pronounced for large inhibition zone diameters (Figure 1). For cefotaxime, cefoxitin, ceftriaxone and amoxicillin/clavulanic acid, the Q–Q plots also highlighted discrepancies for low inhibition zones despite the fact that the mixture model had a separate component for diameters of 6 mm. The Q–Q plots indicate that the distributions of diameters were difficult to fit for ertapenem, meropenem and gentamicin (Figures 1j, k and l). For ertapenem and meropenem, this reflects the small number of strains with observed diameters below the CBPs defining resistance (53 and 3 strains out of 9766 for ertapenem and meropenem, respectively). The observed lack of fit disappeared if the clinical data were extended by 200 simulated resistant isolates (Figure S4). For gentamicin, the main peak of the distribution (centred at 22.5 mm, i.e. the part of the distribution encompassing the WT strains) seemed too heavy-tailed to be described by a normal distribution.

Forecast probabilities and ZMUs

The forecast probability, i.e. the probability that no MEs or VMEs occurs due to methodological variation, is shown in Figure 1. We emphasized that the forecast probability, as defined here, is a function of the observed diameter, i.e. it reflects how far a specific measurement is from the CBPs. Indeed, the forecast probability was 1 within the intermediate zone and for observed diameters sufficiently distant from the CBPs. However, the forecast probability was <1 in the immediate vicinity of the intermediate zone or close to the CBP in settings without an intermediate zone (ampicillin, cefoxitin, amoxicillin/clavulanic acid and cefuroxime). A forecast probability different from 1 can lead to major or very major misclassification errors.

We use the term ZMU for the diameters in the intermediate zone or with a forecast probability <0.99. The median width of the ZMU was 4.5 mm. For antibiotics without an intermediate zone defined by EUCAST, the width of the ZMU varied between 5 and 7 mm (Table 1).[11] The ZMU coincided with the intermediate zone for ertapenem, gentamicin and tobramycin. For the remaining antibiotics, the ZMU extended the intermediate zone by 1 or 2 mm. MASTER predicted that <5% of all strains have observed diameters within the ZMU for most antibiotics. The exceptions were amoxicillin/clavulanic acid (41%), cefoxitin (6%) and piperacillin/tazobactam (5%).

Rates of methodological misclassification errors

MASTER predicted ME and VME rates <0.1% for 9 out of the 13 drugs analysed, based on current EUCAST CBPs for most antibiotics (Table 2). Ampicillin, cefoxitin and cefuroxime had ME and VME rates between 0.1% and 1%, and amoxicillin/clavulanic acid had ME and VME rates >1%. Of note, the CBP for amoxicillin/clavulanic acid is located within the admissible EUCAST range for QC strain *E. coli* ATCC 25922, which is considered a typical representative of the species (Figure S1). The mE rates were between 1% and 3% for piperacillin/tazobactam, ceftazidime, cefepime and tobramycin. The remaining drugs had mE rates between 0% and 1%.

