Open Access Dissertations   Theses and Dissertations

Fall 2013

Disk Diffusion Breakpoint Determination Using a Bayesian Nonparametric Variation of the Errors-in-Variables Model

Glen Richard DePalma
Purdue University

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By _Glen DePalma_

Entitled _Disk Diffusion Breakpoint Determination Using a Bayesian Nonparametric Variation of the Errors-in-Variables Model_

For the degree of _Doctor of Philosophy_

Is approved by the final examining committee:

Bruce Craig
_____
Chair

Xiao Wang
_____

Jayanta Ghosh
_____

_____

Chuanhai Liu
_____

_____

George McCabe
_____

_____

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): _Bruce Craig_____

_____

Approved by: _Jun Xie_____   08/21/2013
Head of the Graduate Program   Date

DISK DIFFUSION BREAKPOINT DETERMINATION USING A BAYESIAN

NONPARAMETRIC VARIATION OF THE ERRORS-IN-VARIABLES MODEL

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Glen DePalma

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2013

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

I have had the privilege to meet and interact with many talented students and faculty throughout my time as a Ph.D. student. I wish to highlight and thank a select few that have had a dramatic impact on my education during this time.

First and foremost, I would like to express my gratitude to my advisor Professor Bruce A. Craig. His advice and ideas not only helped me as a Ph.D. student, but will be used throughout my entire professional career. This thesis is a direct result of his mentorship and research guidance.

I want to thank Professor Rebecca Doerge who, as department head, also provided mentorship and guidance throughout my time as a Ph.D. student. She has been a source of great advice and has generously supported me with funding to showcase my research at various conferences.

I want to thank Professor Laura P. Sands for research assistant support the past three years. She has been a great mentor and I learned a lot about interdisciplinary research, a must for today's statistician. In addition, she has supported my attendance at conferences to showcase my interdisciplinary research.

I would also like to thank Regina Becker who, as a staff member for the statistical consulting service and consultant to Statistics in the Community (StatCom), provided unforgettable advice on what it means to be a statistical consultant.

I want to thank my committee members: Professor Chuanhai Liu, Professor Xiao Wang, Professor George McCabe, and Professor Jayanta K. Ghosh. I have been inspired by their research and work ethic. Of particular inspiration are Professors

Liu and Ghosh who, after taking several of their courses, have challenged me to think deeply on many current statistical topics.

I want to thank the statistics department staff. Doug Crabill and Jarrod Welsh have been terrific in providing computing resources and helped with several special requests. Mary Roe, Becca Pillon, Diane Martin, Teena Erwin, and Marian Duncan have always been extremely receptive to my requests.

Finally, I would like to acknowledge my parents who have supported me generously throughout this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure   Page

ABSTRACT

DePalma, Glen Ph.D., Purdue University, December 2013. Disk Diffusion Breakpoint Determination Using a Bayesian Nonparametric Variation of the Errors-in-Variables Model . Major Professor: Bruce A. Craig.

Drug dilution (MIC) and disk diffusion (DIA) are the two most common antimicrobial susceptibility tests used by hospitals and clinics to determine an unknown pathogen's susceptibility to various antibiotics. Both tests use breakpoints to classify the pathogen as either susceptible, indeterminant, or resistant to each drug under consideration. While the determination of these drug-specific MIC classification breakpoints is straightforward, determination of comparable DIA breakpoints is not. It is this issue that motivates this research.

Traditionally, the error-rate bounded (ERB) method has been used to calibrate the two tests. This procedure involves determining DIA breakpoints which minimize the observed discrepancies between results generated from both tests over a wide range of pathogen strains (or isolates). While simple and intuitive, this approach is very sample dependent and lacks precision.

Model-based approaches were first proposed in 2000. These approaches model the underlying true relationship between the two tests and thus focuses on calibrating the probabilities of classification rather than the observed test results. Both a Bayesian parametric (2000) and a frequentist nonparametric (2008) procedure have been proposed. However, due to various computational difficulties and an absence of easy to use software for clinicians, neither approach has been adopted for use.

In this thesis, we present a novel Bayesian nonparametric model that combines the strengths of the previous two model-based approaches. The resulting approach

provides the flexibility of a nonparametric model to describe the true DIA/MIC relationship within a Bayesian framework in order to extract as much information as possible from the observed data. We demonstrate the strength of this approach via a series of simulation studies comparing it to the ERB and previous model-based approaches using breakpoint determination accuracy and model fit statistics as the comparison criteria. We conclude with applications to several real data sets and a discussion regarding software implementation and future work.

1. ANTIMICROBIAL SUSCEPTIBILITY TESTING

1.1 Introduction to Antibiotics

An antibiotic is an antimicrobial agent that fends off bacteria by either killing it or stopping its reproduction. Antibiotics are considered one of the greatest advancements in modern medicine and is one of the most frequently prescribed classes of medications. They have saved countless lives by treating bacterial infections and preventing bacterial diseases [1].

Antibiotics first wide-spread use was in the 1940s to treat Allied soldiers struck with life threatening infections [2]. Now, we are completely dependent on them, not only for the prevention of infections during most surgical procedures, including organ and prosthetic transplants, but also for the treatment of infectious diseases [3]. There are currently over 100 different antibiotics available to treat major and minor infections.

Antibiotic action was first witnessed in 1895 by Vincenzo Tiberio, physician of the University of Naples, from a mold culture in a waterwell. Tiberio noticed every time the waterwell was cleaned of mold, drinkers would get sick. It turns out the mold was actually killing bacteria in the water, thereby, making it drinkable.

In 1928, Alexander Fleming again observed the antibiotic action from a similar fungus after accidentally leaving a Petri dish uncovered [4]. Fleming hypothesized the antibiotic action was caused by a compound in the fungus he named Penicillin. He characterized the biological properties of this compound and was awarded the Nobel Prize in Medicine along with Howard Florey and Ernst Chain in 1945. Around this

same time, Penicillin became commercially available and was quickly exploited for the benefit of medicine. It is still one of the most commonly prescribed antibiotics.

The first commercially available antibacterial agent was Prontosil (a sulfonamide drug), developed at Bayer Laboratories in Germany by Gerhard Domagk in 1932. Domagk received the Nobel Prize for Medicine in 1939 for his work [5]. Penicillin was in use during this time (Cecil George Paine recorded the first cure with Penicillin in 1930 [6]), but many more studies of its properties were made before it became commercially available.

1.1.1 Antibiotic Resistance

After the introduction of Penicillin and Prontosil into human therapeutics, other compounds soon followed. As with every living organism, however, bacteria continue to evolve and strains resistant to antibiotics emerge [7]. In the last couple of decades, bacteria have become more and more resistant at an alarming rate. It is estimated that seventy percent of the bacteria strains that were once treatable by major antibiotics are now resistant to at least one of them.

Two main factors have catalyzed this recent increase in antimicrobial resistance:

1. Increased antimicrobial use
2. Misuse of antibiotics

The use of antibiotics has increased substantially over the past couple of decades. In 2009, more than 3 million kilograms (kg) of antibiotics were administered to human patients. In 2010 13 million kg were administered to animals [8]. As populations continue to increase, we can only expect usage to increase. Resistance cannot be slowed until we stop exposing the environment to massive amounts of antibiotics. Paul Marino from the ICU Book [9] recommends: "The first rule of antibiotics is try not to use them, and the second rule is try not to use too many of them."

If antibiotics are inappropriately used - such as to treat viral infections - not only are they ineffective, but they also become less effective on the bacteria they intend to treat [10]. Not taking antibiotics exactly as prescribed also leads to issues because the bacteria may not be completely wiped out. The surviving bacteria can become more resistant and also be transferred to other people. Finally, self-prescribed antibiotics such as when doctors erroneously prescribe medications for themselves, family members, or friends, is another example of antibiotic misuse [11].

The ongoing issue of antimicrobial resistance has been a significant public health issue throughout the world. The Chief Medical Officer of England, Professor Dame Sally Davies was recently quoted: "It is clear that we might not ever see global warming. The apocalyptic scenario is that when I need a new hip in 20 years I'll die from a routine infection because we've run out of antibiotics" [12]. In addition, the Centers for Disease Control and Prevention (CDC) calls antibiotic resistance "one of the world's most pressing public health problems." The World Economic Forum recently stated "arguably the greatest risk...to human health comes in the form of antibiotic resistant bacteria. We live in a bacterial world where will never be able to stay ahead of the mutation curve. A test of our resilience is how far behind the curve we allow ourselves to fall" [13].

1.1.2   Current efforts to control resistance

Reversing this antibiotic resistance trend is expected to be slow or non-existent [14]. To help combat the resistance issue, a number of government programs have been formed to monitor the evolving resistance and hopefully hasten its progress. The European Center for Disease Prevention and Control (ECDC) publishes an annual surveillance report on the latest developments in antimicrobial resistance. The National Antimicrobial Resistance Monitoring System (NARMS) is a U.S. national

health agency that tracks antibiotic resistance. Established in 1996, the primary goals of NARMS are to:

1. Monitor trends in antimicrobial resistance;

2. Produce timely reports regarding resistance;

3. Conduct antimicrobial research;

4. Assist the FDA in making decisions regarding safe and effect drugs for animals.

In addition to monitoring resistance, the problems of misuse and overuse of antibiotics have been addressed by the formation of the U.S. Interagency Task Force on Antimicrobial Resistance. This task force aims to actively address antimicrobial resistance, and is coordinated by numerous US agencies [15].

Unfortunately, the situation is grim. There are no simple solutions to the problem of antimicrobial resistance and we are already experiencing the negative consequences. According to the Center for Disease Control (CDC), in 2010 about 99,000 people died in American hospitals from bacterial infections they acquired in the hospital. Part of the cause of this alarmingly high number is that many people enter the hospital with weakened immune systems and are susceptible to infections. These patients may end up dying before a suitable antimicrobial can be found. To compound the issue, there hasn't been a new antibiotic developed since 1987 and there are currently none in the development pipelines of major pharmaceutical companies.

Public health researchers must work on a better understanding of antibiotic resistance and the appropriate choice of antibiotics for a given infection. **The goal of this thesis is to assist with the latter, to propose a more precise and accurate method to determine classification breakpoints.**

1.2 Susceptibility Testing

The goal of susceptibility testing is to predict the success or failure of antibiotic therapy. When a patient enters a hospital with an unknown infection, susceptibility testing is used to generate a list of potential drugs that should result in successful treatment. Testing is done in a laboratory or clinic by exposing the unknown bacteria to a variety of antimicrobials across a range of concentrations. For each drug/bug combination, a susceptibility test classifies the bacterial pathogen to the drug as susceptible (high probability of successful treatment), indeterminant (marginal probability of successful treatment), or resistant (low probability of successful treatment).

It is important to note here that susceptibility testing does not determine the best drug to use but rather serves as a guide for selecting the correct antibiotic. Clinical experience and the status of the patient are additional key factors in the final decision of antibiotic.

Given the list of drugs for which the pathogen is susceptible, a doctor can then select the appropriate one for the patient. If there are no drugs on this list, the doctor may select a drug in the indeterminant classification. This usually results in a drug being prescribed at a higher dose level than normal. If all results come back as resistant, the doctor would have to selectively choose a drug to administer at a very high dose level or seek alternate treatment options.

In a typical hospital setting, one of two common susceptibility tests is administered. The **drug dilution test**, also known as the minimum inhibitory concentration (MIC) test, directly estimates the concentration of the drug needed to kill the bacteria. The **diffusion test**, also known as the disk diffusion (DIA) test, indirectly estimates the concentration via the diameter of a clear zone around a drug disk. Hospitals prefer the DIA test because it uses less lab space, provides faster results, and costs less to implement [16].

These two tests are performed under standardized conditions so that the results are as reproducible as possible. But as with any test there is experimental variation. Fortunately these two tests are well monitored so we have a good idea of the inherent variation of each test. We will discuss the variation of each test when we detail each procedure.

1.2.1 Minimum Inhibitory Concentration (MIC) Test

The minimum inhibitory concentration (MIC) test determines the minimum concentration of a specific antibiotic that will inhibit growth of the isolated pathogen. The test uses two-fold dilutions of the drug (e.g., 1, 2, 4, 8 and 16 $\mu g/ml$) applied to wells in a 96-well plate (Figure 1.1). This standard plate allows for up to 12 antibiotics tested over a range of 8 dilutions or up to 8 antibiotics tested over 12 dilutions.

The pathogen is placed into the wells containing an assay broth (food for growth) and a particular concentrate of a drug. The plates are incubated overnight to allow for bacterial growth. The observed MIC is the lowest concentration with no visible bacterial growth (i.e., the well is not cloudy) after incubation. Few laboratories actually prepare their own plates and instead purchase them from one of several commercial suppliers.

Because the procedure considers two-fold dilutions, it is common to report results on the $\log_2$ scale and the smaller the test result the more likely the pathogen is susceptible to the drug. Due to the evaluation process, however, the actual MIC will be rounded up to the nearest two-fold concentration under consideration. For example, if the actual $\log_2(\text{MIC})$ is 0.6, the observed concentration will be reported as 1 because growth will occur in any well with a dilution below 0.6. These reported MICs are determined manually or by using an automatic reader. Classification of the bac-

teria to each drug is based on drug-specific breakpoints. Details of this classification procedure are described in Section 1.3.

A large array of quality control experiments (repeated tests across numerous labs) have revealed a common three-fold dilution range in reported results within a lab and a similar variance between labs [17] [18]. Depending on where these true MICs are relative to the classification breakpoints, these variabilities can be quite important. For example, it is not highly unlikely for a susceptible drug to be classified as resistant.

The primary advantages of the MIC test is having prepared trays and producing results in concentrations, which makes interpretation and breakpoint determination simple. There is also the advantage of potential automatic readings and report generation. The primary disadvantage of the microdilution method is that not all drugs are available in the standard commercial trays. In addition, the incubation process takes longer and the process requires increased lab space compared to the DIA method.
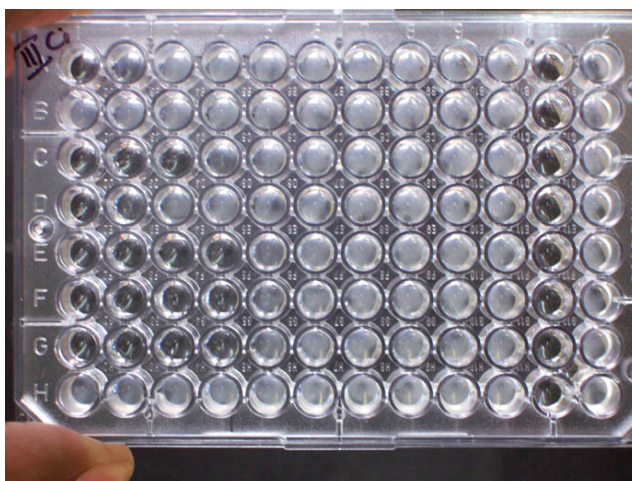
Figure 1.1. Example of a 96-well plate, where rows are drugs and columns are concentrations increasing from right to left. Notice that for some drugs the wells start off cloudy and eventually become clear. The MIC is the concentration where this transition occurs. [1]

---

[1]https://www.boundless.com/microbiology/antimicrobial-drugs/measuring-drug-susceptibility/minimal-inhibitory-concentration-mic/

1.2.2 Disk Diffusion Test

The disk diffusion (DIA) test is more popular in hospitals and clinics because it is simple, faster, and requires less lab space. For this test, a commercially-prepared disk containing a drug is placed on a Petri dish that is smeared with the pathogen (Figure 1.2). Over time the drug diffuses from the disk creating a gradient of concentrations. If the concentration is high enough, the pathogen will not grow. This results in a "clear zone" around the disk, which represents these concentrations. The test result is the diameter of the clear zone around the disk. The larger the diameter of the zone, the more likely the pathogen is susceptible to the drug.

Similar quality control experiments have shown a six-fold millimeter (mm) range both within and across labs. Similar to the MIC test, pathogens near the classification breakpoints are more likely to be misclassified compared to those further away.

Advantages of the DIA test include flexibility in the selection of disks for testing and that the test does not require any special equipment. The disadvantages are the lack of automation and the fact that there is no easy conversion between the clear zone diameter and concentration. It turns out the relationship between the two tests is very drug/bug specific, and, most important to this work, **there is no known functional form.** It is known, however, that the relationship between the two must be monotonically decreasing.

Figure 1.2. Example of the DIA test. The agar plate has several different drug disks. The clear zone around each of the disks is the test result.[2]

1.3 Determining Classification Breakpoints

For a particular drug/pathogen combination, breakpoints take the observed test result and classify the bacteria relative to the drug as susceptible (high probability the drug will be successful), indeterminant (medium probability the drug will be successful), or resistant (low probability the drug will be successful). Since the MIC test deals directly with concentrations of the drug, MIC breakpoints are based on the drug's pharmacokinetics and pharmacodynamics (i.e., how the drug moves through the human body and what concentrations of the drug the body can safely withstand). Agencies such as the Clinical and Laboratory Standards Institute (CLSI) or the Food and Drug Administration (FDA) set these breakpoints.

Establishing breakpoints for the DIA test is a much more difficult task. It is not possible to determine breakpoints based on the drug's pharmacokinetics and

---

[2]http://canadianhealthandcareonline.com/antibiotic-susceptibility-and-resistance-testing-determining-the-minimum-inhibitory-concentration-part-1.html

pharmacodynamics as there is no easy way to convert the concentrations to diameters and vice versa. While it is clear that there is a 1-1 inverse relationship between the true MIC and associated true DIA, the functional form is unknown and drug/bug specific. **Because of this, DIA breakpoints are estimated using the MIC breakpoints and joint MIC and DIA test results from a large number of bacterial strains.** These are the data sets we address in this thesis.

