MS 1: HOMER – Description of a new mathematical tool to calculate categorization error rates of antimicrobial susceptibility testing from population distributions

**MS 2: The use of HOMER, a new mathematical tool to calculate antibiotic susceptibility forecast probabilities in optimizing clinical breakpoints**

**Nicolas Blöchliger, Peter M. Keller, Michael Hombach**

*Institut für Medizinische Mikrobiologie, Universtät Zürich, 8006 Zürich, Switzerland*

Short title: Forecast probability-based optimization of clinical breakpoints

Keywords: antimicrobial susceptibility testing, clinical breakpoints, forecast probability, susceptibility, mathematical model, error rates

‡Corresponding author:

Michael Hombach, M.D.

Institut für Medizinische Mikrobiologie

Universität Zürich

Gloriastr. 30/32

8006 Zürich

Switzerland

Phone: 0041 44 634 27 00

Fax: 0041 634 49 06

email: mhombach@imm.uzh.ch

**Abstract**

**Introduction**

Antibiotic susceptibility testing (AST) results are generally used by clinicians as a guide to select the most promising drugs in the treatment of infectious diseases for individual patients. The laboratory, however, does not provide raw results of AST such as minimum inhibitory concetrations or disk zone diameters. Instead, results are already categorized using clinical breakpoints (CBPs) according to the expected clinical success, which is helpful in daily clinical practice but also means a loss of information between the laboratory and the clinician. CBP setting has traditionally been a hardly standardized procedure: Susceptibility population data of pathogens, pharmacodynamic/pharmacokinetic values, and clinical outcome studies were taken into account, but a set of rules or target parameters that would make CBP categorizations comparable, e.g. for their level of reliability, has not been established. In contrast, in clinical chemistry raw results are transmitted to the clinicians together with reference ranges for the normal values for the patient’s population.

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 has been used in an ambiguous way: EUCAST recently re-defined the intermediate category removing the concept of a technical buffer zone from the definition (EUCAST Document!). Originally, the EUCAST definition contained both traditional 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.” The use of a mixed definition for the intermediate category resulted in confusion regarding its meaning and hampered understanding of the primary addressees of AST reports, i.e. the clinicians selecting the most appropriate drug in antibiotic treatment. To clarify the meaning of the intermediate category, it is now 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 significant methodological variation of AST remains and is no longer covered by intermediate categorization nor is it indicated on AST reports. As a consequence, erroneous forecasts of therapeutic success could occur (Hombach/Roos, Maurer/Hombach, Ochoa/Furrer papers).

Tools that optimize CBPs by reducing error rates seem beneficial to ensure an adequate reliability of antibiograms and could provide a basis for standardizing the process of CBP setting. If acceptable categorization error rates cannot be achieved by CBP changes, zones of methodological uncertainty (ZMUs) may be helpful to indicate a less reliable range of measurements while avoiding the term “intermediate” and the associated confusion. Results in ZMUs may either not be reported at all or be reported as resistant depending on the regional epidemiology and antibiotic policy in the given medical setting (Valsesia??).

In a companion article we have described a new method that allows comparing CBPs and CBP changes with respect to associated forecast probabilities for clinical categories (CIT). The new tool, which we named “HOMER” (Helper for Optimizing Major Error Rates) takes into account technical variation, biological, and epidemiological factors and is therefore suitable to study CBP changes and the introduction of ZMUs in individual epidemiological setting as well as in aggregated datasets.. In this proof-of-principle study use HOMER to systematically evaluate the current CBPs issued by EUCAST. We employ a broad set of inhibition zone diameters of non-duplicate, non-outbreak clinical *E. coli* strains originating from our clinical laboratory, i.e. representing our local epidemiology.

**Materials and Methods**

**Clinical isolates.** Antimicrobial susceptibility data of xxxx non-duplicate, non-outbreak clinical *E. coli* strains were used included in this study. All clinical *E. coli* isolates included in this study were isolated over a five-year period from 2010 until 2014 in the clinical microbiology laboratory of the Institute of Medical Microbiology, University of Zurich. Isolates of the same species were considered duplicates and discarded if they i) originated from the same patient, and ii) showed at most one major and two minor differences in AST interpretation. Isolates with inhibition zone sizes in the left tail area of the distribution curves were verified by visual inspection of the automated antibiogram pictures. Isolates were eliminated from the data set if errors were found (e.g. no plates, contaminated plates, wrong identification, inhibition zones wrongly measured, or incompatibility with wild-type definition due to resistance to other beta-lactams).

