

DRUG DILUTION VS DRUG DIFFUSION:
CALIBRATING THE TWO SUSCEPTIBILITY TESTS

by
and

Bruce A. Craig
Purdue University

Technical Report #99-15

Department of Statistics
Purdue University
West Lafayette, IN USA

DRUG DILUTION VS DRUG DIFFUSION: CALIBRATING THE TWO SUSCEPTIBILITY TESTS

by

Bruce A. Craig
Purdue University

Abstract

To determine the susceptibility of an unknown pathogen to a specific drug, hospital laboratories perform either a drug dilution or disk diffusion test. Test/Drug-specific breakpoints then classify the pathogen as either susceptible, intermediate or resistant to the drug. Since only one of these tests will be used in practice, it is imperative to have the two tests give similar results. Currently, pharmacokinetics and pharmacodynamics of the drug are used to set drug dilution breakpoints and comparable disk diffusion breakpoints are determined by generating pairs of test results for a wide range of pathogens and finding breakpoints which minimize the classification discrepancies. While this procedure is very fast and simple to implement, it does not adequately take into account the inherent variability of each test nor the underlying distribution of pathogens. As a result, the choice of breakpoints is very sample dependent. In this paper, a hierarchical errors-in-variables model is developed to explicitly describe the various factors of uncertainty in these pairs and probabilities from this model are used to determine breakpoints. Bayesian inference is used to combine the model with a scatterplot of test results. While this procedure is more time consuming to implement (particularly the inference component), it is shown that by accounting for the various uncertainties in the results, much of the sample dependency is eliminated giving more consistent and interpretable results.

1. Introduction

Susceptibility testing is used to aid a physician in prescribing a drug to eliminate an unknown pathogen in the body. To determine which drugs are appropriate, hospital laboratories perform one of two tests. Drug dilution involves the addition of two-fold dilutions of the drug to separate vials of broth containing a standardized number of the pathogen (just below visible detection) and after incubation, examining each vial for visible growth (vial becomes cloudy). The minimum inhibitory concentration (MIC) is defined to be the lowest of these two-fold dilutions that prevents visible growth. Disk diffusion involves the placement of a disk with a standardized concentration of the drug on an agar plate covered with the pathogen. The drug diffuses from the disk creating a gradient of concentrations. After incubation, a clear zone around the disk will result from drug concentrations high enough to kill the pathogen in that area. The smaller the MIC, or larger the diameter of the clear zone (DIA), the more likely this pathogen can be successfully treated with the drug.

This likelihood is classified for the physician into one of three categories. Susceptible (S) means that there is a very high likelihood of successfully treating the pathogen by the usual FDA approved dosage. Intermediate (I) means that the pathogen can likely be treated, especially if doses are higher than usual, and resistant (R) means the likelihood is low using approved dosage levels. These drug-specific classification regions, described by a lower and upper breakpoint, are defined by subcommittees of the Food and Drug Administration (FDA) and the National Committee on Clinical Laboratory Standards (NCCLS). Since the MIC test involves concentrations of the drug, its breakpoints are largely based on the pharmacokinetics and pharmacodynamics of the drug. Determination of the DIA breakpoints, however, is not as straightforward because there is not a simple relationship between concentration levels and zone diameter. While there is often an assumed quadratic or linear relationship between the log of the MIC and the diameter, the parameters of this relationship are drug-specific.

The current DIA breakpoint determination procedure, known as the error-rate bounded method (Metzler and DeHaan, 1974; Brunden, Zurenko, and Kapik, 1992), involves investigating a wide range of pathogens using both tests and then finding DIA breakpoints which satisfy certain discrepancy restrictions. In other words, the MIC breakpoints classify each pathogen as either susceptible, intermediate or resistant and the DIA breakpoints are adjusted until the observed percentage of discrepancies between the DIA and MIC classifications satisfy certain restrictions. Specifically, less than 1.5% of the pairs result in a very

major discrepancy (MIC test = R and DIA test = S) and less than 3% of the pairs result in a major discrepancy (MIC test = S and DIA test = R) (NCCLS, 1995). There are currently no restrictions on minor discrepancies (one of the tests = I) but often less than 10% is used.

Because of testing variabilities, rounding and the inverse relationship between MIC and DIA, a discrepancy is most likely to occur when the pathogen is near the MIC intermediate region. When there are very few pathogens in this region, numerous breakpoint sets satisfy the discrepancy restrictions and other decision mechanisms, such as diameter width restrictions, are needed. On the other hand, when there are numerous pathogens in this region, there may be no breakpoint sets that satisfy the discrepancy restrictions (minor included). While previously the distribution of pathogens was bimodal with very few pathogens near the intermediate zone, in recent years there has been a steady increase in marginally susceptible and resistant pathogens making this latter situation more of a problem and has resulted in a reevaluation of this procedure.

Rather than attempt to modify the error-rate bounded method, an alternative model-based approach to breakpoint determination is proposed. This approach eliminates the need for discrepancy restrictions and does not depend on the distribution of pathogens. A hierarchical model explicitly accounts for the various factors of uncertainty in a scatterplot thereby serving as a filter to remove as much explainable variation as possible. Since the model distinguishes between true and observed results, test-specific probabilities from this model are then used to determine DIA breakpoints. To obtain model estimates, Bayesian inference, enabled by Markov chain Monte Carlo, is used. Thus, instead of a single estimate, a distribution of potential breakpoint sets is obtained thereby allowing the experimenter to assess the remaining uncertainty in the scatterplot. The paper is organized as follows. In section 2, the hierarchical model is developed and Section 3 describes how certain probabilities are used to determine DIA breakpoints. In section 4, the Bayesian method of inference is described and in Section 5, the results of a simulation study to compare the two procedures and an application of the new procedure to three published scatterplots are presented. A discussion follows.

2. The Model

The model is constructed to describe a scatterplot of MIC/DIA pairs in terms of Normal distributions thereby allowing simple computation of discrepancy percentages and other conditional probabilities of interest. The model separates the scatterplot into three components:

1. The test procedures (i.e., rounding) and experimental variability

2. The drug-specific relationship between MIC and DIA
3. The underlying distribution of pathogens (or MICs).

The first component links the observed MIC/DIA pair with an underlying true value. The second and third components describe the relationship between these true DIAs and MICs. In the MIC test, two-fold dilutions ($\mu\text{g/ml}$) are performed with the MIC being the lowest two-fold dilution without visible growth of the pathogen. For the remainder of this paper, consider the MIC in terms of log base 2 units so that test results are integer values similar to the zone diameter (DIA) which is measured to the nearest millimeter.