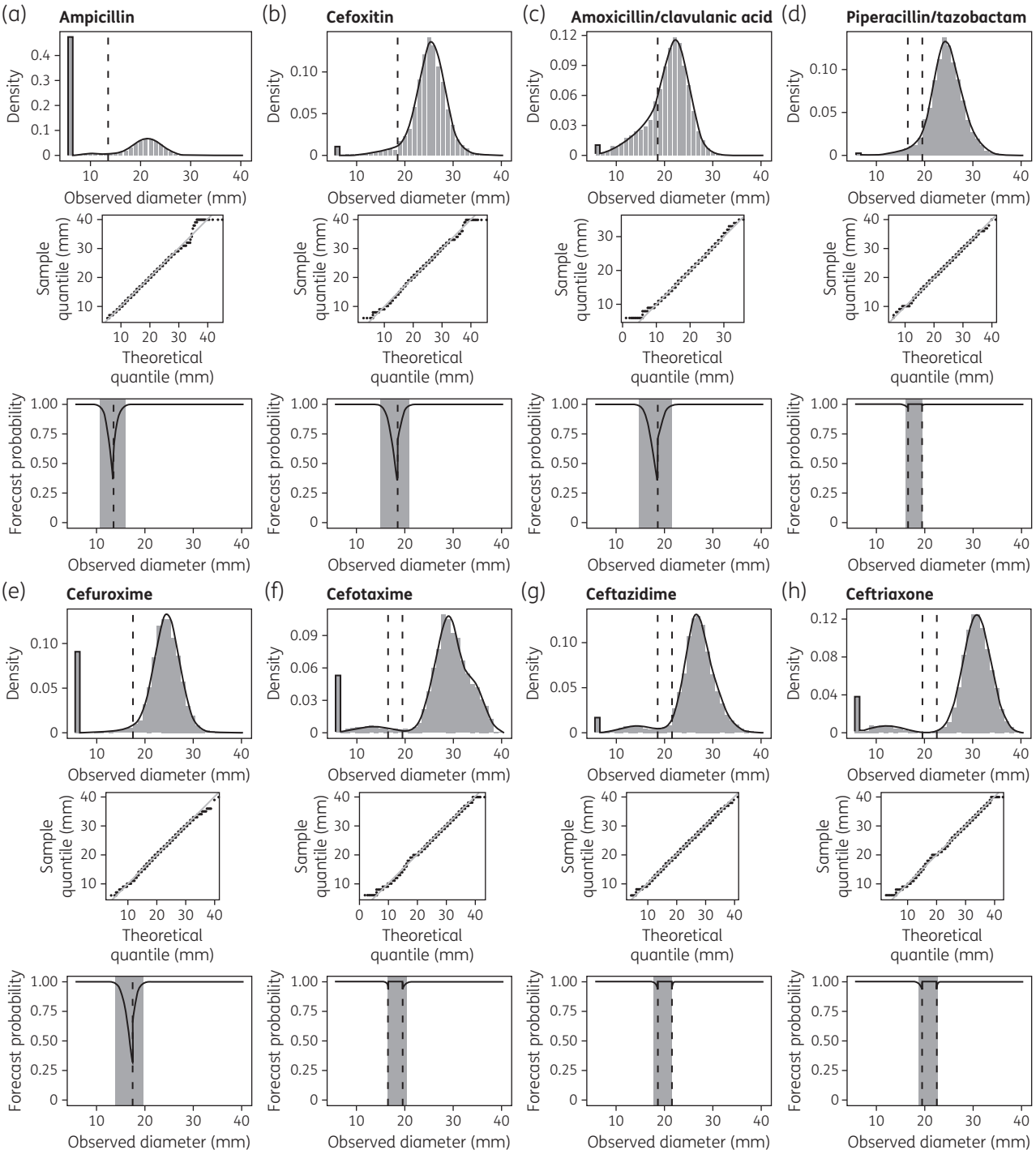

Figure 1. Empirical (histogram) and fitted (black line) distributions for the observed inhibition zone diameters (top), Q–Q plots with identity lines in grey (middle) and forecast probabilities and ZMUs in grey (bottom). CBPs by EUCAST are indicated with dashed lines.[11]

ZMUs reduced ME and VME rates to <0.01% for all drugs (Table 2), and led to mE rates >1% for the drugs without an intermediate zone defined by EUCAST (ampicillin, cefoxitin, amoxicillin/clavulanic acid and cefuroxime). For the remaining drugs, mE rates did not change or increased by a factor of at most 2. The mE rate was highest for amoxicillin/clavulanic acid (20%).

Increasing the CBPs defining resistance by 2 mm reduced ME and VME rates for all drugs (Table 2). For cefoxitin, amoxicillin/clavulanic acid and cefuroxime the ME or VME rates remained >0.1%. The mE rates were >1% for drugs without an intermediate zone. For the remaining drugs, mE rates increased by a factor of at most 5. The mE rate was highest for amoxicillin/clavulanic acid (20%).