1.3.1 Experimental Susceptibility Testing Data

Current CLSI guidelines stipulate that the pool of strains should be at least 500 isolates [19]. Figure 1.3 displays the type of experimental data that are used in this determination procedure. The MIC test is on the x-axis and the DIA test is on the y-axis. Each strain tested contributes one point on this scatterplot. Given the rounding inherent in each test, multiple strains can provide the same results. Because of this, the points on a scatterplot are represented by the number of isolates observed at each combination. The vertical dashed lines represent the established MIC breakpoints (shifted inwards 0.5 units for easy reading). In this example, the MIC test susceptible range is for an observed $MIC \leq -1$, and the resistant range is for an observed $MIC \geq 1$.
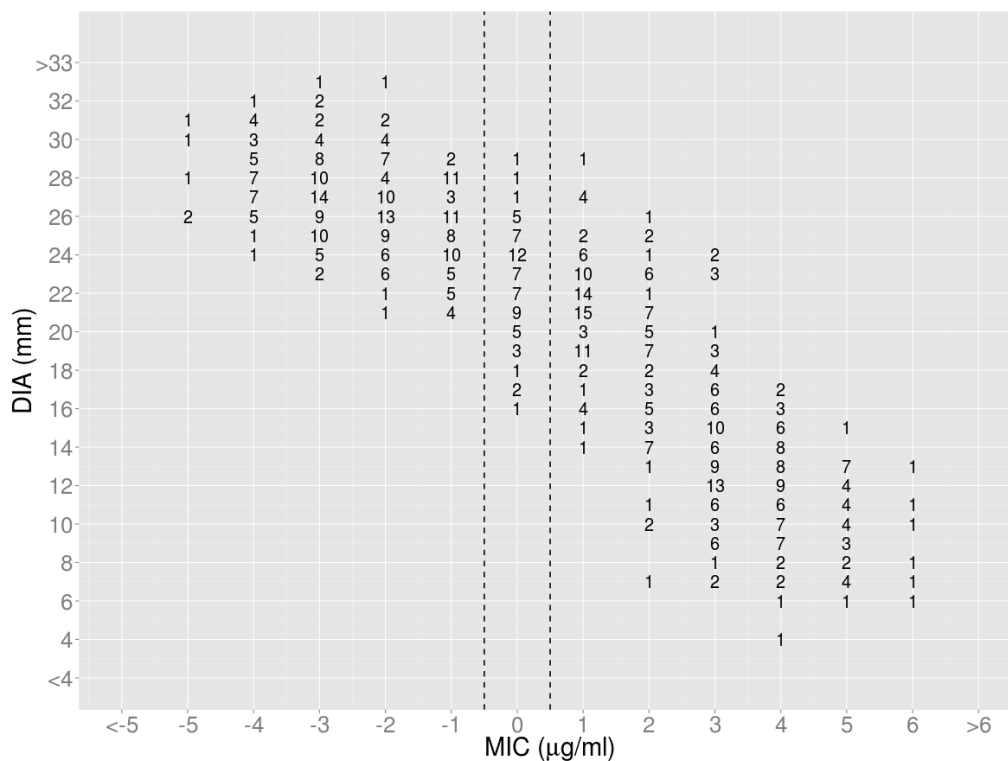
Figure 1.3. Example scatterplot of susceptibility testing data used in the determination of DIA breakpoints. The numbers in the scatterplot represent the number of strains observed at that combination.

Underlying these scatterplots is the fact that there is a 1-1 inverse relationship between the true MIC and DIA values. Obscuring this relationship is the inherent measurement error present in both tests and the fact that the MIC test rounds up to the nearest integer while the DIA test rounds to the nearest integer. Although, one might argue the relationship looks relatively linear in Figure 1.3, there is no known functional form to describe this relationship.

1.4   Prior DIA Breakpoint Determination Methods

The various breakpoint determination procedures attempt to address these scatterplot characteristics in a variety of ways.

1.4.1 Error-Rate Bounded

The long-standing DIA breakpoint determination procedure is the error-rate bounded (ERB) method [20]. This method is fast, intuitive, and assumes no parametric relationship between the MIC and DIA test results. The focus of the procedure is on the observed discrepancies between test classifications. For a given set of DIA breakpoints, the MIC and DIA classifications will either match or not match. When the two classifications do not match there is a discrepancy. The ERB method is a procedure that searches for the DIA breakpoints which minimize these observed discrepancies.

When Metzler and DeHaan proposed the ERB method in 1974, it was typical for there to be only one MIC breakpoint and thus no indeterminant zone. Thus, for a given set of DIA breakpoints, the two test classifications either were identical (both susceptible or resistant) or discrepant. These discrepancies were classified as false resistant (MIC = susceptible, DIA = resistant), false susceptible (MIC = resistant, DIA = susceptible), or minor (DIA = indeterminant).

In order to determine the best DIA breakpoints, a table was produced providing information across many sets of DIA breakpoints. This information included the number of identical classifications as well as the number of false resistant, false susceptible, and minor discrepancies.

In addition, a sampling experiment was performed where half of the observations (training set) was used to determine appropriate DIA breakpoints and the other half (test set) was used to evaluate the selected DIA breakpoint performance. It was thought the test sample might be more representative of breakpoint performance in future experiments as the selected breakpoints are not based on them. Agencies used this table information to choose the best breakpoints.

Brunden et al. modified this procedure by considering two MIC breakpoints [21]. Given this additional indeterminant region for the MIC test, more discrepancies can

occur. They stated a very major error (VM) occurs when the MIC test returns a resistant result and the DIA test returns a susceptible result; a major error (M) occurs when the MIC test returns a susceptible result and the DIA test returns a resistant result; and a minor error (m) occurs when exactly one of the tests returns an indeterminant result. This naming convention was based on the thought that the MIC test was more accurate and thus having the MIC test report resistant and the DIA test report susceptible is more severe then the opposite direction.

To better understand the possible discrepancies given test data, it is helpful to think of the data grid broken into a 3x3 set of regions (Figure 1.4). Isolates in the top-left, center, and bottom right regions are classified the same by both tests. Isolates in the middle box of each axis are minor errors. Isolates in the bottom left box are major errors and isolates in the top right box are very major errors.

To determine DIA breakpoints, Brunden et al. place upper limits on the percentage of allowable very major, major, and minor errors (current percentage limits are listed in Table 1.1). Similar to Meltzler and DeHaan, Brunden et al. produced a table of summary statistics for many sets of DIA breakpoints. Brunden et al. also introduced an index score as an additional statistic. Their proposed index score is

$$Index = \frac{D_U - D_L}{(\%E)}$$

where D_U is the upper DIA breakpoint, D_L is the lower DIA breakpoint, and $\%E$ is the total percent of very major, major, and minor errors. A grid search over all possible DIA breakpoints was performed and the set of DIA breakpoints that maximized the index score was deemed optimal.
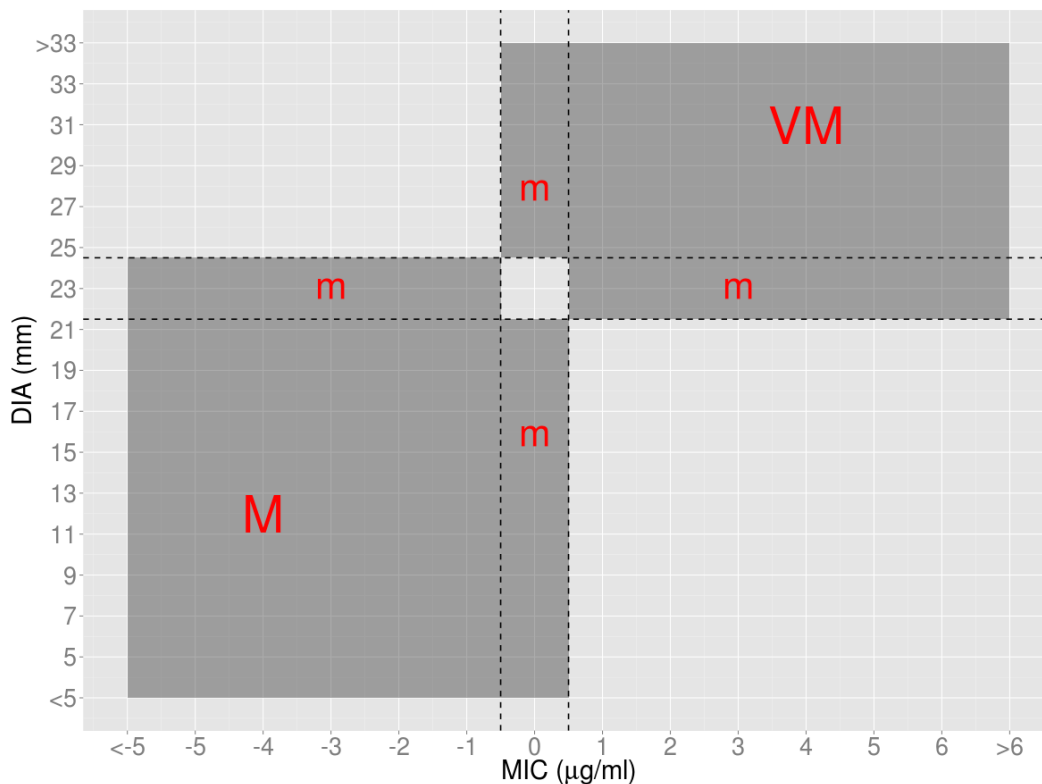
Figure 1.4. Discrepancy regions given a set of MIC and DIA breakpoints. VM stands for a very major discrepancy, while M and m stand for major and minor discrepancies, respectively.

Two other index scores were also considered: $Index = \%E$ and $Index = \alpha(\%M) + \beta(\%VM)$ with $\alpha + \beta = 1$. For the latter two index scores, breakpoints were selected that minimized them. No recommendations were given on how to select α or β, so the choice was left to the practitioner.

Our version of the ERB method is an extension of Brunden et al.'s method utilized by CLSI. They have set percentage limits for discrepancies within and outside one log dilution of the MIC indeterminant range (Table 1.1).

| | Discrepancy Percentages | | |
MIC Range	Very Major	Major	Minor
$\geq I_{High}+1$	$< 2\%$	NA	$< 5\%$
I_{High} to I_{Low}	$< 10\%$	$< 10\%$	$< 40\%$
$\leq I_{Low}-1$	NA	$< 2\%$	$< 5\%$

Table 1.1 CLSI recommended discrepancy percentages for the error-rate bounded method.

Rather than produce a table of results, we consider an index which converts the percents of Table 1.1 into weights that represent their importance. Each discrepancy percentage ($D\%$) is converted to a weight (w) by: $w = \frac{\max(D\%)}{D\%}$, where $\max(D\%)$ is the largest discrepancy percentage. For the percentages given in Table 1.1, the weights are: $VM_1 = 40/10 = 4$, $M_1 = 40/10 = 4$, $m_1 = 10/10 = 1$, $VM_2 = 40/2 = 20$, $M_2 = 40/2 = 20$, and $m_2 = 40/5 = 8$. The subscript 1 indicates the isolate is within 1 log dilution of the MIC indeterminant range and the subscript 2 indicates the isolate is outside 1 log dilution of the indeterminant range. Thus a very major error that occurs outside one dilution of the MIC indeterminant range is considered to be 5 times more serious than one that occurs within one dilution of the indeterminant range. Note the percentages in Table 1.1 treat VM and M errors equally.

Our ERB index is simply an accumulation of these weights and can be expressed as:

$$index = (w_{VM_1} \times \#VM_1/N_1) + (w_{M_1} \times \#M_1/N_1) + (w_{m_1} \times \#m_1/N_1) +$$
$$(w_{VM_2} \times \#VM_2/N_2) + (w_{M_2} \times \#M_2/N_2) + (w_{m_2} \times \#m_2/N_2)$$

where w_* represents the weight associated with a discrepancy, N_* represents the number of observations within one of of the indeterminant range (N_1) or outside one of the indeterminant range (N_2), and $\#^*$ represents the counts for a particular

discrepancy. Applying our ERB method to the data in Figure 1.3 results in the choice of 20 mm and 25 mm for DIA breakpoints. These breakpoints are graphically displayed in Figure 1.5.
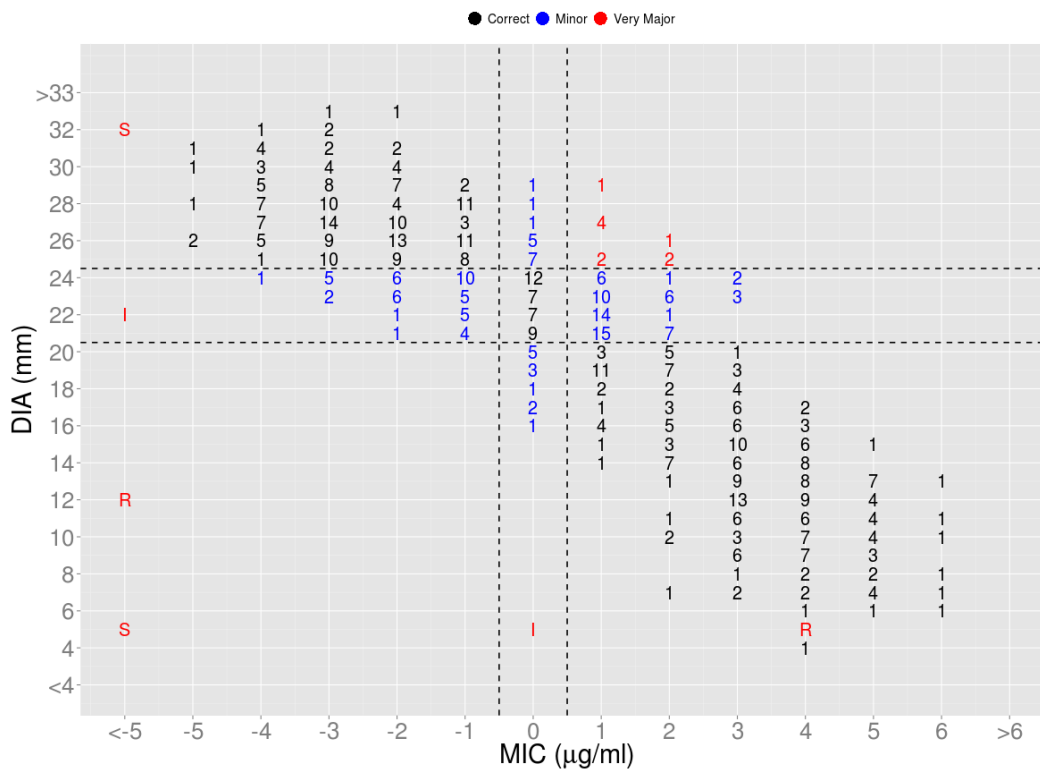
Figure 1.5. Summary plot of the ERB applied to the scatterplot in Figure 1.3. The DIA breakpoints based on our index score is 20 mm and 25 mm.

The color of the points in Figure 1.5 refers to the type of classification. Black points are classified the same by both tests. Blue points represent minor errors and very major errors are in red. For this example and choice of breakpoints, there were no major errors.

The discrepancy index for the selected DIA breakpoints is computed by:

$$1.6129 = (4 \times 7/196) + (4 \times 0/196) + (1 \times 96/196)+$$

$$(20 \times 3/404) + (20 \times 0/404) + (8 \times 42/404)$$

A summary of the data in Figure 1.5 for a variety of DIA breakpoints is presented in Table 1.2. The optimal DIA breakpoints based on the index score are presented in the first row. The subscripts 1 and 2 indicates distance of the MIC from the indeterminant range.

Breakpoints	Agree 1	VM_1	M_1	m_1	Agree 2	VM_2	M_2	m_2	Index
20, 25	**93**	**7**	**0**	**96**	**359**	**3**	**0**	**42**	**1.6129**
19, 24	93	13	0	90	365	6	0	33	1.6750
19, 25	95	7	0	94	353	3	0	48	1.7215
20, 26	92	5	0	99	339	1	0	64	1.9240
21, 26	98	5	4	89	346	1	1	56	1.8457
19, 26	94	5	0	97	333	1	0	70	2.0326

Table 1.2 ERB results for simulated experimental data for a variety of DIA breakpoints. The selected breakpoints and resulting statistics are bolded in the first row.

1.4.2 Concerns with the ERB Approach

A discrepancy is most likely to occur when the pathogen is near the MIC indeterminant range. When there are few pathogens in this region many sets of DIA breakpoints may satisfy the discrepancy percent limits. However when there are many pathogens in the indeterminant region there may not be any breakpoints that satisfy the percentage limits. In both situations it is questionable how to choose appropriate DIA breakpoints. The use of an index avoids this problem.

Because of increased resistance, the situation of having many isolates near the indeterminant range has become more frequent. Before the onslaught of antimicrobial

resistance the distribution was mostly bimodal with only a few pathogens near the indeterminant region and therefore easy to select DIA breakpoints.

Around 10 years ago, agency researchers prompted investigations into the ERB approach due to the increase in the percentage of pathogens around the MIC indeterminant region [22]. It was at this time that agencies modified the ERB method to include a second set of discrepancy percentages to account for more isolates near the indeterminant region. While this modification was an improvement, it also added further complexity to the ERB method and agencies are again revisiting how best to do this breakpoint determination.

In 2000, Craig revealed additional drawbacks with the ERB method [23]. Craig showed that repeat runs of the ERB for the same drug/bug combination can result in very different DIA breakpoints (low precision). This is because the ERB method uses only observed results and does not adequately take into account the test variabilities. In addition, each test rounds differently so the resulting DIA breakpoints are typically biased. While a proper choice of weights for an index can take this rounding into account and may reduce the bias, it does not address the measurement error issue.