2.1 *Experimental Error*

Repeat experiments, same drug and pathogen, have shown a common three-fold range in the observed MIC. This range is attributed to inherent testing variability and rounding. To describe the distribution of observed results, a Normal distribution is discretized by rounding all values up to the next highest integer. For pathogen i , the observed MIC

$$x_i = \lceil m_i + \epsilon_i \rceil, \quad (1)$$

where m_i is the true MIC and ϵ_i , distributed $N(0, \sigma_m)$, is the testing error. The rounding up is due to the fact that lack of growth is only observed in dilutions above the MIC. For example, if the observed MIC were truly -1.78, the lowest test dilution with no growth would be -1. Figure 1 displays the observed MIC distribution for three different choices of the true MIC. The various shapes of the distribution are consistent with what is observed in repeat drug/pathogen trials and the choice of $\sigma_m = 0.5$ is consistent with the common three to four dilution range. One can also see the implications of the variation and rounding in terms of the probability of correct identification. If the MIC breakpoints were -1 and 1, all three of these true MICs (means) are in the intermediate range. However, in terms of the observed results, the second and third pathogens are most likely going to be classified as resistant ($x \geq 1$). While a discrepancy depends on both test results, this rounding up suggests that a very major discrepancy (MIC test = R) is more likely than a major (MIC test = S). A result that does not agree with the current tolerable percentage restrictions.

In a similar fashion, the observed DIA

$$y_i = \lceil d_i + \delta_i \rceil \quad (2)$$

where d_i is the true DIA and δ_i is $N(0, \sigma_d)$. Instead of rounding up, the disk diffusion test rounds the diameter to the nearest millimeter. Because these tests are done separately, we assume that the ϵ_i and δ_i are independent. We also assume that the experimental standard deviations, σ_m and σ_d , remain constant over the range of MICs in the study.

2.2 True MIC/DIA Relationship

Since the model distinguishes between the observed and true results, the relationship between the MIC and DIA is described in terms of the true values. Specifically,

$$d_i = \beta_0 \left(\frac{\exp(\beta_1 - \beta_2 m_i)}{1 + \exp(\beta_1 - \beta_2 m_i)} \right) \quad (3)$$

where m_i and d_i are the true values for pathogen i and the β 's are drug-specific. This one-to-one relationship means that we only need to describe the underlying distribution of MICs to describe the joint distribution of test results. While both linear and quadratic relationships have been used to describe this relationship, both these functions allow negative diameters for large MIC. This function avoids the need to build into the model the restriction $d_i > 0 \forall i$.

2.3 Underlying Distribution of MICs

Just as the β parameters are drug specific, so is the underlying distribution of pathogens. To allow for multi-modality and skewness, this distribution is a mixture of Normals where the number, k , of Normal distributions in the mixture is an unknown. Specifically,

$$\pi(m) = \sum_{j=1}^k w_j N(m; \mu_j, \sigma_j) \quad (4)$$

where μ_j , σ_j , and w_j are the mean, standard deviation, and weight of the j th Normal ($\sum w = 1$).

Figure 2 demonstrate how these two components, combined with experimental variability and rounding, form a scatterplot. The left panel plots 300 true MIC/DIA pairs and the right panel introduces the rounding and experimental variabilities. The true MIC/DIA pairs were obtained by randomly sampling MICs from

$$\pi(m) = .55N(-3.0, 1.00) + .10N(0.5, 0.75) + .35N(5.0, 0.75),$$

which is drawn in the left panel, and determining the true DIA using

$$d_i = 62.8 \left(\frac{\exp(0.27 - 0.28m_i)}{1 + \exp(0.27 - 0.28m_i)} \right).$$

The experimental variabilities were $\sigma_m = 0.5$ and $\sigma_d = 1.8$. Notice that due to the rounding up of the MIC, the scatterplot shifts to the right.

3. Determining DIA Breakpoints

Guidelines state that between 300-500 pathogens “representative of all species likely to be tested” are to be chosen for breakpoint determination (NCCLS, 1995). The MIC distribution in Figure 2 represents the typical bimodal situation where the error-rate bounded method has been used satisfactorily. Since the observed discrepancy percentages are kept low, numerous breakpoint sets satisfy the discrepancy restrictions. Although single numerical summaries of the observed discrepancy percentages have been suggested as a method to determine the “best” set of breakpoints (Brunden, Zurenko, and Kapik, 1992), the final choice of breakpoints is usually a combination of test results, previous breakpoint studies, breakpoints of similar drugs, and experimenter preference. This ad hoc approach to determination has obscured many of the sampling problems with this procedure. It is only recently, due to the increase in pathogens near the MIC intermediate region, that the procedure has been reevaluated.

With the model approach, the performance of each test can be described for a specific set of breakpoints. While similar in approach to the earlier proposed numerical summaries of Brunden et al. (1992), this performance summary does not directly rely on the observed discrepancy percentages (distribution of pathogens). Instead the observed results are used to estimate the model parameters which in turn determine the breakpoints. This not only results in more consistent results but also a more intuitive less sample dependent approach to compare sets of breakpoints. Instead of trying to assess whether seven observed very major discrepancies is much worse than four, one can view how the DIA test’s performance changes with a shift in the upper breakpoint.

Consider the MIC test with breakpoints labeled M_L and M_U . Since these breakpoints are largely based on pharmacokinetics and pharmacodynamics of the drug, they also serve to separate the true MICs into the three classification regions. For example, a true MIC less than M_L would be considered susceptible and a true MIC greater than M_U would be resistant. To describe the test performance, consider the probability that a pathogen with true MIC m is correctly classified. This probability

$$p_{\text{MIC}}(m) = \begin{cases} \text{pr}(x \leq M_L) = \Phi\left(\frac{M_L - m}{\sigma_m}\right) & m \leq M_L \\ \text{pr}(M_L < x < M_U) = \Phi\left(\frac{M_U - 1 - m}{\sigma_m}\right) - \Phi\left(\frac{M_L - m}{\sigma_m}\right) & M_L < m < M_U \\ \text{pr}(x \geq M_U) = 1 - \Phi\left(\frac{M_U - 1 - m}{\sigma_m}\right) & m \geq M_U \end{cases}$$

where Φ is the standard Normal CDF. Figure 3 plots this probability over a range of MIC values using $\sigma_m = .5$ and the common one dilution intermediate region (1,3). Due to the test variability and rounding up, this probability decreases as the MIC approaches a breakpoint from below (see also the last two pathogens in Figure 1).