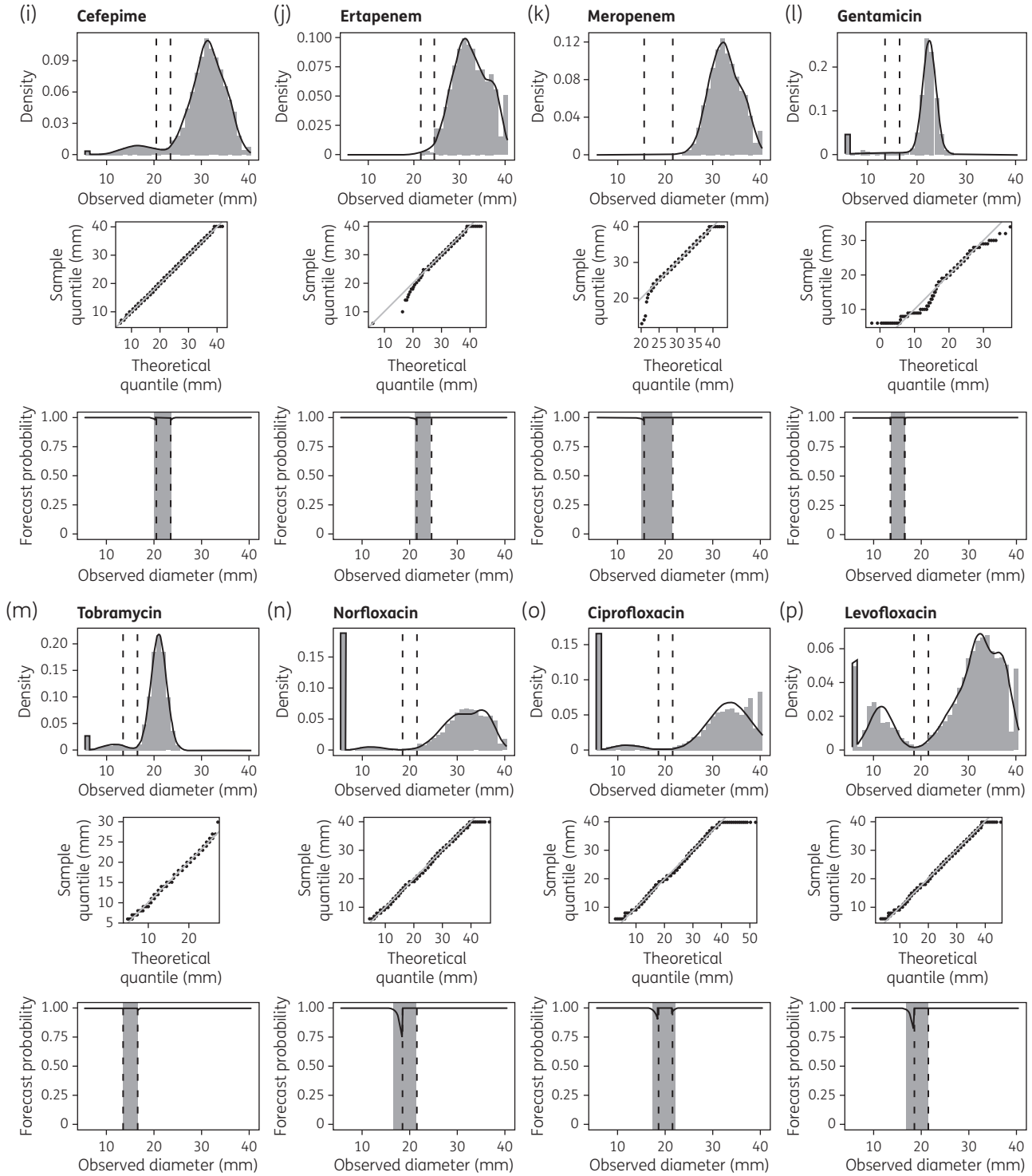

Figure 1. *Continued*

Discussion

In a first step we developed a mathematical model, called MASTER, that estimates the probability that a strain is truly susceptible or resistant given an observed inhibition zone diameter. MASTER is based on two assumptions. First, the model assumes that the observed diameter is the sum of a true, unknown diameter and normally distributed methodological variation. A similar assumption has been made for inhibition zone diameters[22–24] and for log-transformed MIC measurements.[25,26] The assumption that the magnitude of the methodological variation is constant for diameters >6 mm is a common approximation, justified in our

Table 2. Impact of ZMUs and increased CBPs defining susceptibility on rates of methodological misclassification errors