1.4.3   Model-Based Approaches

When there were few isolates near the indeterminant range these issues may not have an impact on the choice of breakpoints. However with the MIC distribution having more indeterminant strains, accurately estimating DIA breakpoints has become more of a statistical problem and the test characteristics and experimental variability must be taken into account. By modeling these characteristics, one can determine the probability of a discrepancy and breakpoint determination can be made using the underlying true relationship rather than observed discrepancies.

Linear Regression Models

Before the ERB method, linear regression models were used to determine DIA breakpoints. If the correlation between the tests was strong enough, clinicians chose the DIA breakpoints where the linear regression line crossed the MIC breakpoints.

There are many problems with using the linear regression model for breakpoint selection, as noted by Metzler and DeHaan. The linear regression model does not take into account the test characteristics nor the measurement errors. Also, much of the data produced today does not follow a linear relationship and therefore using the linear regression model to select DIA breakpoints is not recommended. After the introduction of the ERB method, linear regression was sometimes an additional selection tool [24] but it has since been phased out.

Logistic Model

In 2000, Craig proposed a hierarchical model to account for the measurement error and rounding of each test [23] [25]. Instead of minimizing the observed discrepancies, Craig estimates the DIA breakpoints by finding the set that minimizes a weighted loss function. The loss function is the accumulated sum of the squared difference in the probability of correct classification when the DIA test performs worse than the MIC test. The weights are based on the underlying MIC density. This approach eliminates the need for setting upper limits on discrepancy percentages.

Craig's hierarchical model has three components:

1. The test procedures (i.e., rounding) and experimental variability
2. The drug-specific relationship between the true MICs and DIAs
3. The underlying distribution of pathogens (or MICs)

The first component links the observed MIC/DIA pair with an underlying true MIC value. The second and third components describe the relationship between the true MIC and true DIA. A graphical representation of the model is shown in Figure 1.6. The underlying MIC distribution is the black density shown along the bottom. The true relationship between the MIC and DIA is the curved blue line and the black dots represent experimental values for a given isolate. To generate an observed test pair, a true MIC is randomly drawn from the density and the curve is used to determine the corresponding true DIA. Measurement error is then added to each test and then rounded to get the test results.
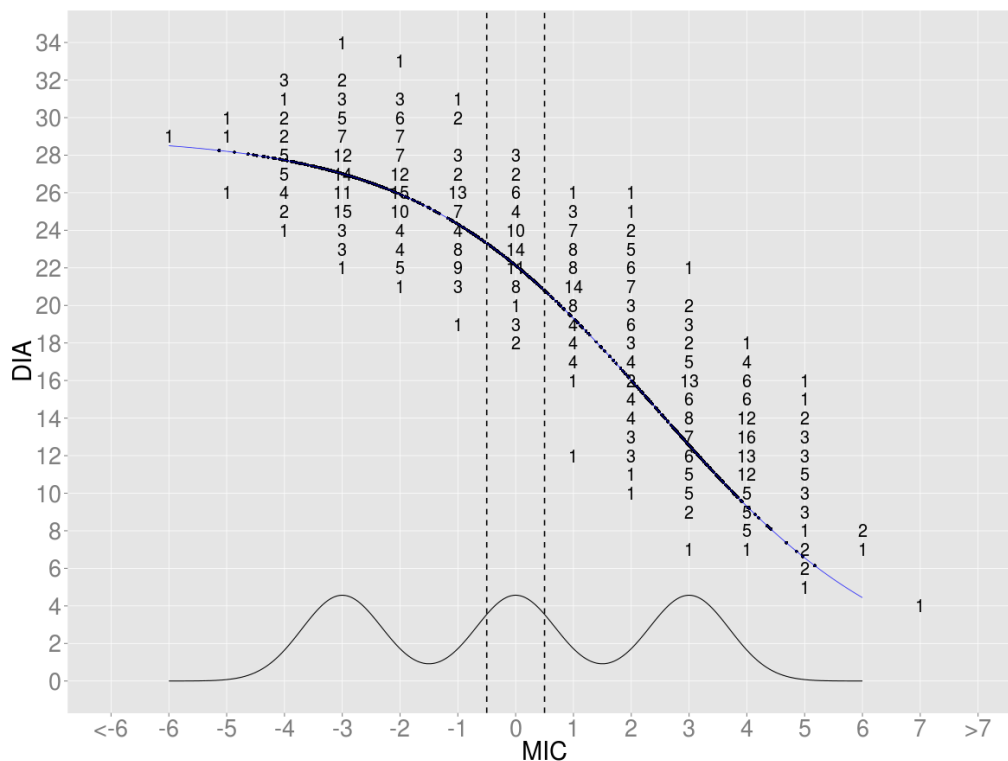
Figure 1.6. Example of Craig's hierarchical model showing the underlying distribution of MICs, the true 1-1 relationship between MIC and DIA, and the observed test results.

The key to Craig's model-based approach is to consider a true MIC/DIA relationship for the population of pathogens, modeled as $d_i = g(m_i)$, where d_i and m_i are the true DIA and MIC values for pathogen i respectively and $g(.)$ is a monotonically decreasing function. This relationship, which was assumed logistic, combined with measurement error models for each test, link the pairs of observed test results.

Two-stage Nonparametric Approach (2008)

While Craig's approach works very well when a logistic curve adequately approximates the true underlying relationship, it also can produce poor and biased results when the function does not fit the true relationship well. In 2008, Qi relaxed Craig's logistic relationship assumption and considered a flexible nonparametric relationship between the true test results [26]. Qi used a two-stage frequentist method. First, he modeled the underlying MIC distribution using nonparametric density estimation based on M-splines. Second, he fit the MIC/DIA relationship using I-splines. For both parts, knot selection was done by placing a knot at each interval MIC value, then doing a backwards selection procedure to remove knots based on AIC. Measurement error and rounding were included in the likelihood function and parameter estimates were determined by maximum likelihood estimation. DIA breakpoint confidence intervals are obtained by nonparametric bootstrap.

While this approach is reasonable, there are concerns that information is lost via this two-stage approach. The MIC density provides information for where the true DIA falls along the fitted curve. By not simultaneously estimating the density and underlying relationship you lose any information the observed DIA values provide in the density estimation.

Simulation studies showed Qi's model was comparable to Craig's logistic model when the true model follows a logistic curve. When the true model deviates from a

logistic curve Qi's model performed substantially better due to the robustness of the nonparametric fit.

1.4.4   Proposed Method for Estimating DIA Breakpoints

We propose a new method that combines the benefits of the previous two model-based approaches. First we relax Craig's parametric assumption by using I-splines to model the relationship between the DIA and MIC, similar to Qi. However, unlike Qi, we do not break the estimation procedure into two steps. We instead take a Bayesian approach to estimate the necessary parameters for the underlying distribution and relationship together. This should provide increased accuracy in estimating DIA breakpoints. This Bayesian approach also provides a posterior distribution of possible DIA breakpoints. Using this posterior distribution, clinicians can combine our recommended breakpoints with clinical expertise to select the appropriate DIA breakpoints.

Neither Qi nor Craig made software available for clinicians to easily implement their approaches. Because of this, clinicians are still using the ERB method. We provide an online application for clinicians to use our method in practice. This software is currently being tested by clinicians from CLSI and the FDA.

The rest of the thesis is organized as follows: Chapter 2 details our approach to the problem, Chapter 3 reports simulation and real data results, and Chapter 4 finishes with details on software implementation, summary, and future work.

2. METHODS

In this chapter we develop our approach to determine DIA breakpoints from a scatterplot of observed (MIC, DIA) pairs. Our model and estimation techniques are presented in Section 2.3. The initial sections provide an overview of key subjects that relate to our model, specifically measurement error (Section 2.1) and nonparametric regression (Section 2.2). Measurement error is a non-ignorable feature of these data and nonparametric monotone regression is used to model the true underlying MIC/DIA relationship.

2.1 Overview of Measurement Error

Measurement error occurs when we don't precisely observe one or more of the model predictors. It occurs in nearly every research discipline and for a variety of reasons. One common reason is instrument/equipment error, where the instrument (e.g. device, survey, or diagnostic test) used to obtain the measurement is not perfect.

The effect of ignoring measurement error on model selection and prediction depends on the magnitude of this error relative to the other variation in the data. If the measurement error is relatively small, the effect is minimal. However if the measurement error is relatively large, the effect can be quite severe.

Before discussing how we handle measurement error in our setting, we review the measurement error literature to provide context for our approach. We consider the following variables in this discussion:

- Y: the observed responses

- X: the true covariates/predictors

- W: observed proxies for X (due to measurement error)

- U: random measurement error with $E(U) = 0$

- ϵ: random error with $E(\epsilon) = 0$ and independent of U

We focus here on the additive measurement error model with constant variance. In terms of our variables, we assume $Y = g(X) + \epsilon$ and $W = X + U$ where $Var(U) = \sigma_u^2$. The measurement error is homoscedastic since the variance of $W|x$ is constant and nondifferential since the measurement error does not depend on y (distribution of $W|x, y$ is the same as the distribution of $W|x$).

2.1.1 Classical and Berkson Measurement Error Models

There are two common types of additive measurement error models, the classical measurement error model and the Berkson measurement error model [27]. The decision of which measurement error model is appropriate depends on whether Y is observed as a function of X or W [28].

The classical measurement error model is as follows

$$Y = g(X) + \epsilon \qquad W = X + U$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2) \qquad U \sim \mathcal{N}(0, \sigma_U^2) \tag{2.1}$$

and the goal is to estimate the unknown function $g(.)$ using the observed Y and W. The true predictor X is not observed.

The Berkson measurement error model is similar but Y is a function of W (the unobserved proxy for X):

$$Y = g(W) + \epsilon \qquad W = X + U$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2) \qquad U \sim \mathcal{N}(0, \sigma_U^2) \qquad (2.2)$$

and the goal is to estimate the unknown function $g(.)$ using the observed Y and target values X.

An example of the classical measurement error is measuring blood chemistry. Laboratories calibrate their instruments and values are returned based on the resulting calibration curve. These values will still contain some error unless there is a deterministic relationship between the true and measured values (rarely the case). On the other hand, in some designed experiments there is a target level of a factor an experimenter wishes to assign to the experimental unit but the actual level may differ from the target due to uncontrollable noise. These are Berkson measurement errors. This could happen, for example, with experiments involving temperature and pressure in industrial experiments, fertilizer or watering levels in agricultural settings, and speed on a treadmill. In each case the response is based on the observed level, which is a randomly perturbed valued of the target level.

2.1.2 When Measurement Error is Ignored

In this subsection, we'll consider both linear and nonlinear settings for $g(.)$. In the linear regression setting, Berkson measurement errors model do not bias the

parameter estimates. In other words, the regression model is robust to Berkson errors. Suppose we want to estimate the parameters of the model

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{2.3}$$

but observe (Y^*, X) pairs where $Y^* = \beta_0 + \beta_1 W + \epsilon$. We can rewrite Y^* as

$$Y^* = \beta_0 + \beta_1(X + U) + \epsilon$$
$$= \beta_0 + \beta_1 X + (\beta_1 U + \epsilon) \tag{2.4}$$

Since $E(Y^*)$ is equal to $E(Y)$, we obtain unbiased estimates for β_0 and β_1. There is, however, an increase in the residual variance due to these errors, which reduces the power and alters some summary statistics, such as the coefficient of determination.

When you have classical measurement errors, the effect on a linear regression model is more severe. Not accounting for this error results in:

1. Bias in β_0 and β_1

2. Loss of power for detecting an association between variables

Figures 2.1, 2.2, and 2.3 demonstrate the bias in the intercept and slope parameter estimates for varying degrees of measurement error. In each figure, the blue points represent values observed with measurement error and the red points represent their true values. The two lines in each plot correspond to the least-square fits of Y vs X and Y vs W.
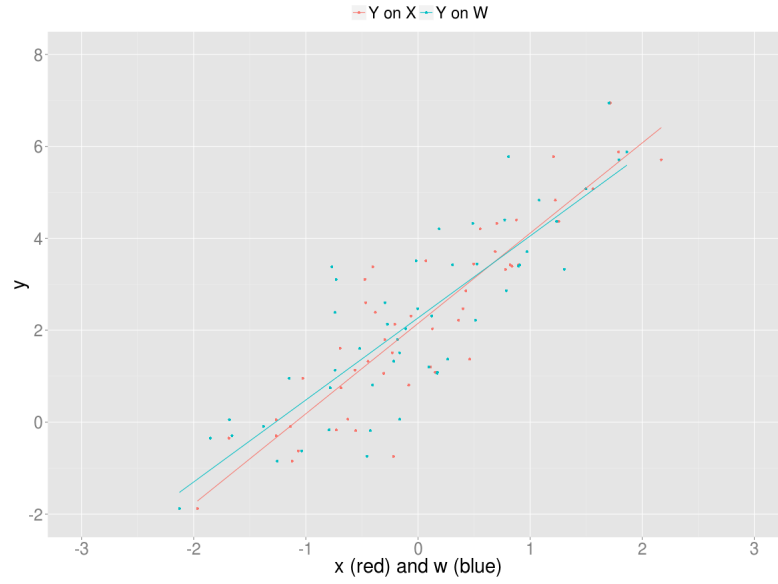
Figure 2.1. Measurement error shrinks the slope towards zero. For the two sets, 50 data points were generated by: $X \sim \mathcal{N}(0,1)$, $\epsilon \sim \mathcal{N}(0,1/3)$, $U \sim \mathcal{N}(0,1/4)$, $Y = 2 + 2X + \epsilon$, $W = X + U$
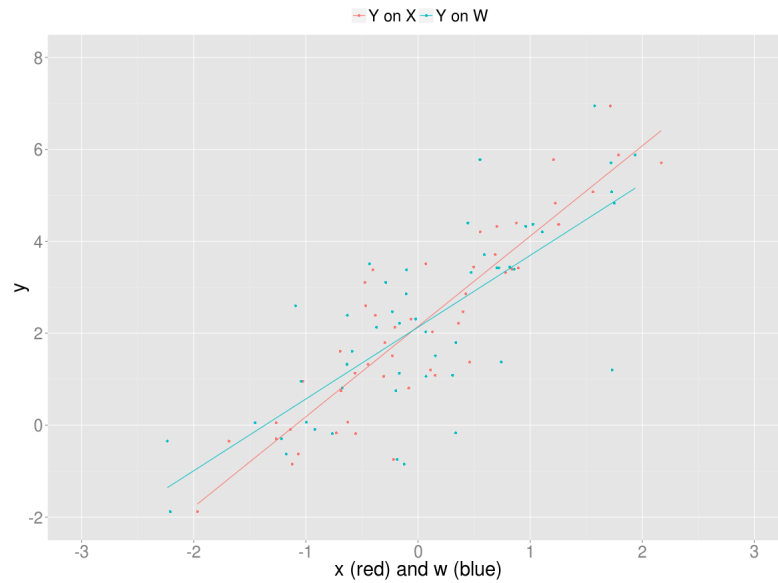
Figure 2.2. Measurement error shrinks the slope towards zero. For the two sets, 50 data points were generated by: $X \sim \mathcal{N}(0,1)$, $\epsilon \sim \mathcal{N}(0,1/3)$, $U \sim \mathcal{N}(0,1/2)$, $Y = 2 + 2X + \epsilon$, $W = X + U$
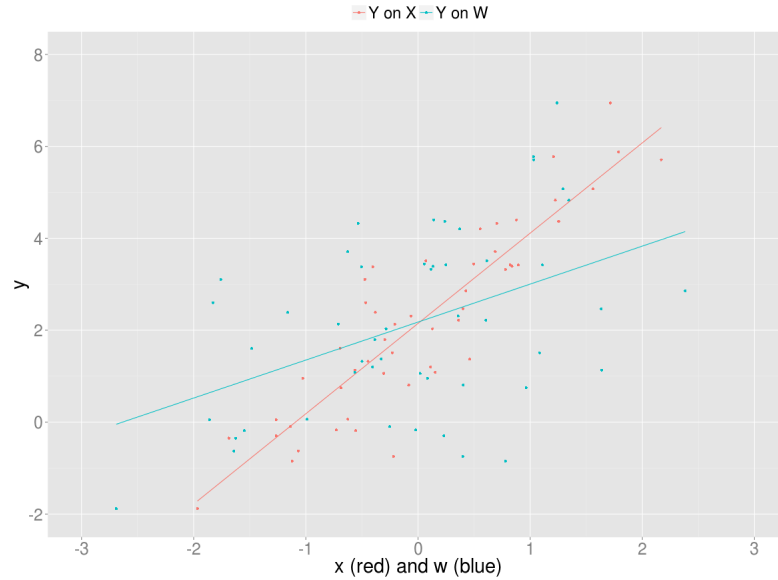
Figure 2.3. Measurement error shrinks the slope towards zero. For the two sets, 50 data points were generated by: $X \sim \mathcal{N}(0,1)$, $\epsilon \sim \mathcal{N}(0,1/3)$, $U \sim \mathcal{N}(0,1)$, $Y = 2 + 2X + \epsilon$, $W = X + U$

Measurement error increases across the three figures. When the measurement error is relatively small (Figure 2.1) the slope is only slightly shrunk towards 0, but as the measurement error increases (Figures 2.2 and 2.3), the difference between the two lines becomes more apparent.