The same curve can be constructed for the DIA test given breakpoints D_L and D_U and (3). The probability

$$p_{\text{DIA}}(m) = \begin{cases} \text{pr}(y \geq D_U) = 1 - \Phi\left(\frac{D_U - .5 - d}{\sigma_d}\right) & m \leq M_L \\ \text{pr}(D_L < y < D_U) = \Phi\left(\frac{D_U - .5 - d}{\sigma_d}\right) - \Phi\left(\frac{D_L + .5 - d}{\sigma_d}\right) & M_L < m < M_U \\ \text{pr}(y \leq D_L) = \Phi\left(\frac{D_L + .5 - d}{\sigma_d}\right) & m \geq M_U \end{cases}$$

Notice that since the diameter test rounds to the nearest millimeter and DIA is inversely related to MIC, the formula is slightly different.

To calibrate the two tests, DIA breakpoints are found which minimize the loss function

$$L = \int_{-\infty}^{\infty} \min(0, p_{\text{DIA}}(u) - p_{\text{MIC}}(u))^2 \pi(u) du$$

where π is a distribution which weights the MIC values. For example, one choice of weights would be the underlying distribution of pathogens. Others would be to more heavily weight those values near the MIC breakpoints or give equal weight to all values. The min function is used because there is interest in determining breakpoints such that the DIA test is at least as good as the MIC test. There is no loss if the probability of the DIA test is higher at a specific MIC.

Figure 4 displays these DIA curves for four sets of breakpoints and the same model parameters as Figure 2. Also shown is $-2\log(L)$ when π gives equal weight to all possible MIC values (i.e., $w_i = 1$). In this case, the best set is (26,32). The DIA test performs similarly in the susceptible and resistant regions and outperforms the MIC test in the intermediate region. Arguably any of these sets could be the best and a different loss function may result in a different choice. From all the sets, one can see that lowering the upper breakpoint or increasing the lower breakpoint results in increased performance in the susceptible or

resistant region but with decrease in performance in the intermediate region. This is similar to the trade-off between observed very major/major and minor discrepancies except in this case the results do not depend on the number and location of the pathogens, these having been factored into the model parameters estimates.

4. Bayesian Inference

To use this breakpoint determination approach in practice, model parameters, specifically σ_m, σ_d , and the β 's, must first be estimated from a scatterplot. Bayesian inference is used to obtain the joint posterior of parameters, $\mathcal{P} = \{\sigma_m, \sigma_d, \beta_0, \beta_1, \beta_2, k, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{p}\}$, where the bold characters represent vectors of length k . Calculation of this distribution by analytic or numeric integration is very difficult due to the varying number of parameters in the posterior, the hidden true MIC and DIA values, and the rounding of both the MIC and DIA tests.

4.1 The Prior

We assume very little a priori knowledge about the parameters except for the testing standard deviations σ_m and σ_d where previous quality control studies (repeat drug/pathogen experiments) provide a range of likely values. We also assume independence among parameters describing different components of the model.

For the parameters used to determine breakpoints, we assume σ_m is Gamma with mean 0.5 and standard deviation 0.204 and σ_d is Gamma with mean 1.8 and standard deviation 0.569. For the drug-specific parameters in (3), we assume $\beta_0 > 0$ and $\pi(\beta_0, \beta_1, \beta_2) \propto 1$.

For the mixture of Normals, we bound the number of Normals such that $k \leq 35$ and assume k is discrete Uniform, $\text{pr}(k) = 1/35$. The choice of 35 was made arbitrarily but no analyses (to date) have resulted in $k > 30$. To keep the Normal mixture identifiable, we assume

$$\mu_1 < \mu_2 < \dots < \mu_k$$

and each mean is marginally Uniform between -12 and 12 so

$$\pi(\boldsymbol{\mu}) = k! \left(\frac{1}{24}\right)^k$$

The choice of (-12,12) was based on the observed range of MICs in scatterplots but can be made wider. For the weights, we assume $\pi(w) = \text{Dirichlet}(1, 1, 1, \dots, 1)$.

Since there is a chance a Normal distribution may not be associated with any observations, the standard non-informative reference priors for the standard deviations will result in an improper posterior distribution. We use the partially proper prior described in Roeder and Wasserman (1997) to avoid this. We assume

$$\pi(\boldsymbol{\sigma}) = \prod^k \pi(\sigma_j|S)$$

where

$$\pi(\sigma_j|S) \sim \chi_1(S) \quad \pi(S) \propto S^{-1}$$

The effect of S is to shrink the σ_j 's down to an unspecified constant S .

4.2 Computing the Posterior

For each pathogen in the scatterplot, there is a hidden true MIC m_i . To more easily keep track of this mean in terms of the Normal mixture, we use a latent group identifier z_i , which denotes from which Normal distribution in the mixture the true MIC is selected. We use Markov chain Monte Carlo (Smith and Roberts, 1993), specifically reversible jump MCMC (Richardson and Green, 1997) to approximate the joint posterior of the model parameters. We run the chain over the set of model parameters, \mathcal{P} and the unobserved true MICs and group identifiers, \mathcal{U} .

$$(\mathcal{U}^1, \mathcal{P}^1), (\mathcal{U}^2, \mathcal{P}^2), (\mathcal{U}^3, \mathcal{P}^3), \dots$$

Each step or “cycle” of the chain is the result of smaller steps, each being a Metropolis-Hastings or Gibbs update which modifies a component of the larger state. Appendix A details these steps.

This Markov chain has equilibrium distribution equal to the posterior

$$\pi(\{x, y\}|m, \beta, \sigma_m, \sigma_d)\pi(m|k, z, \mu, \sigma, w)\pi(\mu|k)\pi(\sigma|k, S)\pi(w|k)\pi(S)\pi(k, \beta, \sigma_m, \sigma_d) \quad (5)$$

where the first term is a product probabilities representing areas under the standard Normal curve. The second term is a product of weighted Normal densities and the remaining terms

are the prior. For each scatterplot, a chain of 6,000 cycles was run on an IBM RISC/6000 computer. Estimates of the DIA breakpoint posterior distribution are based on subsampling every 10th cycle after an initial burn in of 1000; these dimensions were chosen by informal assessment of time-series plots under various starting positions.

5. Results

5.1 Simulation Study

To demonstrate the improved consistency in results (i.e, estimator is more concentrated), two versions of repeated tests were investigated. In the first situation, a set of 300 pathogens were randomly chosen and 100 scatterplots were generated based on this set. In this case, only the experimental variability and relationship between MIC and DIA (in terms of rounding) will affect the results. In the second situation, a new set of pathogens was randomly chosen for each experiment. This introduces another source of variability into the scatterplot.