	VMEs (%)			MEs (%)			mEs (%)			
Antibiotic[a]	EUCAST CBPs	ZMU	CBPs + 2 mm	EUCAST CBPs	ZMU	CBPs + 2 mm	EUCAST CBPs	ZMU	CBPs + 2 mm	Suggested action
Ampicillin	2E−1	9E−5	2E−2	4E−1	7E−5	6E−2	0	2E+0	2E+0	introduce ZMU[b]
Cefoxitin	4E−1	1E−6	4E−2	7E−1	7E−6	1E−1	0	3E+0	4E+0	increase CBPs[c]
Amoxicillin/clavulanic acid	2E+0	1E−6	3E−1	4E+0	6E−6	7E−1	0	2E+1	2E+1	do not test[d]
Piperacillin/tazobactam	3E−3	2E−4	1E−5	1E−2	9E−4	1E−4	3E+0	3E+0	8E+0	–
Cefuroxime	3E−1	6E−7	3E−2	5E−1	5E−6	1E−1	0	3E+0	3E+0	introduce ZMU[b]
Cefotaxime	6E−3	8E−4	8E−5	2E−3	2E−4	2E−5	7E−1	7E−1	8E-1	–
Ceftazidime	3E−3	5E−4	3E−5	1E−2	1E−3	3E−4	2E+0	2E+0	6E+0	–
Ceftriaxone	2E−3	9E−4	6E−5	3E−3	5E−4	9E−4	2E−1	3E−1	1E+0	–
Cefepime	8E−3	1E−3	1E−4	8E−3	1E−3	4E−4	1E+0	2E+0	3E+0	–
Tobramycin	9E−3	9E−3	7E−5	1E−3	1E−3	2E−4	2E+0	2E+0	7E+0	–
Norfloxacin	1E−3	3E−4	7E−5	2E−2	9E−4	3E−3	6E−1	8E−1	2E+0	–
Ciprofloxacin	5E−3	6E−4	3E−4	1E−2	1E−3	2E−3	4E−1	7E−1	1E+0	–
Levofloxacin	2E−3	5E−4	1E−4	3E−2	1E−3	4E−3	9E−1	1E+0	2E+0	–

[a]Agreement between model and empirical distribution was poor for ertapenem, meropenem and gentamicin (Q–Q plots in Figure 1j, k and l). Calculating error rates for these antibiotics was thus not possible.
[b]Introducing a ZMU lowered ME and VME rates more than increasing the CBP defining susceptibility. The ZMU contained <5% of all isolates (Table 1), and mE rates were <5%.
[c]Cefoxitin is used as a screening drug for AmpC β-lactamases. While a ZMU would have decreased ME and VME rates by more than five orders of magnitude, it is unclear what benefit a ZMU would have for a screening drug.
[d]Increasing the CBP defining susceptibility resulted in ME and VME rates >0.1% and in an mE rate of 20%. The ZMU reduced ME and VME by several orders of magnitude. However, it contained 41% of all isolates (Table 1) and entailed an mE rate of 20%.

study by data from repeated measurements for *E. coli* ATCC 25922 and 10 clinical strains (Figure S3 and Table S1). Second, we modelled diameter distributions as mixtures of normal distributions, following an established approach. Inhibition zone diameters of WT isolates are usually modelled as normally distributed, e.g. in the context of setting epidemiological cut-off values.[1,27,28] In addition, several models based on mixtures of normally distributed components have been suggested to describe the full distribution of log-transformed MIC measurements.[22,26,29,30]

The estimated forecast probability comprises technical variation, i.e. how distant a specific diameter is from the CBP, epidemiological factors and unspecific biological factors such as genetic variations or metabolic state. MASTER can be used to derive forecast probabilities complementing AST categorizations with a quantitative measure of uncertainty to guide clinical decisions. To test the applicability of MASTER for CBP optimization we used a large set of inhibition zone diameters of 13 antibiotic drugs ($n = 120698$) originating from up to 9766 non-duplicate, non-outbreak clinical *E. coli* strains that had been isolated in our clinical laboratory from 2010 to 2014.

The Q–Q plots (Figure 1) illustrate that the discrepancies between the fitted models and the data were most pronounced for large inhibition zone diameters and in some cases for diameters of 6 mm. These discrepancies are negligible as they pertained to inhibition zone diameters that were much larger or much smaller than the CBPs. The discrepancies were most likely the inherent consequence of an experimental setting that places lower and upper bounds of 6 and 40 mm, respectively, on the reported diameters. In addition, the Q–Q plots showed that a sufficient amount of data from resistant strains is required for robust modelling by MASTER. For our particular epidemiological situation, this requirement was not met for ertapenem and meropenem (Figure 1j and k).