Unlike the linear model, both classical and Berkson measurement error can mask data features when there is a nonlinear relationship. Figures 2.4, 2.5, and 2.6 demonstrate this effect in a classical error setting using a sinusoidal relationship under varying degrees of measurement error. The trend was estimated using nonparametric loess. The exact (Y, X) data (red points) display an underlying sigmodial relationship. The (Y, W) data (blue points) do not display this relationship, especially when the degree of measurement error is at its largest.
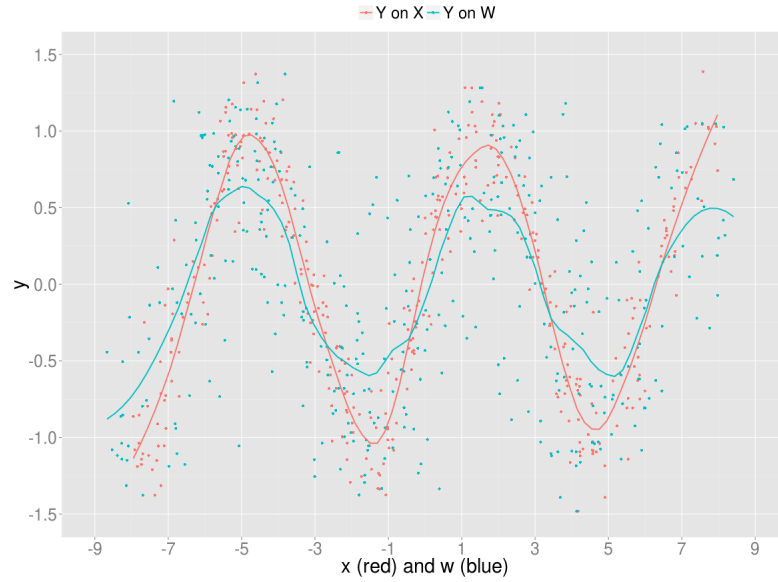
Figure 2.4. Measurement error can mask features of the data. For the two sets, 400 data points were generated by: $X \sim \mathcal{U}(-8, 8)$, $\epsilon \sim \mathcal{N}(0, 1/4)$, $U \sim \mathcal{N}(0, 1)$, $Y = \sin(X) + \epsilon$, $W = X + U$, loess smoother fit of Y vs X and Y vs W
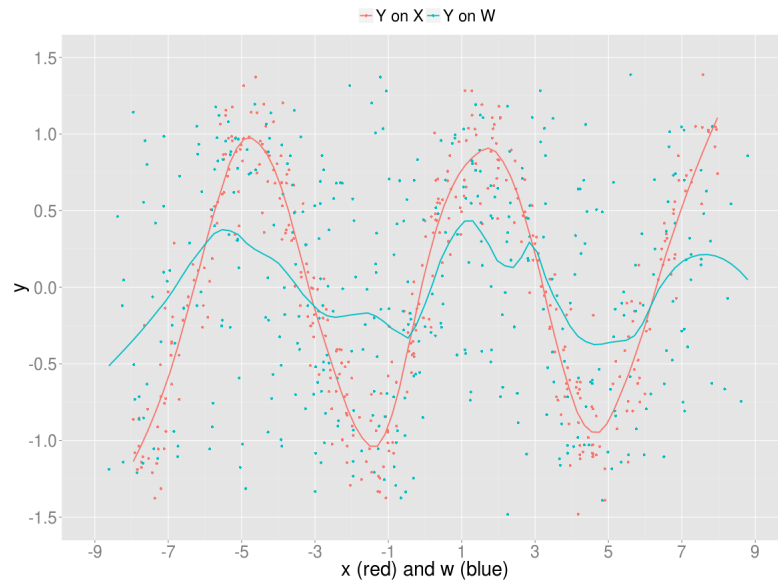
Figure 2.5. Measurement error can mask features of the data. For the two sets, 400 data points were generated by: $X \sim \mathcal{U}(-8, 8)$, $\epsilon \sim \mathcal{N}(0, 1/4)$, $U \sim \mathcal{N}(0, 1.5)$, $Y = \sin(X) + \epsilon$, $W = X + U$, loess smoother fit of Y vs X and Y vs W
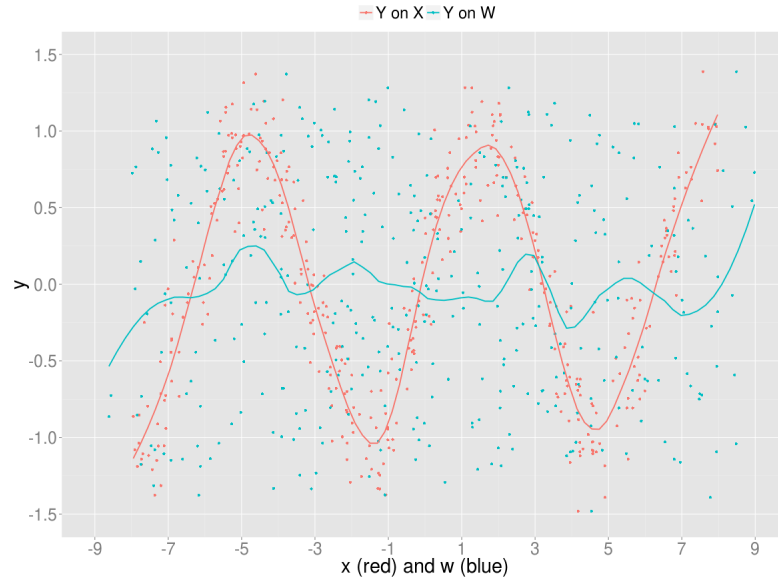
Figure 2.6. Measurement error can mask features of the data. For the two sets, 400 data points were generated by: $X \sim \mathcal{U}(-8, 8)$, $\epsilon \sim \mathcal{N}(0, 1/4)$, $U \sim \mathcal{N}(0, 2)$, $Y = \sin(X) + \epsilon$, $W = X + U$, loess smoother fit of Y vs X and Y vs W

2.1.3 Correcting for Measurement Error

There are two common estimation techniques to account for measurement error, *functional* and *structural* models [29]. In functional modeling the $\boldsymbol{X}$s may be either fixed or random, but in the latter case no, or minimal, assumptions are made about the distribution of the $\boldsymbol{X}$s. In structural modeling, parametric or nonparametric, models describe the distribution of the random $\boldsymbol{X}$s. Functional modeling is useful because it leads to estimates that are robust to the misspecification of $\boldsymbol{X}$. Structural models are more widely applicable but there is a concern with the robustness of inference to the assumptions made about the the distribution of $\boldsymbol{X}$.

2.1.4   Functional Model Approaches

Examples of functional model approaches are:

1. Regression calibration [30]

2. Simulation extrapolation (SIMEX) [31]

3. Instrumental variables [32]

In simple linear regression, we wish to estimate the parameters of the model $Y = \beta_0 + \beta_1 X + \epsilon$. In the classical measurement error setting, the least squares estimate of β_1 (Y vs W) is:

$$\hat{\beta}_1 = \left(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \right) \beta_1^* = \lambda \beta_1^* \tag{2.5}$$

where β_1^* would be the slope estimate if measurement error was not present (Y vs X). The parameter λ is known as the reliability ratio and represents the percent of noise in the data not due to measurement error. Notice that since $\lambda < 1$ the slope of the regression line is shrunk towards 0. One quick way to estimate β_1^* is to multiply the resulting estimate, $\hat{\beta}_1$, by the inverse of the reliability ratio (provided this estimate is known).

Regression calibration is a more general approach to correct for the bias in the parameter estimates. The idea is to replace $\boldsymbol{X}$ with $E(\boldsymbol{X}|\boldsymbol{W})$ and regress Y on $E(\boldsymbol{X}|\boldsymbol{W})$. This approach also requires the measurement error variance to be known (or estimated from data). For example, in simple linear regression, one can show

$$E(\boldsymbol{X}|\boldsymbol{W}) = \bar{W} + \lambda(W - \bar{W}) \tag{2.6}$$

where λ is the reliability ratio (2.5). Regressing Y on $E(\boldsymbol{X}|\boldsymbol{W})$ in 2.6 results in unbiased parameter estimates.

Regression calibration can be extended to multiple and nonlinear regression assuming one can work out the conditional expectation of $\boldsymbol{X}$ given $\boldsymbol{W}$. However, in highly nonlinear models, regression calibration can be quite poor.

The simulation and extrapolation method (SIMEX) is another useful tool for correcting estimates in the presence of measurement error. The measurement error variance is assumed known or estimated from data. SIMEX is based on the fact that the effect of measurement error can be determined experimentally via simulation. The basic idea is to estimate parameters for increasing levels of measurement error and extrapolate back to the case of no measurement error. Let $\beta_j(\lambda)$ be the expected value of β_j if the total measurement error variance is $(1 + \lambda)\sigma_u^2$ (note that λ controls the degree measurement error). For given λ, SIMEX estimates $\beta_j(\lambda)$ by adding more error to W, estimating the model parameters, and repeating. λ is varied from 0 to some upper bound. The SIMEX estimate is obtained by extrapolating back to the case with no measurement error, $\lambda = -1$. The number of simulations run for each λ is usually around 100. The drawback of SIMEX is the computation required in sampling and estimating. There is also the issue of selecting the number of simulations and values for λ. SIMEX can be used in nonlinear problems, however computing may be difficult due to the delicate optimization procedure that must be performed many times.

Both regression calibration and SIMEX rely on knowing, or estimating, the measurement error variance. In addition, the SIMEX method assumes the measurement error to be normally distributed, while regression calibration requires knowledge of the expected value of the measurement error. Without this knowledge accurate parameter estimation may still be possible if the data contain an instrumental variable in addition to $\boldsymbol{W}$. An instrument variable ($\boldsymbol{T}$) must possess the following three properties:

1. T is not independent of X

2. T is uncorrelated with U

3. T must be correlated with $Y - E(Y|X)$

One possible instrumental variable is a second measurement of X obtained by an independent sampling method. For example, a repeat measurement of blood may be considered an instrumental variables. By combining X and T a better estimate of Y can be obtained. In a regression setting this is typically done by using the information in T during parameterization. For example, in simple linear regression the parameter estimates are:

$$\hat{\beta}_1 = \frac{\hat{\sigma}_{ty}}{\hat{\sigma}_{tw}}$$
$$\hat{\beta}_0 = \bar{T} - \hat{\beta}_1 \bar{W} \tag{2.7}$$

where $\hat{\sigma}_{ty}$ and $\hat{\sigma}_{tw}$ are the sample covariances for σ_{ty} and σ_{tw}.

2.1.5 Structural Models

In structural models the unknown Xs are treated as latent or random variables, requiring distributions for the unknown Xs. While often more difficult to implement than functional models, they are more flexible and have a wide variety of applications. Their use has become more common as computing power has increased. In addition, structural models will generally result in increased efficiency (smaller standard errors). The trade-off is the concern of robustness if the distribution placed on X is misspecified. Structural models primarily involve likelihood and Bayesian methods.

The approach is similar for both likelihood and Bayesian structural models. For the likelihood method:

1. Select the likelihood model as if $\boldsymbol{X}$ were observed for every component in the data. That is, set up the likelihood function, $f_{y|x}(y|x, \beta)$, where the interest lies in β.

2. Select the type of measurement error model, Berkson or classical. If classical measurement error select the model for the unobserved $\boldsymbol{X}$s. In the Berkson case a model for the unobserved $\boldsymbol{X}$s may or may not be necessary. This is typically the hardest step.

3. Form the likelihood function from the previous two steps.

4. Maximize the likelihood function. Since $\boldsymbol{X}$ is latent, this step can be difficult and/or time consuming, because one must integrate out the possibly high dimensional latent variables.

The Bayesian method is similar with the addition of selecting priors and computing the posterior distribution. Our approach to DIA breakpoint determination will involve using a Bayesian structural model. Given the abundance of prior data, and knowledge of the test characteristics, we can accurately formulate a parametric relationship between the observed test results and their corresponding true values.

2.1.6 Measurement Error in our Setting

With DIA breakpoint determination, one observes pairs of test results for the MIC and DIA tests. These test results are not perfect and therefore contain measurement error that we need to account for when estimating the true underlying relationship. This fits the classical measurement error model framework except for one added twist, both the observed Y and W are rounded.

For pathogen i, let m_i and $g(m_i)$ denote the true MIC and DIA. The joint distribution of observed MIC ($\boldsymbol{x}$) and observed DIA ($\boldsymbol{y}$) can then be expressed as:

$$x_i = \lceil m_i + \epsilon \rceil \quad y_i = \lfloor g(m_i) + \delta \rceil$$

$$\epsilon \sim N(0, \sigma_m^2) \quad \delta \sim N(0, \sigma_d^2) \tag{2.8}$$

where σ_m and σ_d represent the measurement variability in the MIC and DIA test. We currently assume the experimental variabilities are known given the abundance of quality control data [17] [18], but this restriction can be relaxed (Craig 2000 [25]).

2.2 Nonparametric Regression

Given the observed MIC and DIA results, we want to draw inference on the underlying MIC/DIA relationship, $g(m_i)$. Biologically, this function must be monotonically decreasing. This is due to the fact that as the concentration of the drug increases the pathogen must become more susceptible, and, for the DIA test, the drug concentration decreases as the distance from the disk increases. Because of this monotonically decreasing restriction, and the fact the function can take any form, we consider estimating this relationship using monotonic nonparametric regression.

Nonparametric regression is a flexible technique used to estimate the relationship between variables by relaxing the need to specify a parametric model. Some techniques, such as locally weighted scatterplot smoothing (LOESS [33]), assume no parametric form. Other techniques consider an "infinite parameter" approach and use linear combinations of spline bases [34]. These techniques are becoming increasingly popular as many relationships cannot be well approximated using parametric forms.

In general, regression entails modeling the expected value of Y given X

$$E(Y|X) = g(x)$$

Parametric regression assumes a specific form for $g(.)$. Nonparametric regression relaxes this requirement thereby providing more flexibility and avoiding or reducing possible biases in parameter estimates and predictions. The trade-off is that when

the true model follows the specified parametric form parametric regression is more powerful.

A vast array of research is available on nonparametric regression techniques [35]. Existing techniques include: kernel methods (Nadaraya-Watson [36]), LOESS [33] [37], and spline techniques [38] [39]. Kernel regression is a nonparametric technique that estimates $E(Y|X)$ as a locally weighted average of (x, y) pairs, using a kernel as the weight function. For any x in the support of $\boldsymbol{X}$, the estimate is

$$\hat{g}_h(x) = \frac{\sum K_h(x - X_i)Y_i}{\sum K_h(x - X_i)}$$

where K_h is a kernel with bandwidth h. A popular approach to select h is through the use of cross-validation.

LOESS is a nonparametric technique that was developed by Cleveland, Devlin, and Grosse in 1988 and is a staple in nonparametric regression. The basic idea of LOESS is to fit a weighted least squares regression model for each data point (or a given sample of points) at the center of a given neighborhood. The fraction of the data used to estimate the regression, specified as a smoothing parameter, in each local neighborhood controls the overall smoothness of the fit. The advantage to LOESS is it does not require a specification of any function, only the smoothing parameter must be specified. A disadvantage of LOESS is that it requires a dense sample, as the fit relies on the local data in each neighborhood when performing the local fitting. Another disadvantage is there is no easy way to represent LOESS as a regression function. This limits portability as one would need LOESS software and the data to reproduce the regression function.

Splines are another nonparametric estimation technique that is becoming more and more popular. To establish the intuition behind a spline, first let us consider

a polynomial of order k, written as $f(x) = \sum_{i=0}^{k-1} a_i x^i$. Polynomials are simple to implement and a flexible approach for function approximation. However they often suffer from severe overfitting. Several outliers or fluctuations in the polynomial can severely impact the entire fitted curve. To address this, one can consider piecewise polynomials that are only defined in certain regions of X. The places where each piece connects are known as knots and the entire piecewise function is known as a spline [40].

Spline methods can be written as a linear combination of basis functions, and thus carry a certain parametric flavor [41]. Consider the spline $S(x)$ and a $n+1$ knot sequence $\boldsymbol{t}$, with knots at $a = t_0 < t_1 < t_2 < \cdots < t_n = b$. $S_i(x)$ is a polynomial on each subinterval $[t_i, t_{i+1}]$ for $i = 0, 1, \ldots, n-1$. Spline bases, $\boldsymbol{I}$, are constructed from the given knot sequence and specify where each $S_i(x)$ is defined. Each spline basis is associated with a coefficient β_j. The spline estimate for a particular point, x_i, is:

$$S(x_i) = \sum_{j=1}^{j=B} S_j(x_i) = \sum_{j=1}^{j=B} \beta_j I_j(x_i)$$

where B is the number of spline bases.

A key component to any spline procedure is the knot sequence. If knots are not carefully selected, fit can suffer. Spline methods will typically overfit the data if too many knots are selected and underfit the data if too few of knots are selected, or not put in optimal positions.

2.2.1 Nonparametric Monotone Regression

In many problems, including ours, there is prior knowledge that the underlying relationship must be either monotonically increasing (or decreasing). Incorporating

this information into the fit procedure improves estimation efficiency because noise in the data can often mask this feature [42].

To handle monotonicity in the parametric case, one can select an appropriate monotonically increasing (or decreasing) function. In nonparametric regression, one approach is to adjust the unconstrained fitted curve. This method is known as isotonization and dates back to the 1950's. The basic idea is to adjust the curve in regions where the fit is not monotonic. This idea is expressed in detail by Mukerjee [43]. Along these same lines, Neelon and Dunson (2004 [44]) proposed a Bayesian isotonic piecewise linear model. Dette, Neumeyer and Pilz (2006 [45]) introduced a nonparametric monotone regression estimator based on kernel density estimation. They showed, through a simulation study, their approach was superior to that of Neelon's and Dunson's. We discuss this approach further in Section 3.1.2.

None of these methods incorporate measurement error and/or rounding into the estimation process. Qi (2000 [26]), however, showed I-splines work well for interval-censored data with measurement error in a frequentist and non-isotonization setting. We will utilize I-splines in our approach.

2.2.2   I-Splines

I-splines are a monotonically decreasing (or increasing) function and therefore a linear combination of bases with only positive coefficients will ensure a monotonically decreasing (or increasing) function [46]. Note that while positive coefficients ensure a monotonically decreasing relationship, it is possible to have a monotonically decreasing function with some coefficients negative [47].