For each of the 200 scatterplots, breakpoints were determined using both the error-rate bounded method and the new methodology. For the error-rate bounded method, the best set of breakpoints was the set which maximized $1/E$ where

$$E = .6(\%VMaj) + .3(\%Maj) + .1(\%Min)$$

and the percentage of very major and major discrepancies were below 1.5% and 3.0% respectively. This index gives more weight to a “more severe” discrepancy and roughly follows the discrepancy limits 1.5%, 3.0%, and 10.0%. For the new methodology, breakpoints were determined for each set of parameters in the posterior using the loss function and the mode of the breakpoint set posterior distribution (never had ties) was selected.

The scatterplots were created using the model parameters in Figure 2. Recall this is a situation where there were a few pathogens near the MIC breakpoints which is favorable to the error-rate bounded method with the current tolerable percentages. Table 1 summarizes the results where the same set of pathogens are used throughout. Table 2 summarizes the situation when a new set of pathogens was chosen each experiment. The first number in each cell represents the frequency this pair of breakpoints was chosen using the error-rate bounded procedure and the second number (bold) represents the new methodology.

In both cases, there is far less variability in the selection with the new methodology. Since the parameters that generated the scatterplots were used in Figure 4, the ideal breakpoints in terms of performance should be around (26,32) and in terms of the index E around (25,32). This latter set is based on using the known model parameters to compute the probability of a

very major, major, and minor discrepancy. The model approach selected a breakpoint set in the rectangular region $(25 \pm 2, 32 \pm 1)$ 100% of the time while the error-rate bounded selected a set in this region 66% and 54% of the time. Since the model explicitly describes sources of variability in the scatterplot, there is less variability in the results. Also, one of the major limitations of the error-rate bounded procedure is its dependence on the pathogens selected and this is demonstrated by the increased variability in situation 2. For the new procedure, however, only the experimental variabilities and the true DIA/MIC relationship are used to determine breakpoints so as long as the new set of observed results provide similar estimates of these parameters, the breakpoint set will be similar. Thus, for the model-approach, there is only a slight difference in the two situations and this is reflected in the results.

Application to Published Scatterplots

This methodology was applied to three published scatterplots of the drug Lomefloxacin (Jones et al., 1988; Cormican and Jones, 1995; Cormican et al., 1996). The 1988 scatterplot gave preliminary recommendations for zone diameter breakpoints which were later accepted by NCCLS. The latter two were the result of a reevaluation of Lomefloxacin and both articles suggest a 2mm decrease in the breakpoints. The results of the first scatterplot, for several choices of DIA breakpoints, are summarized in Table 3 and the latter two scatterplots are summarized in Table 4.

In Table 3, each set of breakpoint satisfies the 1.5% and 3.0% restrictions but all have a minor percentage slightly greater than 10%. While the set (18,22) was chosen, any of these other sets would have sufficed. Was the set (18,22) chosen because it had the fewest minor errors or the lowest upper breakpoint? Why is it better than the set (19,23) which has over half as many very major discrepancies? In Table 4, similar breakpoint sets are presented. In these cases, the set (16,20) was selected but there are similar questions concerning this choice.

While an index could be formulated to include the factors that were used in this decision, the simulation study has shown that any decision based on observed discrepancy percentages is highly variable and can likely change with a new study. On the other hand, the model approach was shown to have consistent results and allows comparisons of breakpoints sets using performance curves. The model approach results, based on 500 likely sets of parameters, are summarized in Table 5. For the first scatterplot, over 70% of the time the set (18,23) resulted in the most favorable performance curve. In reference to the choice (18,22), this suggests that a reduction in the performance in the susceptibility region ($22 \rightarrow 23$) is outweighed by the increase in performance in the intermediate region. Without plotting the

curves, however, it is difficult to know exactly what this trade-off entails.

The main reason for choosing these three scatterplots was to see whether the new methodology when applied to the latter two scatterplots would agree with a reduction in the breakpoints. As shown in Table 5, this is the case but the two scatterplots do not suggest the same choice of breakpoints. Table 6 summarizes the marginal posterior distributions of the parameters that are used in determining the breakpoints. The key differences are the estimates of experimental variability and the asymptote (β_0). Since the 1996 scatterplot is a collection of results performed at numerous labs and the model does not factor in lab to lab variability, it is not surprising the variance results are different. Why the MIC variability would be smaller is somewhat surprising but this may be due to the confounding of test and lab variability. As shown in Figure 5, this increase in variability combined with the higher asymptote explains the higher upper breakpoint and similar lower breakpoint. It is interesting to note that the 1988 curve appears parallel with the 1998 curve but about 3 mm higher. This is also reflected in the breakpoint results. Two possible reasons for this could be a change in the disk or media used in the diffusion test or a lab to lab difference. In either case, further investigation is warranted.

6. Discussion

Because of the increasing number of moderately susceptible and resistant pathogens, choosing appropriate breakpoints is becoming more and more a statistical problem. The error-rate bounded method, although simple to implement, suffers from being too sample dependent. In this paper, a model has been presented which explicitly describes certain uncertainties in a scatterplot and is part of an alternative approach to breakpoint determination. Instead of using the observed results directly in breakpoint determination (e.g., very major, major, and minor percentages), they are used to estimate the model parameters which in turn describe the performance of each test. Breakpoints are selected such that the test performances are as similar as possible. In a sense, the model is filtering out as much explainable noise as possible in the scatterplot before determining breakpoints while the previous method does not.

While eliminating disagreements between two tests (error-rate bounded method) is often an appropriate method of calibration, it does not necessarily work when the two tests behave differently. Since the MIC test rounds up, it is more conservative (less likely to say a pathogen is susceptible) than the disk test. Simply bounding the discrepancy percentages does not take this behavior into account. For example, this rounding means that it is more natural

to observe a very major discrepancy than a major discrepancy but this currently has the lowest percentage restriction. The performance curves, on the other hand, are based on each test's ability (in terms of probability) to classify an isolate and therefore uses these rounding procedures directly in the calculations.

While only one simulation study has been performed, there are many factors that will affect the results. Specifically, the inherent variability of each test, the number and location of the isolates investigated, and the MIC/DIA relationship. However, by demonstrating how much more variable the results are when the error-rate bounded method supposedly works satisfactorily, it is not felt that other simulation studies are necessary. For example, other small scale simulation experiments where there was a larger percentage of pathogens near the intermediate range resulted in even greater disparity in the results. The error-rate bounded method resulted in very wide zones due to the fact that the current discrepancy restrictions of 1.5% and 3.0% were too low for this situation.