**Antibiotic susceptibility testing.** Antimicrobial susceptibility testing was performed by the disk diffusion method following EUCAST standard procedures (13-15) on Müller-Hinton II agar (Beckton-Dickinson, Franklin Lakes, NJ, USA) and with antibiotic discs from i2a (Montpellier, France). Inhibition zone diameters were recorded automatically using the Sirscan/Sirweb system (i2a, Montpellier, France) (REF Zbinden/Hombach).

**Model for true and observed inhibition zone diameters.**

We developed a model to determine the methodological error for categorisation of Kirby-Baur inhibitory zone diameters. Our model distinguishes between an observed inhibition zone diameter *Y* that suffers from methodological variation and an underlying true diameter *X* that is not observed. The relationship between *X* and *Y* is given by *Y* = *X* + *E*, where *E* models the methodological variation.We use a normal mixture model to describe the distribution of the true diameter *X* and we assume *E* to be normally distributed and independent of *X* with zero mean. We estimate the standard deviation of E from independent repeated measurments of inhibition zone diameters for the quality control strain ATCC 25922. Observed diameters of 6 mm are treated differently, as described in detail elsewhere (Cite MS1).

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

**Forecast probabilities and zones of technical uncertainty.**

Given an observed diameter *y*, what is the probability that no major or very major error occurred due to methodological variation? We term this probability, which is a function of *y*, forecast probability (*pf* (*y*)) and it can be readily calculated based on our model. We can identify inhibition zone diameters for which the AST categorisation is ambiguous due to methodological

variation based on the forecast probabilities *pf* (*y*). In particular, we define the ZMU to encompass all the observed diameters for which the risk of a major or very major error is higher than 1% (i.e. *pf* (*y*) *<* 0*.*99) or which are in the intermediate zone.

**Probabilities of methodological misclassification errors.** Differences in clinical categorization are referred to as “errors”, which are traditionally are split according to their therapeutic implications: Erroneous categorization of true-susceptible isolates as resistant are referred to as “major errors” (ME) leading to unnecessary restriction of therapeutic options, whereas the most serious clinical implications do result from “very major errors” (vME), i.e. categorization of true-resistant isolates as susceptible, as there is a high likelihood of therapeutic failure.

Our model allows the calculation of the probabilities of major and very major misclassification errors only as far as they are due to methodological variation. In the present work, the probability for a very major methodological error is defined as the probability that the observed diameter *Y* is above the CBP defining susceptibility while the true diameter *X* is below the CBP defining resistance. Similarly, the probability of a major methodological error is the probability that the strain is classified as resistant based on the observed diameter while it is susceptible according to the true but unknown diameter *X*.

**Results**

For 14 out of the 18 drugs analyzed, our model predicted ME and vME probabilities in the desirable range, i.e. <0.1% (<1E-03), if current EUCAST CBPs were applied (Table 1). For these 14 drugs vME probabilities ranged from 3E-14 to 3E-05, and ME probabilities ranged from 1E-10 to 1E-04. The CBPs for all of the 14 drugs that displayed error probabilities in the desirable range define intermediate zones of 3 mm to 6 mm width. The lowest ME/vME probabilities were calculated for imipenem and meropenem, for which CBPs contain the widest intermediate zones.

Three drugs displayed ME/vME probabilities higher than the desirable range, i.e. >0.1% (>1E-03), if current EUCAST CBPs were applied, i.e. ampicillin, cefoxitin, and cefuroxime, whereas amoxicillin-clavulanic acid was the only drug for which HOMER predicted ME/vME probabilities that were not acceptable, i.e. >1% (Table 1).

Calculated ZMUs ranged from 3 mm width for tobramycin to 8 mm width for meropenem (Table 1). The majority of calculated ZMUs resembled the intermediate zones if present in EUCAST CBPs. The relative number of isolates that were situated in calculated ZMUs ranged from one-digit percentages (15 out of 18 drugs) over 10% for ceftazidime, 15% for piperacillin-tazobactam up to 41% for amoxicillin-clavulanic acid. While for piperacillin-tazobactam the comparably high number of isolates in the ZMU did not result in particularly high ME/vME rates due to the presence of an intermediate zone, the significant part of the total population for amoxicillin-clavulanic acid that was situated in the ZMU lead to unacceptably high ME/vME rates as no intermediate zone is defined by current CBPs.