The calculated forecast probabilities were used to define ZMUs. The ZMUs encompassed <5% of all strains for most of the antibiotics investigated in this study. However, the ZMUs of amoxicillin/clavulanic acid and cefoxitin contained 41% and 6% of all strains, respectively (Table 1 and Figure 2). This finding is particularly important given that methodological variation is no longer covered by the proposed modification in the definition of the intermediate zone by EUCAST.[31]

The error rates predicted by MASTER can lead to several practical consequences: (i) CBPs that have been derived from aggregated datasets can be checked for error rates in a particular epidemiological setting; (ii) the model can provide a rationale to decide whether actions should be taken on CBPs to ensure optimal forecast probabilities for therapeutic success, and/or general CBPs may even be individually adjusted for specific epidemiological settings; and (iii) MASTER can be used to calculate the effect of CBP changes and/or the implementation of ZMUs on ME/VME rates and provides a rationale to decide whether action should be taken to ensure optimal forecast probabilities for therapeutic success, and if so what action should be taken.

To optimize CBPs for acceptable ME/vME rates, we applied a set of three rules as triggers for action:

1. ME and VME rates of <1% are mandatory; ME and VME rates of <0.1% are preferable.
2. (a) Increasing the susceptible CBP or implementation of a ZMU is recommended if these actions decrease ME and/or VME

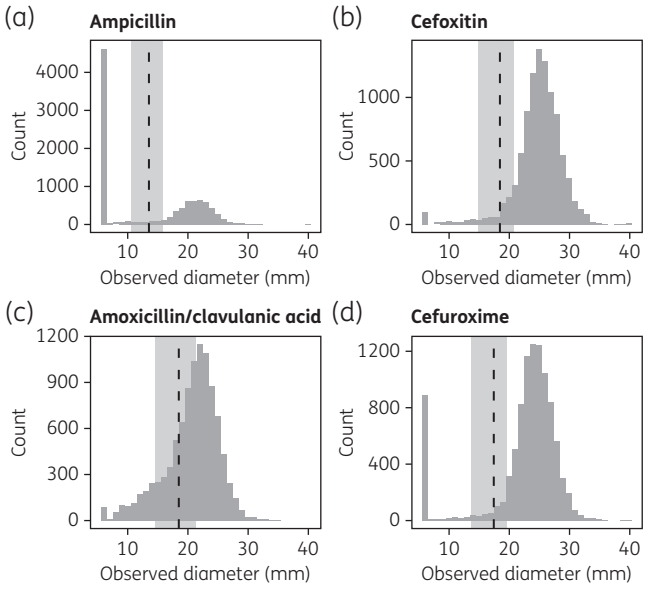

Figure 2. Empirical distribution of observed diameters (histogram), ZMUs (grey) and CBPs by EUCAST[11] (dashed lines) for drugs with suggested actions on CBPs.

rates by at least one order of magnitude. (b) ZMUs must not contain >10% of the isolates.

3. If neither increasing the susceptible CBP nor implementation of the ZMU results in ME and VME rates of at most <1% or if the ZMU contained >10% of the isolates, testing of the drug is of limited value for a given epidemiology.

For our epidemiological setting, ampicillin (10 μg disc), cefoxitin, amoxicillin/clavulanic acid and cefuroxime had ME and VME rates >0.1%, indicating the need for action according to rule 1 (Table 2). For these drugs, implementation of ZMUs decreased error rates to a greater extent than increasing susceptible CBPs (Table 2) and ZMUs would thus be recommended, applying rule 2(a). Cefoxitin, however, is used as a screening parameter for the presence of AmpC β-lactamases to ensure maximum sensitivity, but is not reported to clinicians for therapeutic purposes.[32,33] Although a ZMU for cefoxitin would decrease ME and VME rates by more than five orders of magnitude, it would provide no benefit for a screening drug that is used for diagnostic purposes only. For amoxicillin/clavulanic acid, the ZMU contained 41% of all isolates (Table 1) and would thus not be practically useful (rule 2b). Furthermore, an amoxicillin/clavulanic acid ZMU would lead to mE rates of 20% (Table 2). Since ME and VME <0.1% could also not be achieved by an increased susceptible CBP for amoxicillin/clavulanic acid, reporting of the drug may be of limited clinical value, at least in an epidemiological setting like that of this study, in which many isolates were situated close to the CBP (rule 3). Since prediction of therapeutic success is frequently uncertain, amoxicillin/clavulanic acid may no longer be recommended for clinical use in *E. coli*. Ampicillin and cefuroxime are reported for treatment purposes, and <5% of the isolates were contained in their respective ZMUs (Table 1). Therefore, the implementation of ZMUs for ampicillin and cefuroxime can be recommended.