The performance curve depends not only on the choice of breakpoints but also the model selected. While it is felt this model adequately describes scatterplots, variations such as allowing for increasing variance in the DIA test or a different relationship between DIA and MIC can be included and are currently under review. This same methodology can also be applied in device testing where NCCLS approved procedures (usually MIC lab tests) are compared with device results. At the present time, a variation of the error-rate bounded procedure is performed where device and lab results must be more than one dilution apart in order to be considered a discrepancy. With this methodology, one would assume that the MIC and device (d) results are linearly related

$$d_i = \beta_0 + \beta_1 m_i$$

and could use this same procedure to see if the device is unbiased ($\beta_0 = 0$ and $\beta_1 = 1$). It would also take into account the location of the isolates investigated and eliminate the need for similar tolerable percentages.

ACKNOWLEDGMENTS

I wish to thank the M23 working group of NCCLS for introducing me to this problem and the helpful comments throughout the process. I'd also like to thank George McCabe, George Casella, Ray Carroll, Michael Black, and Michael Newton for helpful discussions and comments.

REFERENCES

- Brunden, M.N., Zurenko, G.E., and Kapik, B. (1992). Modification of the error-rate bounded classification scheme for use with two MIC breakpoints. *Diagnostic microbiology and infectious disease* **15(2)**, 135-140.
- Cormican, M.G., Erwin, M.S. and Jones, R.N. (1996). Multi-center validation of proposed disk diffusion susceptibility testing interpretive criteria for Lomefloxacin using more than 1,500 clinical isolates. *Diagnostic Microbiology & Infectious Disease* **24(3)**, 169-172.
- Cormican, M.G. and Jones, R.N. (1995). Reevaluations of disk diffusion susceptibility testing interpretive criteria for Lomefloxacin and Norfloxacin using fluoroquinolone-resistant isolates. *Diagnostic Microbiology & Infectious Disease* **21(4)**, 227-230.
- Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711-732.
- Jones, R.N., Aldridge, K.E., Barry, A.L., Fuchs, P.C., Gerlach, E.H., and Pfaller, M.A., Washington II, J.A. (1988). Multicenter in vitro evaluation of Lomefloxacin (NY-198, SC-47111), including tests against nearly 7,000 bacterial isolates and preliminary recommendations for susceptibility testing. *Diagnostic Microbiology & Infectious Disease* **10(4)**, 221-240.
- Metzler, C.M. and DeHaan, R.M. (1974). Susceptibility Tests of Anaerobic Bacteria: Statistical and Clinical Considerations. *The Journal of Infectious Diseases* **130(6)**, 588-594.
- National Committee for Clinical Laboratory Standards (NCCLS) (1994). Development of in vitro susceptibility testing criteria and quality control parameters; Approved guideline M23-A. Villanova, PA: NCCLS.
- Richardson, S. and Green, P.J. (1997). On Bayesian analysis with an unknown number of components. *Journal of the Royal Statistical Society, Series B* **59**, 731-792.
- Roeder, K. and Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association* **92**, 894-902.
- Smith, A.F.M. and Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society Series B* **55**, 3-24.

APPENDIX I

Metropolis-Hastings Algorithms

We run a Markov chain over the set of model parameters and unobserved variables. One complete step in the Markov chain is produced by a sequence of smaller Metropolis-Hastings (MH) steps which modify different aspects of this set. Each MH algorithm is defined by a proposal distribution, having density $q(s, s')$, which indicates the probability density of sampling the new state s' given the current state s . The MH ratio,

$$r = \frac{\pi(s')q(s', s)}{\pi(s)q(s, s')}$$

is calculated and with probability $\min(r, 1)$ the Markov chain moves to s' , otherwise it stays put. The following sections describe the proposal distribution for each of the MH steps and the form of the MH ratio.

Updating the β 's, σ_m , σ_d

We update each parameter individually using a normal centered at the current value. In the case of the two standard deviations, this normal is folded over zero to avoid negative standard deviations. Both the normal and folded normal centered at the current value are symmetric proposal distributions so the MH ratio is simply the ratio of the likelihood and prior. For example, for σ_d , the ratio of likelihood and prior reduces down to

$$\frac{\pi(\{x, y\} | m, \beta, \sigma_m, \sigma_d^*) \pi(\sigma_d^*)}{\pi(\{x, y\} | m, \beta, \sigma_m, \sigma_d) \pi(\sigma_d)},$$

where the first term is a product of probabilities representing the area under the standard Normal as defined by the DIA test and the second term is the prior.

Normal Distributions in Mixture

A Gibbs step is used to update the weights w , means μ , and standard deviations σ . For the weights, the full conditional is Dirichlet($n_1 + 1, n_2 + 1, \dots, n_k + 1$), where n_j is the number of MICs sampled from the j th Normal. For each mean, the full conditional is a truncated Normal with mean \bar{m}_j and standard deviation σ_j/n_j (if $n_j > 0$) or truncated Beta distribution defined by the j th order statistic of k Uniforms. In this situation, it is assumed $\mu_0 = -12$ and $\mu_{k+1} = 12$.

The full conditional of σ_j is a scaled inverse-chi distribution with $n_j + 1$ degrees of freedom where the scale factor is

$$S + (n_j - 1)s_j^2 + n_j(\bar{m}_j - \mu_j)^2.$$

A Gibbs step is also used to update the standard deviation scaling parameter S . The full conditional is Gamma($\frac{k-1}{2}$, $2 \sum \sigma_j^{-2}$).

Latent Group Identifiers

The full conditional for variable z_i is Multinomial where

$$\text{pr}(z_i = j) = \frac{w_j \text{N}(m_i; \mu_j, \sigma_j)}{\sum_{l=1}^k w_l \text{N}(m_i; \mu_l, \sigma_l)}. \quad (6)$$

Instead of performing a Gibbs step, a Metropolis-Hastings step, proposed by Roeder and Wasserman (1997), is used. Instead of all k groups as potential candidates, the proposal distribution chooses uniformly from the nearest neighbors $z_i - 1$ and $z_i + 1$. When $z_i = 1$, the only potential candidate is 2 and when $z_i = k$, the only potential candidate is $k - 1$. The acceptance probability is based on a ratio of probabilities using (6) and possibly a 2 or $\frac{1}{2}$ if near a boundary group. Note that this avoids calculation of the denominator in (6).