Consider the knot sequence $t = \{t_1, \ldots, t_{n+k}\}$ where n is the number of free parameters (or bases) and k is the degree of the I-spline. Since we are interested in cubic I-splines, $k = 3$. The knot sequence has the following properties:

1. $t_1 \leq \cdots \leq t_{n+k}$

2. $n - k$ interior knots with k endpoint knots at both L and U

For a given x, let j be the index such that $t_j \leq x \leq t_{j+1}$. I-spline bases can easily be constructed by the recursive algorithm:

$$I_i(x|k,\boldsymbol{t}) = \begin{cases} 0 & i > j \\ \sum_{m=i}^{j}(t_{m+k+1} - t_m)\frac{M_m(x|k+1,\boldsymbol{t})}{k+1} & j - k + 1 \leq i \leq j \\ 1 & i < j - k + 1 \end{cases} \quad (2.9)$$

for $i = 1 \ldots n$, where M_* refers to a quadratic M-spline. One can easily manipulate the I-spline bases to force a monotonically decreasing function by taking 1 minus each I-spline basis: $1 - I_i$. Figure 2.7 shows an example of these basis functions.
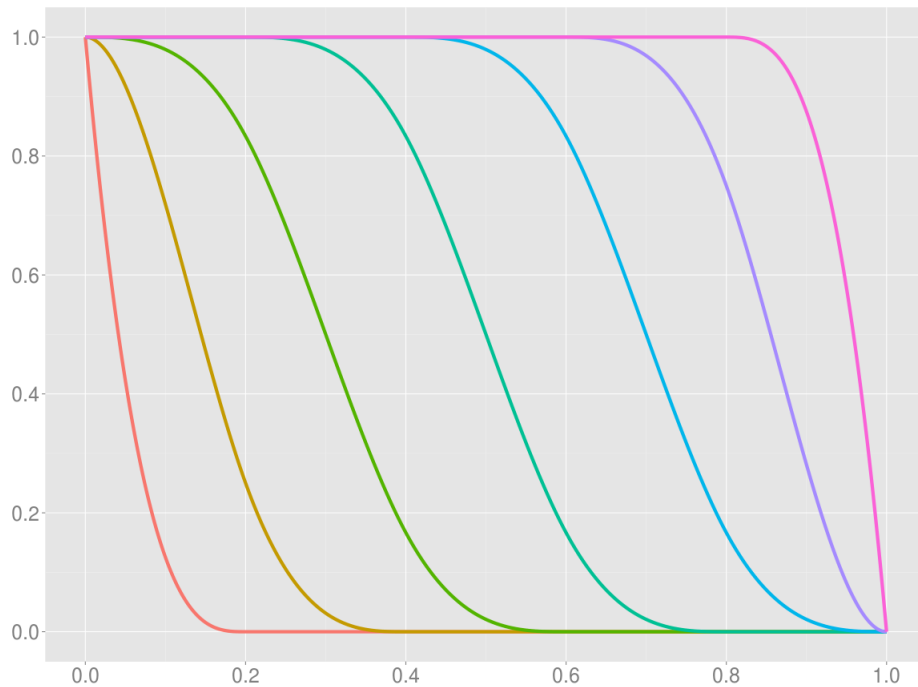

Figure 2.7. Cubic I-splines $(1 - I_i)$ with knots at 0, .2, .4, .6, .8 and 1. Monotonicity is enforced by the shape of the splines.

From the recursion we see I-splines are an extension of M-splines. We consider quadratic M-splines of degree 2 to correspond with cubic I-splines of degree 3. The ith I-spline equals

$$I_i(x|k, \boldsymbol{t}) = \int_L^x M_i(u|k, \boldsymbol{t}) \ du$$

However, for constructing the I-spline bases it is easier to use the recursive algorithm presented previously.

For each M-spline basis, M_i is positive in (t_i, t_{i+k}) and zero elsewhere. In addition, M-splines have the normalization property, $\int M_i(x) \ dx = 1$ and therefore has the properties of a probability density function. M-splines are specified for $t_i \leq x < t_{i+1}$ as:

$$M_i(x|k, \boldsymbol{t}) = \begin{cases} \frac{1}{t_{i+1}-t_i} & t_i \leq x < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{2.10}$$

with $M_i(x|k, \boldsymbol{t})$ defined by the recursion:

$$M_i(x|k, \boldsymbol{t}) = \frac{k[(x - t_i)M_i(x|k - 1, \boldsymbol{t}) + (t_{i+k} - x)M_{i+1}(x|k - 1, \boldsymbol{t})]}{(k - 1)(t_{i+k} - t_i)} \tag{2.11}$$

Before we introduce the full model, we need to discuss incorporating measurement error into a nonparametric approach. Berry et al. 2002 [48], noted that nonparametric estimation in the presence of measurement error is a very difficult subject. It is not clear how to incorporate functional techniques such as, regression calibration and SIMEX, into a nonparametric approach. Kernel regression and LOESS do not provide an easy way to incorporate measurement error into the likelihood function due to their strictly nonparametric nature. The most promising approach to date for handling measurement error within a nonparametric framework is a structural modeling approach using splines.

2.3 Complete Model and Method of Inference

Now that we've described our measurement error setting and have described the nonparametric model for $g(m_i)$, we're ready to describe the full model approach. As stated in Section 1.3.1, we are provided with a scatterplot of (X, Y) test results for a large number of isolates. Our model describes the scatterplot in terms of three components:

1. The test procedures (i.e., rounding and measurement error)
2. The drug/bug-specific relationship between the true MICs and DIAs
3. The underlying distribution of pathogens (or MICs)

The first component links the observed MIC/DIA pair with an underlying true MIC value. The second and third components describe the relationship between the true MIC and its corresponding true DIA.

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

The joint distribution of the observed MIC ($\boldsymbol{x}$) and the observed DIA ($\boldsymbol{y}$) was specified in Section 2.1.6. Past studies have shown that the measurement error distribution for the $\log_2(MIC)$ and DIA are both normal [17] [18]. We introduce latent values, $\boldsymbol{m}$, to represent the corresponding true MIC values. Thus given m_i, the probability we observe x_i is

$$\Phi(x_i; m_i, \sigma_m^2) - \Phi(x_i - 1; m_i, \sigma_m^2)$$

The normal distribution models the measurement error in the MIC test. The difference in normal CDFs accounts for the upwards rounding inherent in the test.

Similarly, for that same isolate, $d_i = g(m_i)$ the probability we observe y_i is

$$\Phi(y_i + .5; d_i, \sigma_d^2) - \Phi(y_i - .5; d_i, \sigma_d^2)$$

Again, the normal distribution describes the measurement error in the DIA test and the difference in normal CDFs account for the rounding. Because the pairs are independent, the complete likelihood of observed test results given $\boldsymbol{m}$ and $g(\boldsymbol{m})$ is

$$\prod_{i=1}^{N}[\Phi(x_i; m_i, \sigma_m^2) - \Phi(x_i - 1; m_i, \sigma_m^2)]\times$$
$$[\Phi(y_i + .5; d_i, \sigma_d^2) - \Phi(y_i - .5; d_i, \sigma_d^2)] \tag{2.12}$$

2.3.2 True Underlying Relationship

We use I-splines to model the underlying relationship between the MIC and DIA test (Section 2.2.2). Given a knot sequence and bases ($\boldsymbol{I}$), we connect the true DIA test results to the true MIC test results by

$$d_i = g(m_i) = \sum_{j=1}^{j=B} \beta_j I_j(m_i) \tag{2.13}$$

where $\boldsymbol{\beta}$ are the unknown spline coefficients and B is the number of bases.

In order to implement I-splines in our model we must select the knot sequence. Common knot selection approaches are based on either fit statistics and/or cross validation [49]. One may select the knot sequence that minimize a goodness of fit statistic, such as AIC or BIC. To find this set, one can place many knots at equally spaced intervals and find the knot sequence that minimizes the goodness of statistic via a backwards selections approach. Cross-validation procedures are similar, for

each specified knot sequence a subset of the data is fit and the knot sequence that minimizes some sort of fit criterion, such as the mean squared error, is selected.

In our approach, the true MIC values are unknown and treated as additional parameters. Since we are not dealing with a fixed set of data points, this makes the knot selection process difficult as selection techniques based on fit statistics do not suffice. Ruppert et al. [34] proposed using penalized B-splines (known as P-splines). This approach can be thought of as the equivalent of least absolute shrinkage and selection operator (LASSO) for regression [50]. The main idea was to introduce a penalty parameter into the estimation that serves to limit the number of knots. This idea is more promising, however enforcing the monotonicity constraint within this framework is not straight-forward.

We propose two solutions for knot selection: (1) consider a set of equally-spaced knots across the range of $\boldsymbol{x}$, and (2) consider knot locations and the number of knots as additional unknown parameters. The method for each proposed solution is described in detail in Section 2.6.

The collection of pathogen strains used in the DIA breakpoint determination process is commonly considered to be a random sample of the pathogens that would be seen at a hospital or clinic. Thus, we use the sample of test results to estimate an underlying population distribution.

Craig (2000) used a mixture of normals model to describe this distribution. This is a flexible model that allows for multi-modality and/or skewness in the fitted distribution. The mixture of normals model can be written as:

$$\pi(\boldsymbol{m}|\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{\sigma^2},K) = \prod_{i=1}^{N}\sum_{j=1}^{K} p_j\,\mathcal{N}(m_i|\mu_j,\sigma_j^2) \tag{2.14}$$

where K is the number of normals in the mixture and $\boldsymbol{p},\boldsymbol{\mu},\boldsymbol{\sigma^2}$ are the K-dimensional vector of parameters for each component.

A key issue when using this mixture model is determining the number of components in the mixture. Standard methods involve investigating various K and selecting the best one based on a fit statistic, such as AIC, or using cross validation [51]. In our case, however, the true MIC values are unknown parameters in the model, so these methods are not feasible. We instead use a Dirichlet Process Mixture of Normals (Ghosh and Ramamoorthi, 2003 [52]) to address the determination of K.

The nonparametric Dirichlet Process (DP) avoids the problem of having to specify the number of mixture components. The Dirichlet Process is a distribution over discrete probability distributions. This restriction makes modeling a continuous distribution difficult using a DP. However, this restriction is resolved by the Dirichlet Process Mixture (DPM) that draws parameters of a mixture model based on a draw from the DP.

Intuitively the Dirichlet process can be thought of as a Bayesian mixture model consisting of K components, where K is chosen based on the data being analyzed:

$$
\begin{aligned}
\theta_k &\sim H \quad k = 1 \ldots K \\
\boldsymbol{p}|\alpha &\sim Dirichlet(\alpha/K, ..., \alpha/K) \\
\text{for} \quad i &= 1 \ldots N \\
z_i|\boldsymbol{p} &\sim Mult(\boldsymbol{p}) \\
m_i &\sim F(\theta_k)
\end{aligned}
$$

where z_i is the group indicator, the vector $\boldsymbol{p}$ is the mixing proportion, α is the concentration of the Dirichlet prior, H is the prior distribution over component parameters θ (known as the central distribution), and $F(\theta)$ is the component distribution parameterized by θ. Since we are modeling a mixture of normals, F is the normal distribution and θ is a vector containing the unknown mean and variance. We choose

H as the conjugate prior of F, the normal-inverse-gamma distribution, for computation convenience and analytic tractability. Marginalizing out $\boldsymbol{p}$ and $\boldsymbol{z}$ leads to a Dirichlet-Normal distribution.

The DPM assumes that the prior on θ can have any form. The terms in the form of a general DP are

$$
\begin{aligned}
\theta_i | G &\sim G \\
m_i | G &\sim F(\theta_i) \\
\text{for} \quad i &= 1 \ldots N
\end{aligned}
$$

where G is any probability distribution drawn from a DP, H is the central distribution of the DP, and α is the concentration parameter. Smaller values of the concentration parameter result in fewer discrete distributions and vice versa. Integrating out G leads to the DPM.

We use JAGS (Just Another Gibbs Sampler) to approximate the Dirichlet mixture of normals process [53]. It is only an approximation as the number of possible mixture components should be unbounded, however we limit the number of components to a maximum of 20. In practice the number of components rarely exceeds 8. We refer to the resulting density as $f(\boldsymbol{m})$.

2.4 Determination of DIA Breakpoints

Given our structural model (2.8), we can compute the probability that the observed MIC will fall in any region. Traditionally, MIC breakpoints only referred to the test results. Given our modeling of the underlying truth, we need to make a distinction between test MIC and true MIC breakpoints. Due to the upwards rounding of the MIC test, we consider the true MIC breakpoints to be shifted down 0.5. Thus

if the observed MIC breakpoints are -1 and 1 the true MIC breakpoints are -1.5 and 0.5. The test MIC breakpoints are referred to as M_L and M_U while the true MIC breakpoints are referred to as M_L^* and M_U^*. The probability of correct MIC and DIA classification for given a MIC, m, is

$$
p_{\text{MIC}}(m) = \begin{cases} \Pr(x \leq M_L) = \Phi\left(\frac{M_L - m}{\sigma_m}\right) & m \leq M_L^* \\ \Pr(M_L < x < M_U) = \Phi\left(\frac{M_U - 1 - m}{\sigma_m}\right) - \Phi\left(\frac{M_L - m}{\sigma_m}\right) & M_L^* < m < M_U^* \quad (2.15) \\ \Pr(x \geq M_U) = 1 - \Phi\left(\frac{M_U - 1 - m}{\sigma_m}\right) & m \geq M_U^* \end{cases}
$$

$$
p_{\text{DIA}}(m) = \begin{cases} \Pr(y \geq D_U) = 1 - \Phi\left(\frac{D_U - .5 - g(m)}{\sigma_d}\right) & m \leq M_{L^*} \\ \Pr(D_L < y < D_U) = \Phi\left(\frac{D_U - .5 - g(m)}{\sigma_d}\right) - \Phi\left(\frac{D_L + .5 - g(m)}{\sigma_d}\right) & M_{L^*} < m < M_{U^*} \quad (2.16) \\ \Pr(y \leq D_L) = \Phi\left(\frac{D_L + .5 + g(m)}{\sigma_d}\right) & m \geq M_{U^*} \end{cases}
$$

We estimate the DIA breakpoints by finding the set that minimize a weighted loss function. Our loss function is the accumulated squared difference in the probability of correct classification when the DIA test performs worse than the MIC test. We use the underlying MIC density for our weights. This can be expressed

$$
L = \int_{-\infty}^{\infty} \min\left(0, p_{\text{DIA}}(g(u)) - p_{\text{MIC}}(u)\right)^2 w(u)\, du \qquad (2.17)
$$

where $w(.)$ are the weights. Alternative loss functions may be considered. For example, one may wish to weigh the resulting fit more heavily near the MIC indeterminant range. We also investigate an alternative loss function that closely resembles the target of the ERB method in Section 3.3.1.

2.5 Model Parameters

The following is a list of unknown model parameters that must be estimated from the scatterplot of data:

1. $\boldsymbol{m}$ - true MIC values

2a. $\boldsymbol{\beta}$ - I-spline coefficients

2b. smoothing parameter, λ, for a fixed number of equally spaced knots **or** the number of knots and their locations (k and t_j)

3. $f(\boldsymbol{m})$ -the underlying pathogen distribution

To estimate these parameters from a scatterplot, we take a Bayesian approach. Qi proposed a similar model but estimated the parameters using a two-stage frequentist approach. We wish to incorporate the full information in the data when estimating the underlying MIC density and underlying MIC/DIA relationship. To do this in a frequentist setting is very difficult as there are many unknown parameters, and therefore a maximum likelihood approach may become computationally infeasible. We believe this to be one of the reasons Qi used two stages in his approach.

Parameter estimation is much easier in a Bayesian framework using Markov Chain Monte Carlo (MCMC). In addition, a Bayesian analysis allows us to easily construct credible intervals and thereby produce a distribution of possible DIA breakpoints. Qi resorted to a bootstrap procedure to obtain confidence intervals, and this involved even more computation.

2.6 Bayesian Computation

When considering a fixed set of equally spaced knots, we're interested in the posterior $\pi(\boldsymbol{\beta}, f(\boldsymbol{m}), \lambda | \boldsymbol{x}, \boldsymbol{y}, \sigma_m, \sigma_d)$, where $f(\boldsymbol{m})$ is the distribution of the true MIC and $\{\beta, \lambda\}$ define the relationship between the true MIC and DIA. To make this

computation easier we consider the individual true MICs as additional parameters. By doing so, the posterior can be broken down as follows

$$\pi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{m})\pi(\boldsymbol{\beta}, \lambda)\times$$
$$\pi(\boldsymbol{m}|f(\boldsymbol{m}))\times$$
$$\pi(f(\boldsymbol{m})|\boldsymbol{\mu}, \boldsymbol{\sigma^2}, K) \tag{2.18}$$

2.7 Priors

With any Bayesian analysis, prior distributions are required on the parameters [54]. When possible, we have chosen relatively noninformative priors for each of the parameters.

When considering many knots at equally spaced intervals, we can constrain the coefficients via a random walk prior (2.19) [55].

$$\beta_t|\beta_0...\beta_{t-1} \sim \mathcal{N}(\beta_{t-1}, \lambda) \tag{2.19}$$

Each coefficient has a normal prior with mean equal to the current coefficient estimate and λ serves as a smoothness parameter. A non-informative prior is used for β_0, $\beta_0 \sim \mathcal{N}(0, 100)$. We place knots every 0.5 $\log_2$ units. This prior forces each spline coefficient to be relatively close to the previous spline coefficient to avoid overfitting.

We also investigated constraining the coefficients using the prior:

$$\beta_t \sim ddexp(0, \sigma_\beta^2) \tag{2.20}$$

where *ddexp* represents the double exponential distribution. The β's are apriori assumed independent. The double exponential distribution (or Laplace distribution) is

the Bayesian equivalent of LASSO [56]. This prior shrinks the coefficients towards zero to prevent overfitting. However we found the random walk prior resulted in more accurate and stable results.