**Discussion**

In a first step we developed HOMER, a new tool for calculating SIR-classification error rates described in the companion of this article that covers both technical variation as well as biological and epidemiological factors (CIT). HOMER estimates the probability that a strain is truly susceptible given an observed diameter that suffers from technical uncertainty.

To test for the principal applicability and usefulness of HOMER for CBP setting in this study, we used a large set of inhibition zone diameters (N=xxxxxx data points) originating from xxxx non-duplicate, non-outbreak clinical *E. coli* strains and 18 antibiotic drugs that had been isolated in our clinical laboratory from 2010 to 2014.

We here used our local epidemiology as a paradigmatic example to test for the principle usefulness of the mathematical model. However, HOMER is not restricted to a particular epidemiology and can thus be used to calculate expected error rates of CBPs in different epidemiological settings. The predicted error rates that are generated by the model can lead to several practical consequences: i) General CBPs that have been derived from aggregated datasets can be checked for error rates in all epidemiological settings covered by guideline issuing societies; ii) Official, uniform CBPs may be individually adjusted for specific epidemiological settings and/or laboratory environments to account for varying local susceptibility prevalence and associated differences in AST data distributions; iii) HOMER can be used to calculate the effect of CBP changes and/or the implementation of zones of methodological uncertainty (ZMUs) on ME/VME rates and provide a rationale to decide, whether and, if necessary, which actions should be taken on CBPs to ensure optimal forecast probabilities for therapeutic success.

ZMUs will be of particular importance as technical measurement variation and the resulting error probabilities are no longer contained in the EUCAST definition of the intermediate zone that indicates the need of high dose therapy or on-site concentration of the drug only for sake of an unambiguous clinical recommendation. To standardize CBP setting according to acceptable ME/VME rates, we used a set of rules as triggers for action on CBPs:

1. ME and vME rates of 1.0e-02 (<1%) should be mandatory and ME and vME rates of 1.0e-03 (<0.1%) are desirable. To avoid unnecessary changes in guidelines ME and vME rates of ≤1.0e-04 (<0.01%) are accepted.
2. Increasing S CBP and/or ZMU are recommended if any of these actions decrease ME and/or vME probability at least 1 order of magnitude.
3. Decreasing vME outweighs increasing ME if the expected overall rate does not exceed 1%.
4. ZMUs are recommended if not more than 10% of the isolates are situated in the proposed ZMU range.
5. If more than 10% of the isolates are situated in the proposed ZMU range AND vME probability of official CBPs is >1% AND increasing the S CBP does not result in decreasing the vME rate to <1%, testing of the drug should be discouraged as no results can be obtained on acceptable probability level.

For our local epidemiology this would have triggered actions for *E. coli* isolates on four drugs with vME/ME rates higher than what we considered acceptable (<1%) and/or desirable (<0.1%), i.e. ampicillin (10 µg disc), cefoxitin, amoxicillin-clavulanic acid, and cefuroxime. Applying the abovementioned rules, the following modifications from current EUCAST CBPs seemed reasonable: Ampicillin (10 µg disc): Introduction of a ZMU of 11 to 15 mm is recommended as this ZMU leads to desirable ME and vME rates lower than that for increased susceptible CBPs and only 3% of isolates are situated in the ZMU; cefoxitin: this drug and the associated breakpoint serve as the AmpC screening parameter to ensure maximal sensitivity, i.e. low vME rates are most desirable. ME rates reflecting specificity are less important as results of the AmpC screening must be confirmed anyway by independent testing (EUCAST ZIT, Polsfuss et al). Introduction of a ZMU would be technically optimal to lower error rates, but it is unclear what value a ZMU would have for a screening drug that is usually not included on clinical reports; amoxicillin-clavulanic acid: Increasing the susceptible CBP did not lead to desirable error rates, while applying a ZMU resulted in acceptable error rates, but 41% of isolates would be situated in the necessary ZMU, which seems of no value. As desirable forecast probabilities cannot be achieved by either an increased CBP nor a ZMU, testing of the drug is discouraged; cefuroxime: Introduction of a ZMU of 14 to 19 mm is recommended as this ZMU leads to desirable ME and vME rates lower than that for increased susceptible CBPs and only 4% of isolates are situated in the ZMU.