Reversible Jump

To update k , the number of Normals in the mixture, a splitting/combining step similar to that proposed by Richardson and Green (1997) is used. This increases/decreases the value of k by 1 and makes appropriate changes to the Normal distribution parameters (w, μ, σ) . In each update, either a split or combination step is chosen with probability 1/2. For the combination step, an adjacent pair (j_1, j_2) of Normal distributions is randomly selected and merged together such that

$$\begin{aligned} w_{j^*} &= w_{j_1} + w_{j_2} \\ w_{j^*} \mu_{j^*} &= w_{j_1} \mu_{j_1} + w_{j_2} \mu_{j_2} \\ w_{j^*} (\mu_{j^*}^2 + \sigma_{j^*}^2) &= w_{j_1} (\mu_{j_1}^2 + \sigma_{j_1}^2) + w_{j_2} (\mu_{j_2}^2 + \sigma_{j_2}^2) \end{aligned} \quad (7)$$

For the split step, a Normal distribution is randomly selected and three Uniform random variables (u_1, u_2, u_3) are generated to specify the new parameters making sure that (7) holds.

In addition, the n_j pathogens are randomly allocated to the two new groups. This is done analogously to the Gibbs update of the group identifiers. For further details, see Richardson and Green (1997).

TABLE 1: Number of times each DIA breakpoint set was selected when using the same set of isolates for each scatterplot^a

Lower Limit	Upper Limit									
	28	29	30	31	32	33	34	35		
22		1		1	1	3	2	1		
23	1	1	4	4	1	7	3	5		
24		1	1	6	12	8	8	4	1	
25		1	2	2	1	6	35	1	1	
26			2	3	7	6	30	5	2	
27				4		3	1	2	2	2
28						3			1	
29								2		

^aThe left number in each cell is for the error-rate bounded procedure and the right for the new methodology.

TABLE 2: Number of times each DIA breakpoint set was selected when using a different set of isolates for each scatterplot^a

Lower Limit	Upper Limit										
	27	28	29	30	31	32	33	34	35		
20								1			
21		1			1			1	1		
22				4	3	2	1	3			
23	2		1	2	3	3	6	1	4	1	
24			1	5	4	2	13	2	3	1	
25				3	5	8	6	23	5	9	1
26				1	4	1	4	34	3	7	
27					3		3		1	1	2
28							1				
29									3		

^aThe left number in each cell is for the error-rate bounded procedure and the right for the new methodology.

TABLE 3: Summary of 1988 Scatterplot ($n = 829$)

Brkpts		Number of Isolates in Each Category					
		SS ^a	RR ^b	VMAJ	MAJ	MIN	INDEX ^c
19	23	587	131	3	2	86	75.36
18	23	587	116	3	1	94	72.09
20	24	569	150	1	6	91	72.09
19	24	569	131	1	2	104	71.47
18	24	569	116	1	1	112	68.51
18	22	598	116	7	1	85	63.77
17	23	587	88	3	1	113	61.87
17	22	598	88	7	1	104	55.64

^aClassified as susceptible on each test^bClassified as resistant on each test^cIndex used in the simulation study**TABLE 4:** Summary of 1995 and 1996 Scatterplots

Brkpts		1995 ($n = 299$)						1996 ($n = 1501$)					
		# of Isolates in Each Category						# of Isolates in Each Category					
		SS ^a	RR ^b	VMAJ	MAJ	MIN	INDEX ^c	SS	RR	VMAJ	MAJ	MIN	INDEX
15	21	79	200	1	0	15	143.52	1237	132	4	1	106	112.86
15	20	82	200	2	0	11	130.00	1285	132	7	1	61	141.60
16	20	82	201	2	0	12	124.58	1285	136	7	2	59	140.28
15	19	83	200	2	0	10	135.91	1300	132	14	1	45	113.71
14	20	82	194	2	0	15	110.74	1285	124	7	1	68	132.83
16	21	79	201	1	0	10	186.88	1237	136	4	2	104	112.01
14	21	79	194	1	0	19	119.60	1237	124	4	1	113	107.21
14	19	83	194	2	0	14	115.00	1300	124	14	1	52	107.99

^aClassified as susceptible on each test^bClassified as resistant on each test^cIndex used in the simulation study

TABLE 5: Potential Breakpoints for Lomefloxacin Using Model-Based Approach

1988		1995		1996	
Brkpts	Freq %	Brkpts	Freq %	Brkpts	Freq %
(18,23)	71.8	(14,19)	37.5	(14,21)	54.0
(17,22)	20.2	(15,20)	35.0	(14,20)	25.5
(18,22)	4.2	(14,20)	17.0	(15,21)	19.5
		(15,19)	6.5		

Table 6: Summary of Marginal Posteriors for 1995 and 1996 Scatterplots

Parameter	1995			1996		
	Percentiles			Percentiles		
	2.5	50.0	97.5	2.5	50.0	97.5
σ_m	0.28	0.47	0.66	0.13	0.26	0.38
σ_d	1.72	2.03	2.35	3.37	3.52	3.68
β_0	27.91	29.26	31.37	30.16	30.72	31.47
β_1	0.79	1.12	1.44	1.02	1.14	1.27
β_2	0.41	0.47	0.54	0.44	0.48	0.52