When considering the number and location of knots as unknown parameters, we add a process to add, delete, or move knots within our Markov chain. This is performed using Reversible Jump MCMC (RJMCMC) [57].

Model 1 refers to using the random walk prior for knot selection. Model 2 refers to using the reversible jump step to update the unknown knot locations and number. $\mathcal{LN}$ is the log normal distribution.

In summary, the priors for each model are as follows

- Priors for the parameters modeling the MIC density: $\pi(f(\boldsymbol{m})|\boldsymbol{\mu}, \boldsymbol{\sigma^2}, K)$

  1. $\pi(\boldsymbol{\mu}) \sim \mathcal{N}(0, 1000)$

  2. $\pi(\boldsymbol{\sigma^2}) \sim InvGamma(.001, .001)$

  3. $\pi(K|\alpha) \sim Dirichlet(\alpha/K)$

- Model 1:

  1. $\pi(\beta_0) \sim \mathcal{LN}(0, 100)$

  2. $\pi(\beta_i|\beta_{i-1}, \lambda) \sim \mathcal{LN}(\beta_{i-1}, \lambda)\ i = 2 \ldots n$

  3. $\pi(\lambda) \sim Uniform(0, 2)$

- Model 2:

  1. $\pi(\boldsymbol{\beta}) \sim \mathcal{N}(15, 500)$

  2. $k \sim TPOI(3)$ where $TPOI(3)$ is a truncated Poisson distribution (no probability is assigned at 0) with $\lambda = 3$. This results in a mean number of knots of 3.15 with a 95% confidence interval of 1 and 7 knots.

3. The interior knot locations are assumed apriori independent with each

$$t_j \sim Uniform(t_{j-1}, t_{j+1})$$

2.8   Markov Chain Monte Carlo (MCMC)

Because the posterior cannot be written in closed form, we approximate the posterior using Markov Chain Monte Carlo (MCMC) to draw samples from the representative posterior [58]. Throughout the chain, a new estimate is introduced and either accepted or rejected. To do this we use the Metropolis-Hastings updates [59] [60].

The Metropolis-Hasting algorithm allows us to update the parameter estimates during the MCMC process. Suppose the specified distribution has unnormalized density $h(.)$. A Metropolis-Hastings update involves the following steps:

1. Given the current state is x, propose a move to y using a proposal density denoted $q(.)$. This density may or may not depend on the current state x.

2. Calculate the ratio:
$$r(x, y) = \frac{h(y)q(x|y)}{h(x)q(y|x)}$$

3. Accept the proposed move to y with probability:

$$a(x, y) = min(1, r(x, y))$$

4. Thus the resulting state is y with probability $a(x, y)$ and x with probability $1 - a(x, y)$

2.8.1   Updating Process

In order to construct our Markov chain the first step is to initialize the unknown model parameters. We then begin updating the parameters. When possible we consider symmetric proposal distributions and which cancel out in the Metropolis-Hasting update procedure. Here are the details of our MCMC processes:

1. Update $\boldsymbol{m}$ (and $\boldsymbol{d}$)

 – Proposal distribution: $q(m_i^*|m_i) = \mathcal{N}(m_i, .5)$

$$\alpha = \frac{\pi(m_i^*|f(\boldsymbol{m}))\pi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\beta}, m_i^*)}{\pi(m_i|f(\boldsymbol{m}))\pi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\beta}, m_i)}$$

2. Update $\boldsymbol{\beta}$

 – For model 1, $\boldsymbol{\beta}$ is restricted to be positive. For model 2 there is no restriction on $\boldsymbol{\beta}$.

 – Proposal distribution - Model 1: $q(\boldsymbol{\beta^*}|\boldsymbol{\beta}) = \mathcal{MLN}(\boldsymbol{\beta}, \Sigma)$

 – Proposal distribution - Model 2: $q(\boldsymbol{\beta^*}|\boldsymbol{\beta}) = \mathcal{MN}(\boldsymbol{\beta}, \Sigma)$

 – Update Model 1:
$$\alpha = \frac{\pi(\boldsymbol{\beta^*}, \boldsymbol{\lambda})\pi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\beta^*}, \boldsymbol{m})}{\pi(\boldsymbol{\beta}, \boldsymbol{\lambda})\pi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{m})}$$

 – Update Model 2:
$$\alpha = \frac{\pi(\boldsymbol{\beta^*})\pi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\beta^*}, \boldsymbol{m})}{\pi(\boldsymbol{\beta})\pi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{m})}$$

 – We consider an adaptive M-H step here where

$$\Sigma = \begin{cases} \text{Diagonal covariance with variance equal to 0.2} & i < 1000 \\ \\ Cov(\boldsymbol{\beta}) \text{ of the past 500 estimates} & i \geq 1000 \end{cases}$$

 – If model 2 check to make that resulting fit is monotonically decreasing

3. Update λ (Model 1)

 – Proposal distribution: $q(\lambda^*|\lambda) = \mathcal{U}(\lambda - .1, \lambda + .1)$

$$\alpha = \frac{\pi(\lambda^*)\pi(\boldsymbol{\beta}|\lambda^*)}{\pi(\lambda)\pi(\boldsymbol{\beta}|\lambda)}$$

    – Note: automatic reject if proposal is negative

3. Update knot sequence $\boldsymbol{t}$ (Model 2)

    – Described in Section 2.8.2

4. Update $f(\boldsymbol{m})$ - Dirichlet Process

    – This is done in JAGS using a variation of the algorithm written by Johnson et al. [55].

Step 4, updating the MIC distribution, is computationally expensive. Therefore we recommend not performing this update at every iteration but instead around every 200 iterations.

2.8.2 Reversible Jump Step

RJMCMC for knot selection has been proposed before in the literature. DiMatteo et al. in 2001 [61] showed how knot selection can be implemented in a reversible jump MCMC setting. This procedure is known as free-knot splines and consisted of adding, deleting, or moving a knot based on the least-square (LS) coefficients.

RJMCMC knot selection for monotonically decreasing splines was proposed by Wang in 2008 [62]. Wang imposed the monotone shape restriction by using cubic splines and second-order cone programming with a truncated normal prior. This approach, however, assumes a fixed set of data and therefore does not directly translate to our situation.

In 2000, Meyer and Hackstadt proposed a Bayesian monotone regression estimation with RJMCMC for knot selection [47]. This method used shape-restricted splines (SRS) to ensure monotonicity. However their approach did not consider measurement error and we encountered it does not work well for our interval censored data. We

found the estimated curve to not be stable or accurate in most situations. This is most likely due to the changing nature of the true data points, $\boldsymbol{m}$ and $\boldsymbol{d}$.

Our RJMCMC procedure is based on the idea that the current spline coefficients, for given $\boldsymbol{m}$ and $\boldsymbol{d}$, should be relatively close to the LS coefficient estimates. We define $(t_0, t_1, \ldots, t_k)$ as the current knot sequence, $\boldsymbol{\beta_{LS}}$ the current LS spline coefficients, $\boldsymbol{\beta}$ as the current spline coefficient estimates, and Σ_{LS} as the LS covariance matrix. The current estimated true DIA values are $\boldsymbol{d}$.

Once a new model has been proposed, we define $\boldsymbol{\beta_{LS}^*}$ as the new LS coefficients with corresponding covariance matrix Σ_{LS}^*, and the proposed new coefficients as $\boldsymbol{\beta^*}$. $\boldsymbol{\beta^*}$ is drawn from a multivariate normal distribution with $\mu = \boldsymbol{\beta}$ and $\Sigma = \Sigma_{LS}^*$. The proposed new DIA values, based on $\boldsymbol{\beta^*}$ and Σ_{LS}^*, are $\boldsymbol{d^*}$.

Birth

We first consider the case of adding a knot to the existing knot sequence. Let $t = (t_1, ..., t_{n+k})$ be the ordered knot locations with $n - k$ interior knots. Recall k represents the degree of the I-splines, and since we use cubic I-splines $k = 3$. A birth move adds a new knot, t^*, selected uniformly from t_k to t_{n-k}. t^* will fall between two knots, t_j and t_{j+1}. The acceptance probability, A, is given by:

$$A = \frac{\pi(\boldsymbol{y}|\boldsymbol{d^*})\pi(\boldsymbol{d^*}|\boldsymbol{m},\boldsymbol{\beta^*})}{\pi(\boldsymbol{y}|\boldsymbol{d})\pi(\boldsymbol{d}|\boldsymbol{m},\boldsymbol{\beta})} \times \frac{\pi(\boldsymbol{\beta^*})\pi(k+1\,knots)}{\pi(\boldsymbol{\beta})\pi(k\,knots)} \times \frac{\mathcal{MN}(\boldsymbol{\beta}|\boldsymbol{\beta_{LS}},\Sigma_{LS})}{\mathcal{MN}(\boldsymbol{\beta^*}|\boldsymbol{\beta_{LS}^*},\Sigma_{LS}^*)} \frac{\frac{1}{k+1}}{\frac{1}{k+1}\frac{1}{t_{j+1}-t_j}} \times |1|$$

Move

A relocation move keeps the same number of interior knots, k, but the location of one knot is moved. From the set of existing knots, an interior knot is randomly

selected to be moved and denoted t_{j+1}. The new knot location is proposed uniformly from t_j to t_{j+2}. The acceptance probability, A, is given by:

$$A = \frac{\pi(\boldsymbol{y}|\boldsymbol{d^*})\pi(\boldsymbol{d^*}|\boldsymbol{m},\boldsymbol{\beta^*})}{\pi(\boldsymbol{y}|\boldsymbol{d})\pi(\boldsymbol{d}|\boldsymbol{m},\boldsymbol{\beta})} \times \frac{\pi(\boldsymbol{\beta^*})\pi(k\,knots)}{\pi(\boldsymbol{\beta})\pi(k\,knots)} \times \frac{\mathcal{MN}(\boldsymbol{\beta}|\boldsymbol{\beta_{LS}},\Sigma_{LS})\,\frac{1}{k}\frac{1}{t_{j+1}-t_{j-1}}}{\mathcal{MN}(\boldsymbol{\beta^*}|\boldsymbol{\beta_{LS}^*},\Sigma_{LS}^*)\,\frac{1}{k}\frac{1}{t_{j+1}-t_{j-1}}} \times |1|$$

Death

A death move removes one of the $n-k$ interior knots from the model. Let t_{j+1} be the randomly selected knot to remove. The acceptance probability, A, is given by:

$$A = \frac{\pi(\boldsymbol{y}|\boldsymbol{d^*})\pi(\boldsymbol{d^*}|\boldsymbol{m},\boldsymbol{\beta^*})}{\pi(\boldsymbol{y}|\boldsymbol{d})\pi(\boldsymbol{d}|\boldsymbol{m},\boldsymbol{\beta})} \times \frac{\pi(\boldsymbol{\beta^*})\pi(k-1\,knots)}{\pi(\boldsymbol{\beta})\pi(k\,knots)} \times \frac{\mathcal{MN}(\boldsymbol{\beta}|\boldsymbol{\beta_{LS}},\Sigma_{LS})\,\frac{1}{k}\frac{1}{t_{j+1}-t_{j-1}}}{\mathcal{MN}(\boldsymbol{\beta^*}|\boldsymbol{\beta_{LS}^*},\Sigma_{LS}^*)\,\frac{1}{k}} \times |1|$$

In the next chapter we investigate the performance of our approach and compare it to the ERB approach, an extended four-parameter logistic approach, and a two-stage frequentist approach. We investigate how well each approach estimates DIA breakpoints for a variety of MIC/DIA relationships and MIC pathogen distributions. We also compare the relative performance of the random walk prior versus the reversible jump step for knot selection. We investigate the BNP approach for unique/unusual underlying MIC densities as well as different number of sample isolates. Finally, we assess the performance of various approaches on three real data sets.

3. RESULTS

In this chapter we consider a series of simulation studies to assess the performance of various approaches for DIA breakpoint estimation. These simulation studies were designed to investigate the performance of particular features of our approach as well as to compare our approach to competing methods.

3.1 Simulation Study Details

DIA breakpoint determination can be affected by many factors, such as the underlying pathogen density or the true underlying relationship between the MIC and DIA. Our simulation study is therefore a series of studies, involving a manipulation of at least one of the following five conditions:

1. The underlying pathogen (true MIC) density
2. The true underlying relationship between the two tests
3. Degree of measurement error in each test
4. Location of MIC breakpoints
5. Number of pathogen strains used in calibration

Since each of these five conditions has numerous possible levels, the potential number of studies (combination of levels) is very large. We, therefore, select only a handful of these combinations to cover the range of likely scenarios. Each chosen combination is shown in a table at the beginning of each study section.

3.1.1 Assessing Model Performance

To compare the methods, we considered several performance summaries. Model performance is assessed by the proportion of time the correct DIA breakpoints are selected, how close the estimated MIC/DIA relationship is to the truth, and how close the MIC density is to the truth. The latter two are assessed using sum of squared errors (SSE)

$$SSE = \int_{c_1}^{c_2} (\hat{f}(c) - f(c))^2 \, dc$$

where $f(.)$ is the true underlying density or relationship, and $\hat{f}(.)$ is the estimated density or relationship. Instead of computing this integral exactly it was approximated using a sum of 500 equally spaced MIC values from c_1 to c_2.

Since nonparametric approaches are known to deviate near the boundary points where there is little or no data, the SSE for the underlying MIC/DIA relationship is only calculated between -2.5 and 2.5 (0 is at the center of the MIC breakpoints), while the underlying MIC density was evaluated from -6 to 6. In all tables, the SSE statistics for the density fit are multiplied by 100 for easier reading.

3.1.2 DIA Breakpoint Estimation Approaches

We consider the following four breakpoint estimation methods in the scenarios of Section 3.3.

Method I - Bayesian Nonparametric Model (BNP and BNPRJ)

This is our Bayesian nonparametric approach. The model is referred to as BNP when we use a fixed number of equally-spaced knots and BNPRJ when we treat the number and location of knots as additional unknown parameters in the model.

Method II - Four-Parameter Logistic Model (L4P)

Because Xi (2005) demonstrated that the three-parameter logistic model can often fit curves poorly, we extend Craig's three-parameter logistic model to a more flexible four-parameter logistic model using the following parameterization

$$g(m_i) = \beta_1 \frac{(fx \exp(\beta_3(\beta_2 - m_i))) + ((1 - fx) \exp(\beta_4(\beta_2 - m_i)))}{1 + (fx \exp(\beta_3(\beta_2 - m_i)) + (1 - fx) \exp(\beta_4(\beta_2 - m_i)))} \quad (3.1)$$

where

$$fx = 1/(1 + \exp(-mb \times (\beta_2 - m_i))) \text{ and } mb = (2 \times \beta_3 \times \beta_4)/(\beta_3 + \beta_4)$$

This four-parameter version allows for asymmetry around the inflection point β_2 and reduces to the three-parameter model when $\beta_3 = \beta_4$. This parameterization is sometimes referred to as a five-parameter logistic model in the literature, because they consider mb as an additional parameter [63]. The overall fit is the result of two logistic curves, with the weights being fx and $1 - fx$.

In addition to the coefficients all being positive, the formulation of mb guarantees a monotonically decreasing function. One can express mb as the inverse of the average reciprocal rate: $mb = (\frac{1/\beta_3 + 1/\beta_4}{2})^{-1}$. This forces the weights to lean towards the smaller of the two rates so that we do not end up weighting the slowly decreasing curve too highly, which could cause an increase in the curve.

Method III - Frequentist Nonparametric Regression (FNP)

Qi was the first to consider using nonparametric regression to describe the true MIC/DIA relationship. He took a two-step approach, first fitting the underlying MIC distribution using M-splines and then, based on this distribution, estimating

the MIC/DIA relationship using I-splines. For estimating the MIC distribution, measurement error and interval censoring are accounted for in the likelihood. The estimated MIC distribution was then used in estimating the MIC/DIA relationship in the second step.

For our comparisons here, we consider a two-step procedure that approximates Qi's method. We first estimate the underlying MIC distribution using a Dirichlet process mixture of normals. To account for the upward rounding of the MIC test, the data are shifted to the left 0.5 before estimation. Measurement error is accounted for in the likelihood equation.

For the second step, the MIC/DIA relationship is fit using a monotonically decreasing nonparametric kernel estimator proposed by Dette, Neumeyer and Pilz (2006) [45]. We again shift the data to the left 0.5 before estimation to account for rounding and measurement error. Unlike Qi's approach, the estimated MIC distribution is not considered in this step.

Dette, Neumeyer, and Pilz's monotonically increasing (decreasing) estimate is constructed in three steps:

1. Start with an unconstrained estimate of the regression function, $\hat{q}$.

2. Calculate a density estimate of the observations $\hat{q}(m_i)$, which is then integrated to obtain an estimate of the inverse of the regression function.

3. Inverse the estimate from step 2.

These three steps are summarized in equations 3.2 and 3.3. Consider the regression function q, then

$$\hat{q}^{-1}(t) = \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{t} K_d \left(\frac{\hat{q}(i/N) - u}{h_d} \right) du \qquad (3.2)$$

is an estimate of $q^{-1}(t)$, and

$$\hat{q}(x) = \frac{\sum\limits_{i=1}^{N} K_r((X_i - x)/h_r)Y_i}{\sum\limits_{i=1}^{N} K_r((X_i - x)/h_r)} \qquad (3.3)$$

is the Nadaraya-Watson estimate where K_d and K_r are symmetric kernels.

Through simulations we found setting both smoothing parameters equal to 0.1 provided consistent and accurate fits to the true underlying relationship. While this method is not the exact method proposed by Qi, we believe it to be roughly equivalent in terms of model performance.