By design, this study is limited to the analysis of the epidemiology of a single geographical location. However, MASTER can readily be applied to other regional epidemiologies to evaluate local adjustments of CBPs, or it could be applied to aggregated datasets, e.g. those used by EUCAST and CLSI to set CBPs. If aggregated data are analysed, EUCAST or CLSI QC ranges can be used as a surrogate for methodological variation since they are based on aggregated data using disc/agar plates from different manufacturers.[21] Bootstrapping can be used to quantify the uncertainty of ZMUs and rates of methodological errors calculated with MASTER.[34] Furthermore, interpretative categories may not be adjusted if the breakpoint is placed in a region of low probability of target attainment as determined by Monte Carlo simulations.

In conclusion, this study demonstrates how CBPs can be optimized in a standardized process aiming to produce lower categorization error rates. If the expected ME and VME rates using a single CBP are not satisfactory, ZMUs may be introduced in case an intermediate zone is not appropriate (e.g. for pharmacokinetic/pharmacodynamic reasons) and ME and VME rates cannot be decreased by a change in the CBP.

Acknowledgements

We thank the technicians of the Institute of Medical Microbiology for expert help and assistance, and we are grateful to the Institute of Medical Microbiology and the University of Zurich for continuous support. We thank Giorgia Valsesia for data preprocessing and Patrice Courvalin and Florian Maurer for valuable discussions on forecast probabilities.

Funding

This work was supported by the University of Zurich.

Transparency declarations

None to declare.

Supplementary data

Figures S1 to S4 and Table S1 are available as Supplementary data at *JAC* Online.

References

1 Kronvall G, Giske CG, Kahlmeter G. Setting interpretive breakpoints for antimicrobial susceptibility testing using disk diffusion. *Int J Antimicrob Agents* 2011; **38**: 281–90.

2 Turnidge J, Paterson DL. Setting and revising antibacterial susceptibility breakpoints. *Clin Microbiol Rev* 2007; **20**: 391–408.

3 Mouton JW, Brown DF, Apfalter P *et al*. The role of pharmacokinetics/pharmacodynamics in setting clinical MIC breakpoints: the EUCAST approach. *Clin Microbiol Infect Dis* 2012; **18**: E37–45.

4 Genzen JR, Tormey CA; Education Committee of the Academy of Clinical Laboratory Physicians and Scientists. Pathology consultation on reporting of critical values. *Am J Clin Pathol* 2011; **135**: 505–13.

5 EUCAST. *Consultation on a Proposed Change in the EUCAST Definition of the Intermediate Category*. 2015. http://www.eucast.org/eucast_news/

news_singleview/?tx_ttnews%5Btt_news%5D=151&cHash=24813f2b9f862507db854ab7cafd2059.

6 Hombach M, Bottger EC, Roos M. The critical influence of the intermediate category on interpretation errors in revised EUCAST and CLSI antimicrobial susceptibility testing guidelines. *Clin Microbiol Infect Dis* 2013; **19**: E59–71.

7 Hombach M, Ochoa C, Maurer FP *et al*. Relative contribution of biological variation and technical variables to zone diameter variations of disc diffusion susceptibility testing. *J Antimicrob Chemother* 2016; **71**: 141–51.

8 Maurer FP, Courvalin P, Bottger EC *et al*. Integrating forecast probabilities in antibiograms: a way to guide antimicrobial prescriptions more reliably? *J Clin Microbiol* 2014; **52**: 3674–84.

9 Dalhoff A, Ambrose PG, Mouton JW. A long journey from minimum inhibitory concentration testing to clinically predictive breakpoints: deterministic and probabilistic approaches in deriving breakpoints. *Infection* 2009; **37**: 296–305.

10 US Department of Health and Human Services FDA. *Class II Special Controls Guidance Document: Antimicrobial Susceptibility Test (AST) Systems*. 2009. http://www.fda.gov/RegulatoryInformation/Guidances/ucm080564.htm.

11 EUCAST. *Breakpoint Tables for Interpretation of MICs and Zone Diameters*. Version 6.0. 2016. http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_6.0_Breakpoint_table.pdf.