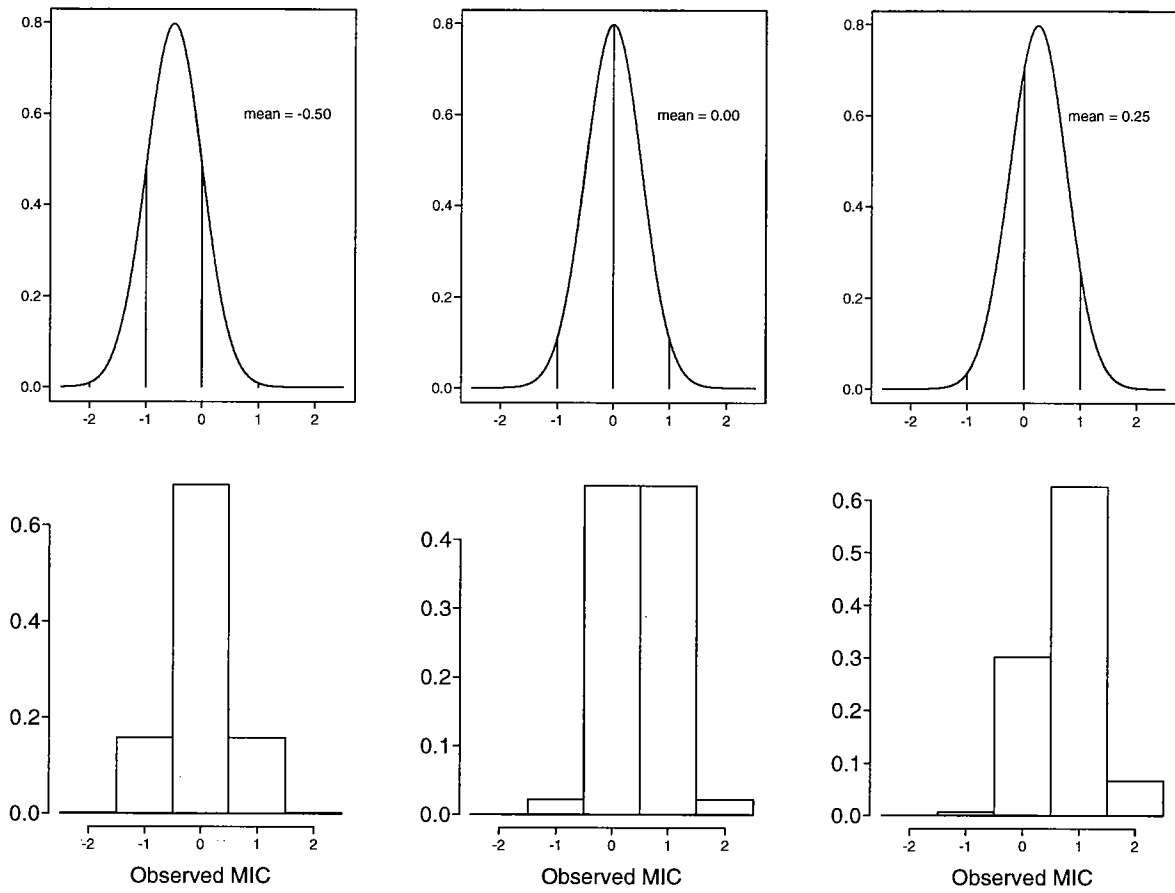


Figure 1: The upper three panels show Normal distributions with various means and standard deviation $\sigma_m = 0.5$. The lower three panels show the observed MIC probability distribution after rounding up.

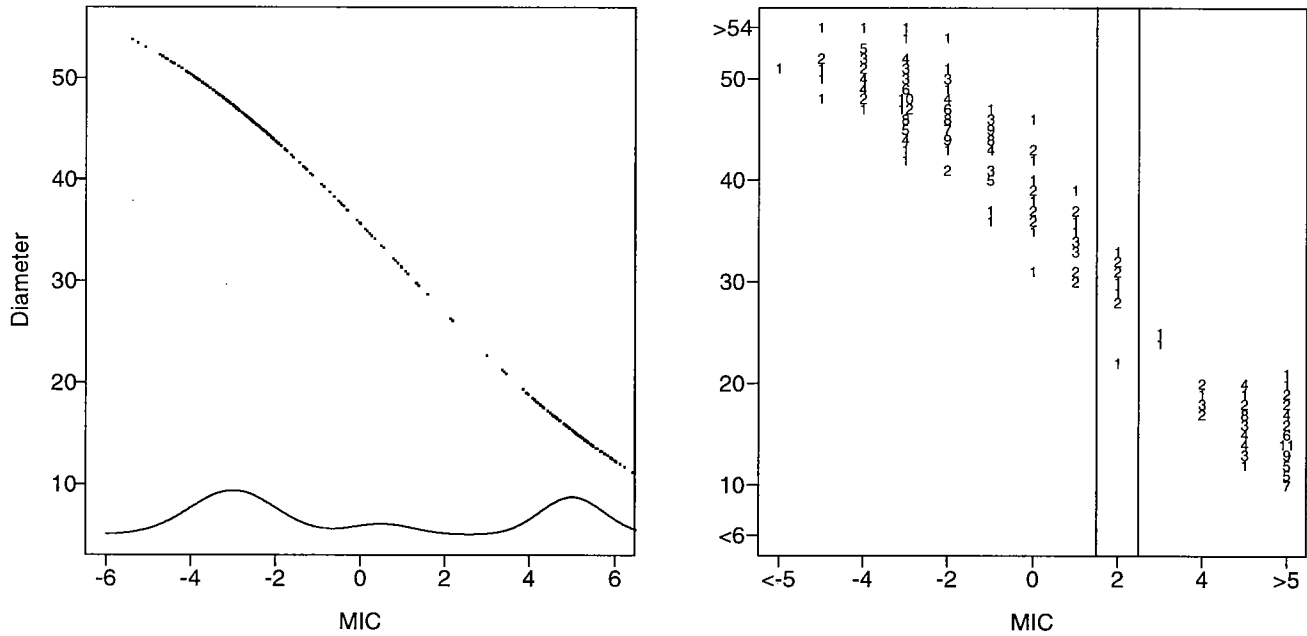


Figure 2: Formation of a scatterplot. The left panel contains 300 randomly sampled MICs and the associated DIA values. The right panel introduces the testing errors and rounding.

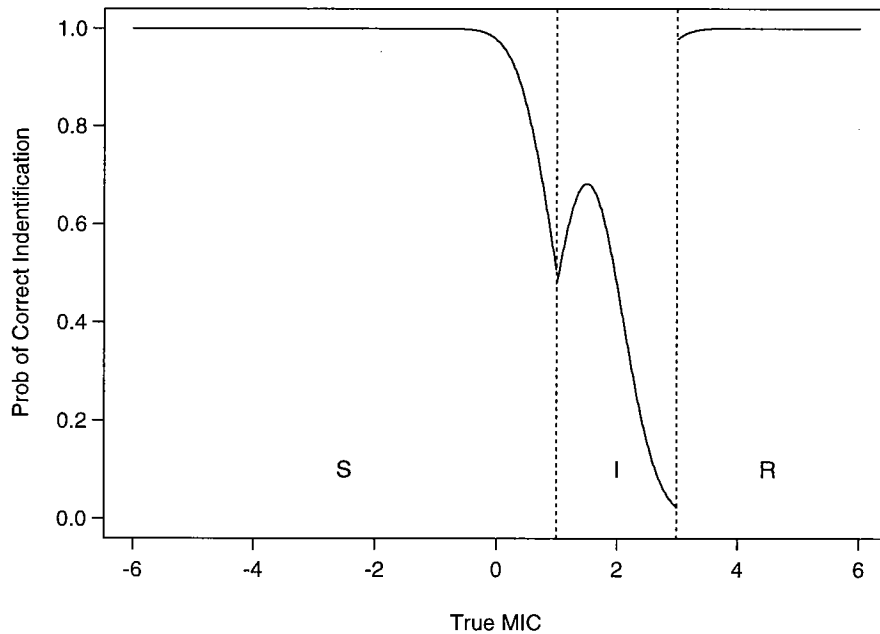


Figure 3: Probability that a pathogen with true MIC m will be classified correctly by the MIC test ($\sigma_m = 0.5$ and $(M_L, M_U) = (1, 3)$).