Method IV - Error-Rate Bounded (ERB)

Method IV is the ERB method as described in Section 1.4.1.

3.1.3 True Relationship between the Two Tests ($g(.)$)

We investigate five possible MIC/DIA relationships. The first relationship is a linear relationship, chosen to represent the most simple relationship and one used in the past. The second and third relationships are logistic relationships that more closely resemble the type of data typically seen in experiments. The fourth and fifth relationships are unusual and, while unlikely to be seen in practice, challenges the flexibility of the approaches. Figures 3.1 - 3.5 visually display the relationships considered.
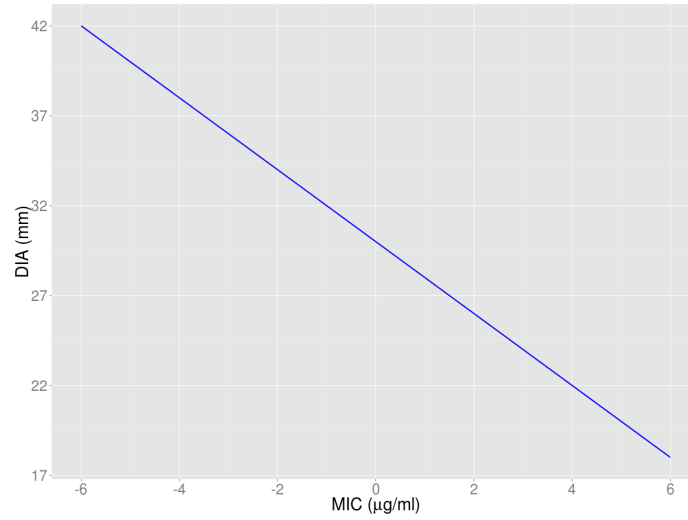
Figure 3.1. MIC/DIA Relationship 1 - Linear: $d_i = 30 - 2m_i$
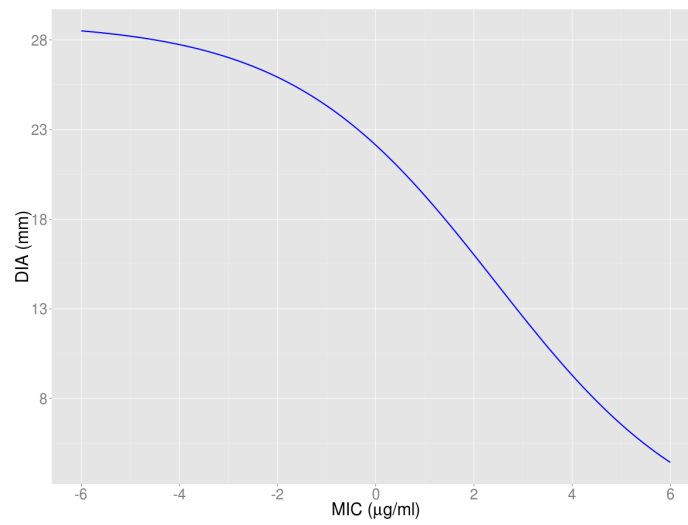
Figure 3.2. MIC/DIA Relationship 2 - three-parameter logistic: $d_i = 29\frac{exp(1.17-0.48m_i)}{1+exp(1.17-0.48m_i)}$
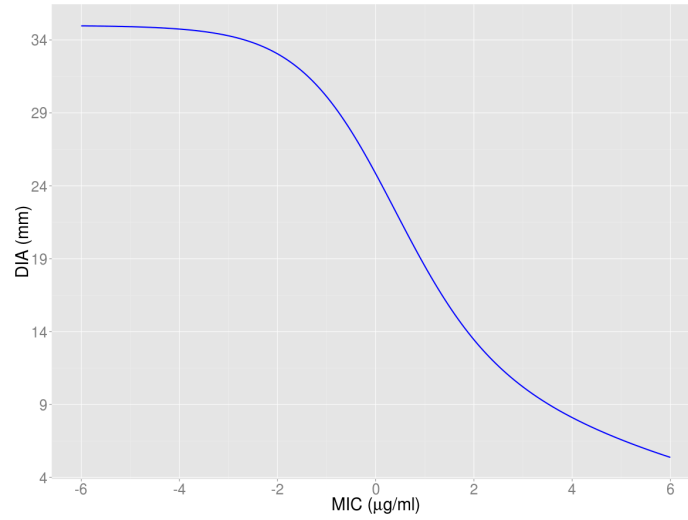
Figure 3.3. MIC/DIA Relationship 3 - four-parameter logistic: coefficients: $35, 1.17, 0.1, 1.2$
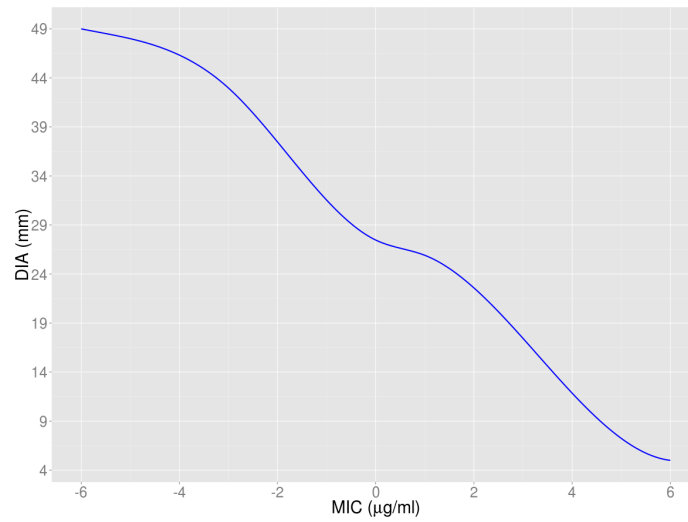
Figure 3.4. MIC/DIA Relationship 4 - Nonparametric 1: I-spline with knots: $-6. - 3, 0, 1, 6$ and coefficients: $1, 10, 1, 25, 1, 1, 10, 1$
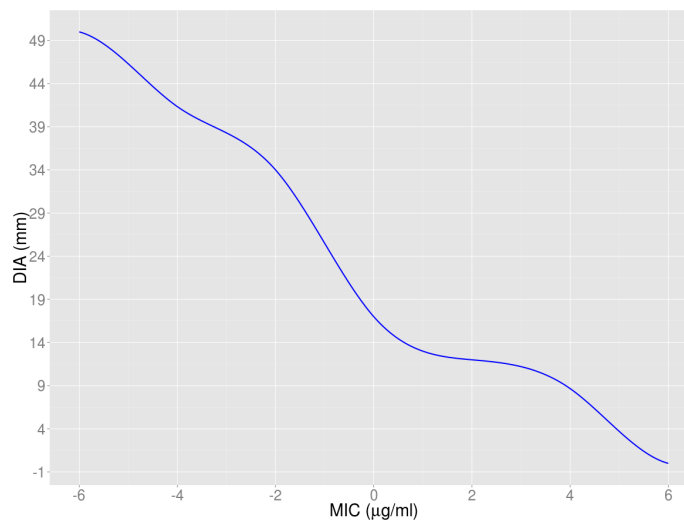
Figure 3.5. MIC/DIA Relationship 5 - Nonparametric 2 - I-spline with knots: $-6. -4, -2, 0, 2, 4, 6$ and coefficients: $1, 10, 1, 25, 1, 1, 10, 1$

3.1.4 True Underlying MIC Distribution

Five different mixture of normal distributions were considered to represent the underlying MIC distribution. The first density has three distinct components and therefore should be accurately estimated. Densities two, three, and four are more likely to be seen in practice. Most of the density is in the susceptible range but there is a gradual decrease towards the resistant range. This is in contrast to density five, in which there is a clear separation between the susceptible and resistant range and little density near the indeterminant range. Figures 3.6 - 3.10 visually display these densities.
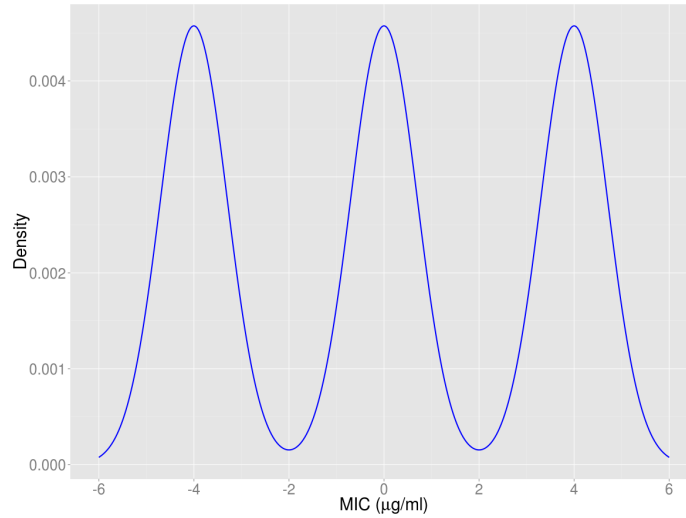
Figure 3.6. MIC Density 1 - Symmetric Three Component Mixture - $\boldsymbol{\mu} = (-4, 0, 4)$, $\boldsymbol{\sigma} = (0.7, 0.7, 0.7)$, and $\boldsymbol{p} = (1/3, 1/3, 1/3)$
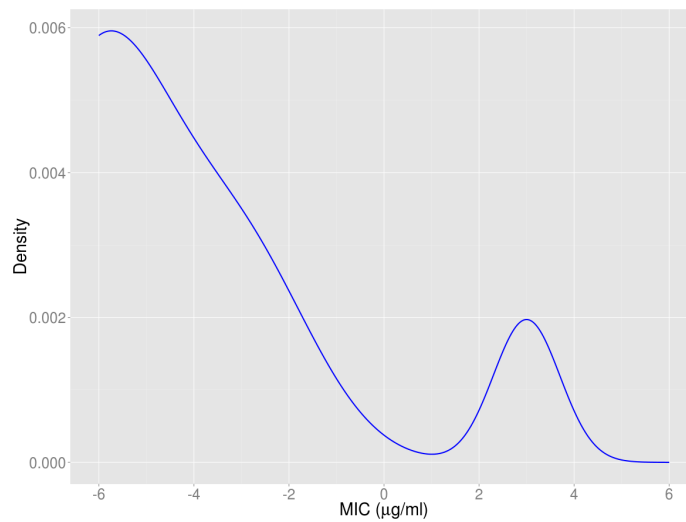
Figure 3.7. MIC Density 2 - Right Skewed Bimodal - $\boldsymbol{\mu} = (-6, -3, 3)$, $\boldsymbol{\sigma} = (1.5, 1.5, 0.7)$, and $\boldsymbol{p} = (0.6, 0.3, 0.1)$
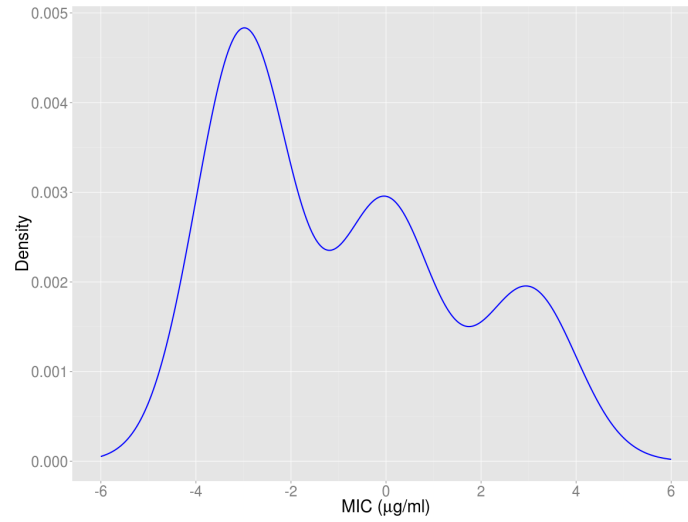
Figure 3.8. MIC Density 3 - Right Skewed Trimodal - $\boldsymbol{\mu} = (-3, 0, 3)$, $\boldsymbol{\sigma} = (1, 1, 1)$, and $\boldsymbol{p} = (0.5, 0.3, 0.2)$
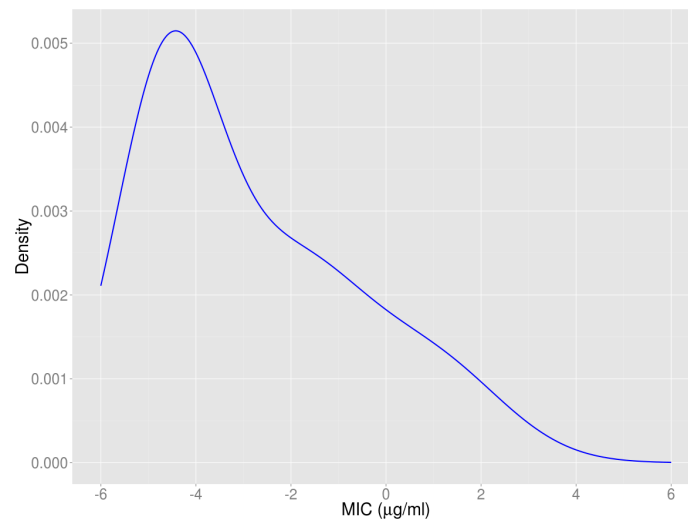
Figure 3.9. MIC Density 4 - Right Skewed Unimodal - $\boldsymbol{\mu} = (-4.3, -1, 2)$, $\boldsymbol{\sigma} = (1, 1.9, 1.5)$, and $\boldsymbol{p} = (0.5, 0.33, 0.17)$
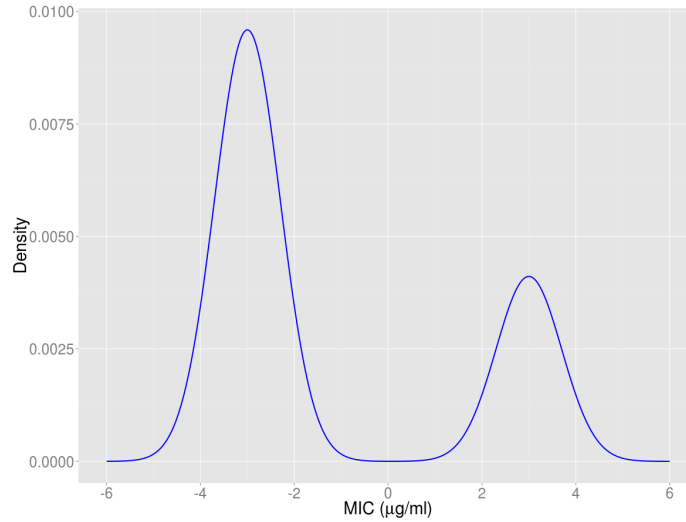
Figure 3.10. MIC Density 5 - Nonsymmetric Bimodal - $\boldsymbol{\mu} = (-3, 3)$, $\boldsymbol{\sigma} = (0.7, 0.7)$, and $\boldsymbol{p} = (0.7, 0.3)$

3.2 Simulation Details

For each scenario we consider 200 scatterplots. The generation of each scatterplot involved:

1. Randomly sampling MIC isolates from the specified MIC density.

2. Determining the true DIA using the known MIC/DIA relationship.

3. Generating the observed test pair by adding independent measurement errors to each true value and then rounding appropriately.

Each scenario was performed using the statistical package R [64] with assistance from the following packages:

- *mvtnorm*: used to simulate data from a multivariate normal distribution [65]

- *multicore*: used for parallel computation [66]

- *mixtools*: used to generate data from a mixture of normal distributions [67]

- *ggplot2*: used for graphing data [68]

- *R2jags*: communicate with JAGS (just another gibbs sampler) for the Dirichlet mixture of normals estimation of the underlying MIC density [69] [53]

- *mclust*: used for parameter estimation for a mixture of normals distribution [70]

- *monreg*: used for quickly estimating a monotonically decreasing relationship [71]

3.2.1   Simulation Scenarios

For the first simulation scenario we compare the BNP approach (Method I) to the error-rate bounded (ERB) approach (method IV). For this case, we modify our loss function to more closely resemble the target of the ERB method. We do this because the ERB target is different and we want to compare apples to apples. Both Craig's and Qi's earlier comparisons did not consider this. This scenario starts on page 67.

The second scenario compares the BNP approach (Method I) to the frequentist two-stage (FNP) approach (Method III). Our hypothesis was that estimating both the MIC density and underlying MIC/DIA relationship in one step results in a better fit and therefore more accurate DIA breakpoints. This scenario starts on page 75.

The third scenario compares the BNP approach (Method I) to the four-parameter logistic (L4P) approach (Method II). When the true underlying MIC/DIA relationship deviates from a logistic fit, we expect the BNP approach to do better. This scenario starts on page 104.

The fourth scenario investigates the two methods of knot selection for the BNP approach. We expect both methods to produce similar results. This scenario starts on page 113.

Scenario five investigates how well the BNP approach does when there are few isolates in the indeterminant range. There are real world examples where this occurs. This scenario starts on page 130.

For scenario six we assess our BNP approach for a small sample of 500 isolates. The previous scenarios considered a sample of 1000 isolates. We compare the results from using a much smaller sample size, 500, to the results using 1000 isolates. This scenario starts on page 130.

Finally, we present results for several real data sets using all approaches. This begins on page 141.

3.3 Simulation Results

3.3.1 BNP vs ERB

Condition	
MIC Density	right skewed unimodal
MIC/DIA Relationship	all five
σ_m, σ_d	0.707, 2.121
MIC Breakpoints	-1, 1
Number of Pathogen Strains	1000
Assessment	DIA Breakpoint Accuracy

Table 3.1 Scenario Conditions and Assessments

To compare our model-based approach to the ERB method we must make sure both methods are aiming at the same target. The ERB method attempts to minimize the observed discrepancies, or equivalently, maximize the observed classification agreements (subject to weights). Our loss function (3.4) doesn't directly aim at this target. Therefore for this scenario we use a different loss function that more closely aligns to the target of the ERB method.