12 EUCAST. *Antimicrobial Susceptibility Testing. EUCAST Disk Diffusion Method*. Version 4.0. 2014. http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Disk_test_documents/Manual_v_4.0_EUCAST_Disk_Test.pdf.

13 Hombach M, Zbinden R, Böttger EC. Standardisation of disk diffusion results for antibiotic susceptibility testing using the sirscan automated zone reader. *BMC Microbiol* 2013; **13**: 225.

14 Hombach M, Courvalin P, Böttger EC. Validation of antibiotic susceptibility testing guidelines in a routine clinical microbiology laboratory exemplifies general key challenges in setting clinical breakpoints. *Antimicrob Agents Chemother* 2014; **58**: 3921–6.

15 Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Series B Methodol* 1977; **39**: 1–38.

16 Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978; **6**: 461–4.

17 Fraley C, Raftery AE. Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc* 2002; **97**: 611–31.

18 Wilk MB, Gnanadesikan R. Probability plotting methods for the analysis of data. *Biometrika* 1968; **55**: 1–17.

19 R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2015.

20 Fraley C, Raftery AE, Murphy TB *et al*. mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation. *Technical Report No. 597*. Department of Statistics, University of Washington, Seattle, 2012.

21 EUCAST. *Routine and Extended Internal Quality Control for MIC Determination and Disk Diffusion as Recommended by EUCAST*. Version 6.1. 2016. http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/QC/v_6.1_EUCAST_QC_tables_routine_and_extended_QC.pdf.

22 Craig BA. Modeling approach to diameter breakpoint determination. *Diagn Microbiol Infect Dis* 2000; **36**: 193–202.

23 Lamy B, Carret G, Flandrois J-P *et al*. How does susceptibility prevalence impact on the performance of disk diffusion susceptibility testing? *Diagn Microbiol Infect Dis* 2004; **49**: 131–9.

24 Annis DH, Craig BA. Statistical properties and inference of the antimicrobial MIC test. *Stat Med* 2005; **24**: 3631–44.

25 van de Kassteele J, van Santen-Verheuvel MG, Koedijk FDH *et al*. New statistical technique for analyzing MIC-based susceptibility data. *Antimicrob Agents Chemother* 2012; **56**: 1557–63.

26 Annis DH, Craig BA. The effect of interlaboratory variability on antimicrobial susceptibility determination. *Diagn Microbiol Infect Dis* 2005; **53**: 61–4.

27 Kronvall G, Kahlmeter G, Myhre E *et al*. A new method for normalized interpretation of antimicrobial resistance from disk test results for comparative purposes. *Clin Microb Infect* 2003; **9**: 120–32.

28 Turnidge J, Kahlmeter G, Kronvall G. Statistical characterisation of bacterial wild-type MIC value distributions and the determination of epidemiological cut-off values. *Clin Microb Infect* 2006; **12**: 418–25.

29 Meletiadis J, Mavridou E, Melchers WJG *et al*. Epidemiological cutoff values for azoles and *Aspergillus fumigatus* based on a novel mathematical approach incorporating cyp51A sequence analysis. *Antimicrob Agents Chemother* 2012; **56**: 2524–9.

30 Cantón E, Pemán J, Hervás D *et al*. Comparison of three statistical methods for establishing tentative wild-type population and epidemiological cutoff values for echinocandins, amphotericin B, flucytosine, and six *Candida* species as determined by the colorimetric Sensititre YeastOne method. *J Clin Microbiol* 2012; **50**: 3921–6.

31 EUCAST. *Proposed Modifications to EUCAST Definition of the Intermediate Category*. 2015. http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Consultation/2016/20160919_Intermediate_consultation_comments_and_responses.pdf.

32 EUCAST. *EUCAST Guidelines for Detection of Resistance Mechanisms and Specific Resistances of Clinical and/or Epidemiological Importance*. http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Resistance_mechanisms/EUCAST_detection_of_resistance_mechanisms_v1.0_20131211.pdf.

33 Polsfuss S, Bloemberg GV, Giger J *et al*. A practical approach for reliable detection of AmpC β-lactamase producing Enterobacteriaceae. *J Clin Microbiol* 2011; **49**: 2798–803.

34 Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Boca Raton, FL, USA: Chapman & Hall, 1994.