A limitation of this study is the analysis of one regional epidemiology. However, this study was intended to serve as a proof of principle on the example of a single clinical AST dataset. The model HOMER, however, can easily be applied on other local epidemiologies to check for reasonable individual actions on CBPs, or it can be used on aggregated datasets such as those used by EUCAST and CLSI to set their CBPs.

To apply our model to data collected by different laboratories it is mandatory to apply their individual technical variation as reflected e.g. by repetitive measurements of QC strains. If aggregated data are used, the EUCAST or CLSI QC ranges could be used as a surrogate for the magnitude technical variation as QC ranges usually reflect the 2-fold standard deviation next to a given target, i.e. a mean value. Moreover, if an individual laboratory changes the manufacturer of antibiotic discs or agar plates the methodological variation will also change (EUCAST Dok ZIT) and the model needs to be adjusted.

In conclusion, this study demonstrates that CBPs can be optimized in a standardized process with the view to keep methodological categorization error rates as low as possible. Something to write on ZMU….

**References**

[1]F.P. Maurer, P. Courvalin, E.C. Böttger, M. Hombach, Integrating forecast probabilities in antibiograms:

A way to guide antimicrobial prescriptions more reliably?, J. Clin. Microbiol. 52 (2014) 3674–3684.

[2]A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm,

J. Roy. Stat. Soc. B. (1977) 1–38.

[3]R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical

Computing, Vienna, Austria, 2015. https://www.R-project.org/.

[4]C. Fraley, A.E. Raftery, T.B. Murphy, L. Scrucca, 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. (2012).

[5]B.A. Craig, Modeling approach to diameter breakpoint determination, Diagn. Microbiol. Infect. Dis. 36

(2000) 193–202.

[6]B. Lamy, G. Carret, J.-P. Flandrois, M.L. Delignette-Muller, How does susceptibility prevalence impact on

the performance of disk diffusion susceptibility testing?, Diagn. Microbiol. Infect. Dis. 49 (2004) 131–139.

[7]D.H. Annis, B.A. Craig, Statistical properties and inference of the antimicrobial MIC test, Statist. Med.

24 (2005) 3631–3644.

[8]J. van de Kassteele, M.G. van Santen-Verheuvel, F.D.H. Koedijk, A.P. van Dam, M.A.B. van der Sande,

A.J. de Neeling, New statistical technique for analyzing MIC-based susceptibility data, Antimicrob. Agents

Chemother. 56 (2012) 1557–1563.

[9]D.H. Annis, B.A. Craig, The effect of interlaboratory variability on antimicrobial susceptibility determination,

Diagn. Microbiol. Infect. Dis. 53 (2005) 61–64.

[10]G. Kronvall, G. Kahlmeter, E. Myhre, M.F. Galas, A new method for normalized interpretation of

antimicrobial resistance from disk test results for comparative purposes, Clin. Microbiol. Infect. 9 (2003)

120–132.

[11]J. Turnidge, G. Kahlmeter, G. Kronvall, Statistical characterisation of bacterial wild-type MIC value

distributions and the determination of epidemiological cut-off values, Clin. Microbiol. Infect. 12 (2006)

418–425.

[12]J. Meletiadis, E. Mavridou, W.J.G. Melchers, J.W. Mouton, P.E. Verweij, Epidemiological cutoff values

for azoles and Aspergillus fumigatus based on a novel mathematical approach incorporating cyp51A sequence

analysis, Antimicrob. Agents Chemother. 56 (2012) 2524–2529.

[13]E. Cantón, J. Pemán, D. Hervás, C. Iñiguez, D. Navarro, J. Echeverría, J. Martínez-Alarcón, D. Fontanals,

B. Gomila-Sard, B. Buendía, L. Torroba, J. Ayats, A. Bratos, F. Sánchez-Reus, Fernández-Natal, the

FUNGEMYCA Study Group, 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. 50 (2012) 3921–3926.

[14]G. DePalma, J. Turnidge, B.A. Craig, Determination of disk diffusion susceptibility testing interpretive

criteria using model-based analysis: Development and implementation, Diagn. Microbiol. Infect. Dis. in press

(2016) ??–??

[15]G. DePalma, J. Turnidge, B.A. Craig, dBETS: An online software package for diffusion test breakpoint

determination, User Manual. in press (2016) ??–??

4