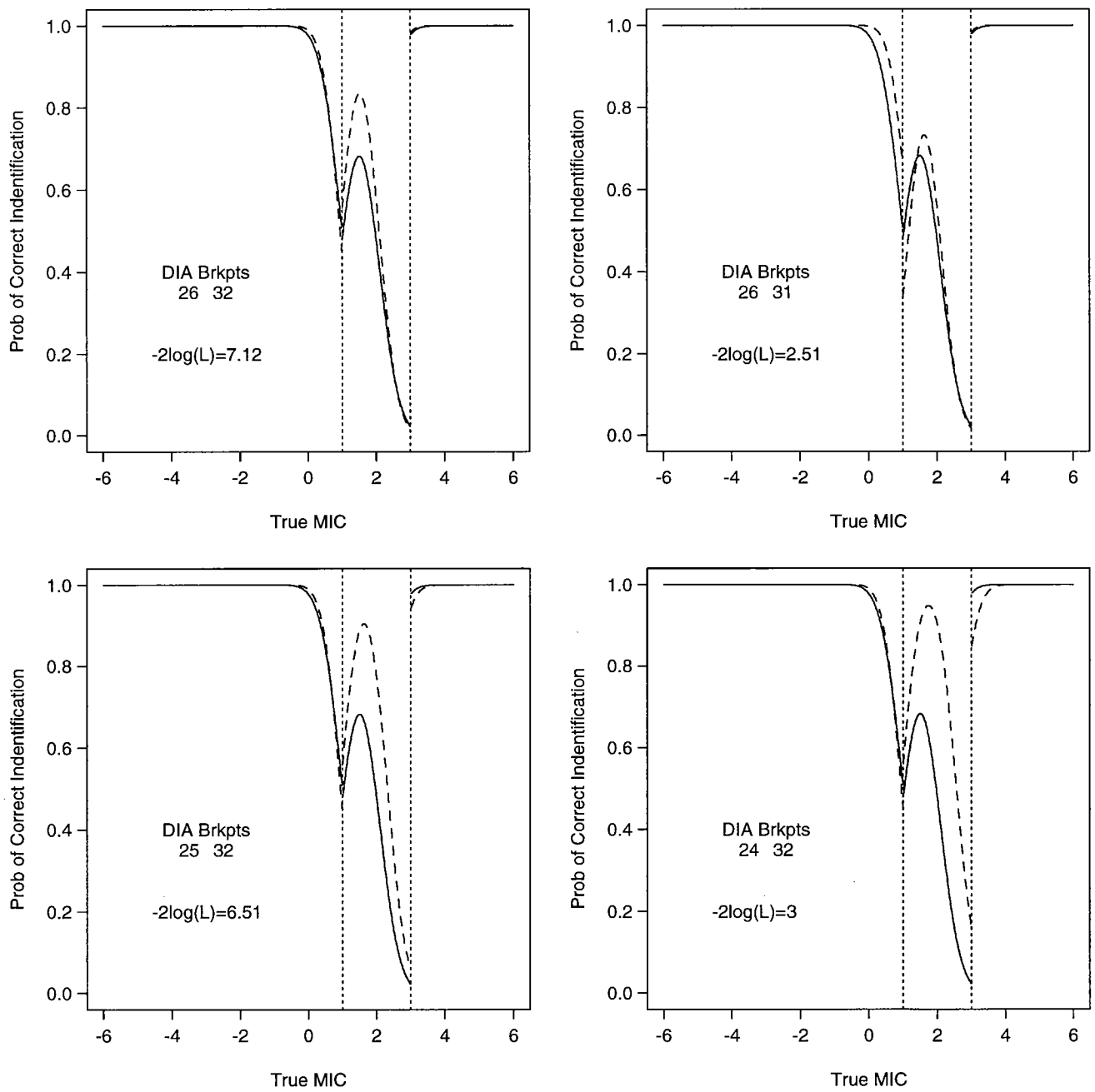


Figure 4: DIA test probability curves (dashed) for various breakpoints sets, $\sigma_y = 1.8$ and $d = 62.8(\exp(.27 - .28m)/(1 + \exp(.27 - .28m)))$. The MIC performance curve (solid) is shown for reference.

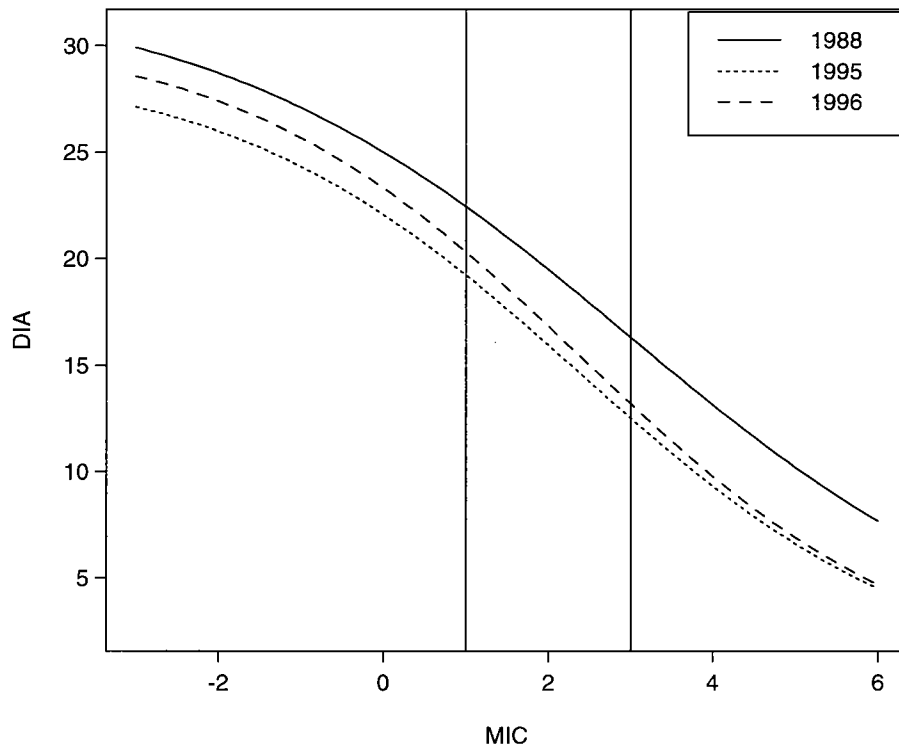


Figure 5: Estimated MIC/DIA relationship for the three scatterplots.