Assuming the MIC breakpoints to be correct, for every MIC value m we calculate the probability the two tests return the correct result (both susceptible, both indeterminant, or both resistant), $p_c(.)$. This probability is calculated as:

$$p_c(m) = \begin{cases} \Phi(\frac{M_L-m}{\sigma_m}) \times (1 - \Phi(\frac{D_U-.5-g(m)}{\sigma_d})) & m \leq M_L^* \\ \left(\Phi(\frac{M_U-1-m}{\sigma_m}) - \Phi(\frac{M_L-m}{\sigma_m})\right) \times \left(\Phi(\frac{D_U-.5-g(m)}{\sigma_d}) - \Phi(\frac{D_L+.5-g(m)}{\sigma_d})\right) & M_L^* < m < M_U^* \\ (1 - \Phi(\frac{M_U-1-m}{\sigma_m})) \times \Phi(\frac{D_L+.5-g(m)}{\sigma_d}) & m \geq M_U^* \end{cases}$$

where m represents the current MIC value, M_L is the lower MIC test breakpoint, M_L^* is the true lower MIC breakpoints, M_U is the upper MIC test breakpoint, M_U^* is the true upper MIC breakpoint, D_U is the upper DIA breakpoint, D_L is the lower DIA breakpoint, σ_m is the MIC measurement error, σ_d is the DIA measurement error, and $g(m)$ is the function the links the true MIC to the true DIA.

We estimate the DIA breakpoints by finding the set that maximize the weighted probability of correct classification. This can be expressed as

$$P_C = 1/N \int_{-\infty}^{\infty} p_c(u) \times w(u) \, du \tag{3.4}$$

where $w(.)$ are the weights and the term $1/N$ normalizes the probability. For this scenario we use the underlying MIC distribution as the weights.

The true DIA breakpoints and correct classification probabilities for the five MIC/DIA relationships are shown in Table 3.2.

Relationship	True DIA Breakpoints	Correct Classification Probability
Relationship 1	27, 32	0.803
Relationship 2	18, 23	0.816
Relationship 3	22, 28	0.869
Relationship 4	26, 31	0.843
Relationship 5	16, 22	0.870

Table 3.2 True DIA Breakpoints based on Probability of Correct Classification

Breakpoint performance between the BNP approach and ERB are shown in Figures 3.11 - 3.15. The BNP results are on the right while the ERB results are on the left. Note on some of the ERB figures many of the estimated DIA breakpoints are outside the graph range.
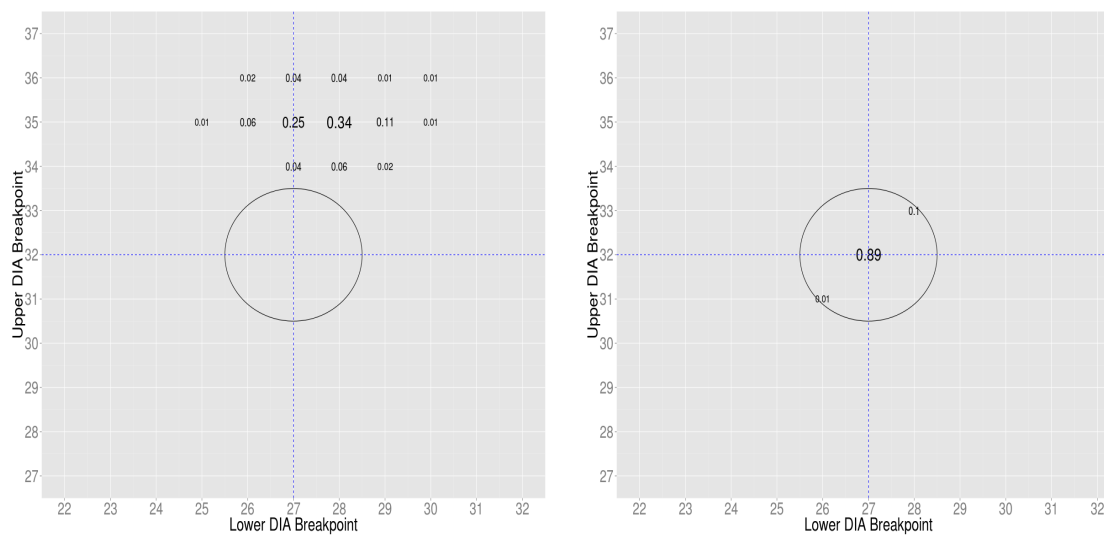
Figure 3.11. ERB and BNP Breakpoints for MIC/DIA Relationship 1
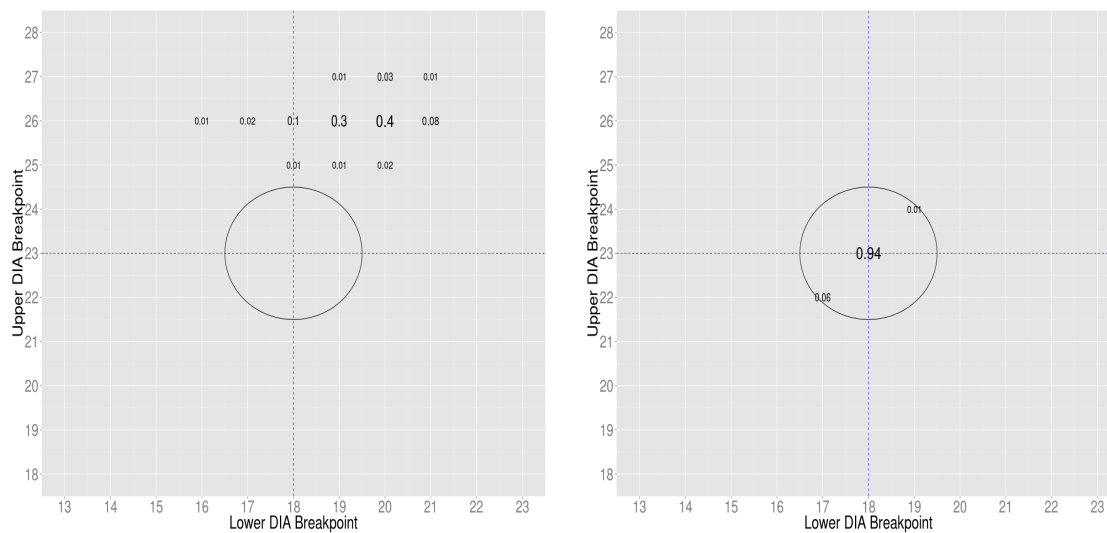
Figure 3.12. ERB and BNP Breakpoints for MIC/DIA Relationship 2
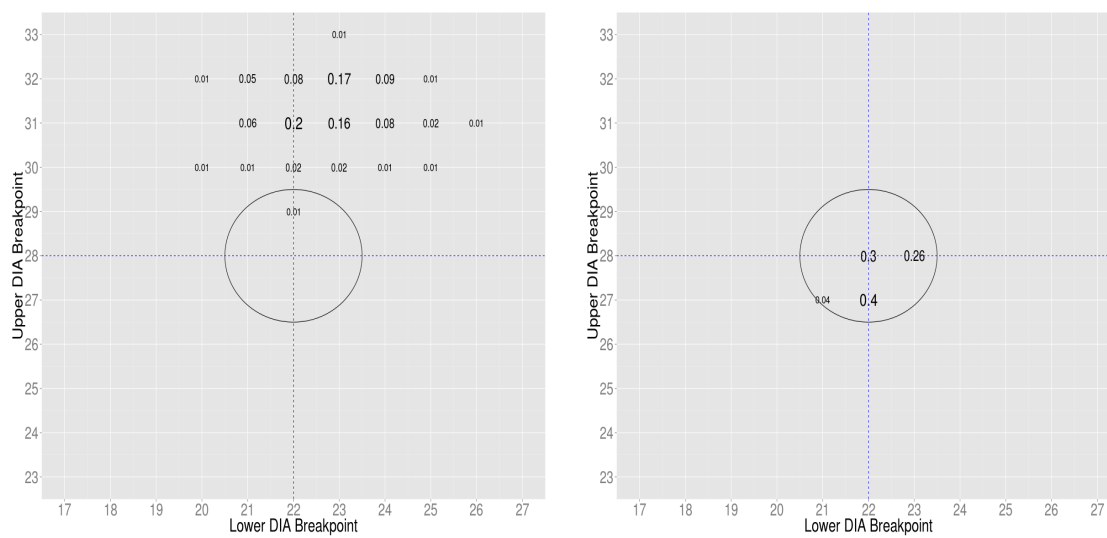
Figure 3.13. ERB and BNP Breakpoints for MIC/DIA Relationship 3
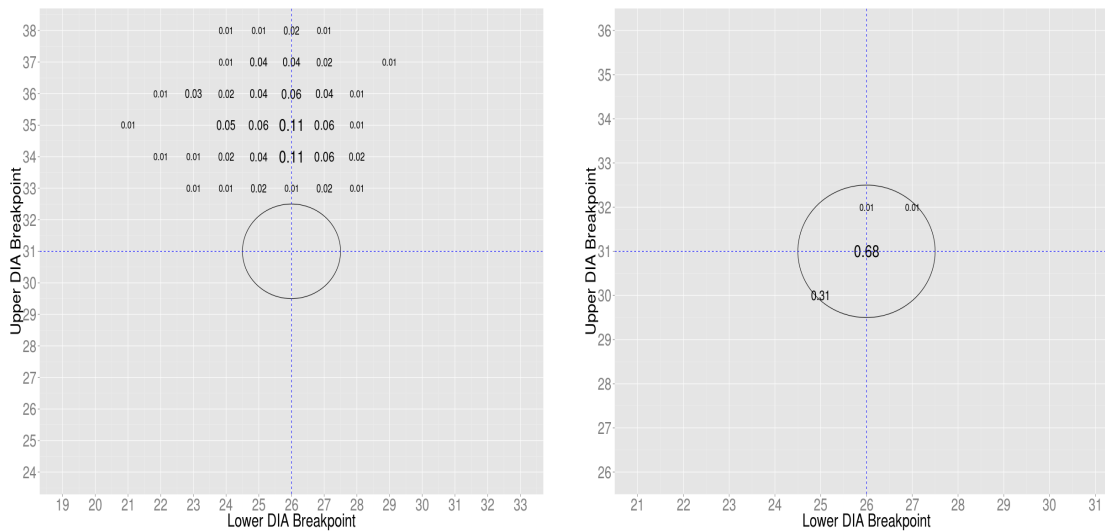
Figure 3.14. ERB and BNP Breakpoints for MIC/DIA Relationship 4
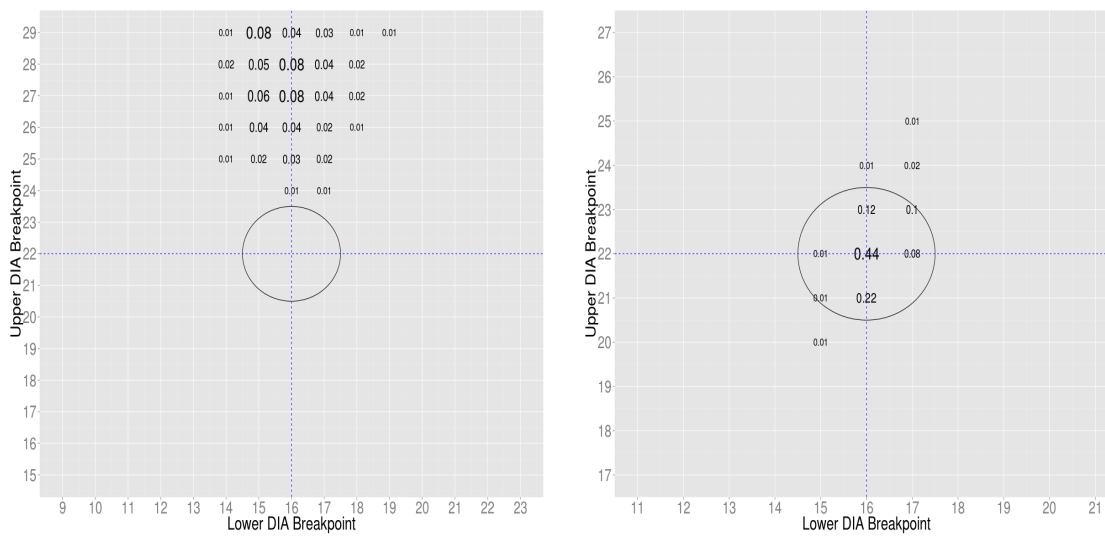
Figure 3.15. ERB and BNP Breakpoints for MIC/DIA Relationship 5

We see the BNP approach is always more precise and less biased. The BNP approach was able to estimate the exact true DIA breakpoints with 90% accuracy for relationships one and two, and was within ± 1 of the true DIA breakpoints almost 100% of the time for all five relationships. The resulting DIA breakpoints for the ERB approach were quite variable and rarely within ± 1 of the true breakpoints.

BNP vs ERB - Increase Measurement Error

We now investigate performance for the BNP and ERB approaches after increasing the measurement error 50%: $\sigma_m = 1.06$ and $\sigma_d = 3.18$. The true DIA breakpoints and correct classification probabilities for the five MIC/DIA relationships are shown in Table 3.3. Notice the probability of correct classification has decreased as the measurement error increased.

Relationship	True DIA Breakpoints	Correct Classification Probability
Relationship 1	25, 30	0.759
Relationship 2	16, 21	0.770
Relationship 3	21, 26	0.832
Relationship 4	26, 31	0.806
Relationship 5	17, 22	0.835

Table 3.3 True DIA Breakpoints based on Probability of Correct Classification for Increased Measurement Error

Breakpoint performance between the BNP approach and ERB approach are shown in Figures 3.16 - 3.20. The BNP results are on the right while the ERB results are on the left. Note in most of the ERB figures many of the estimated DIA breakpoints are outside the graph range.
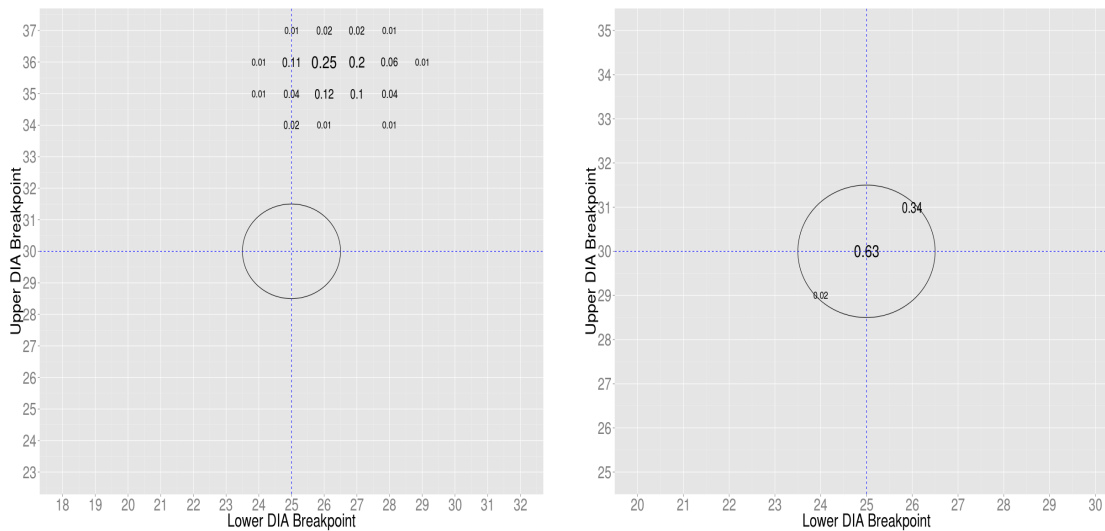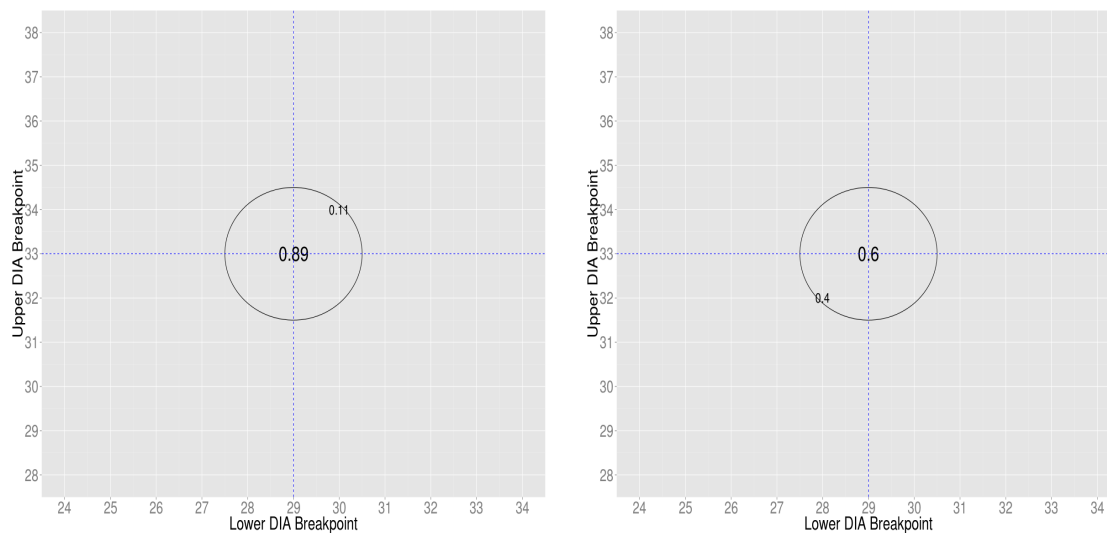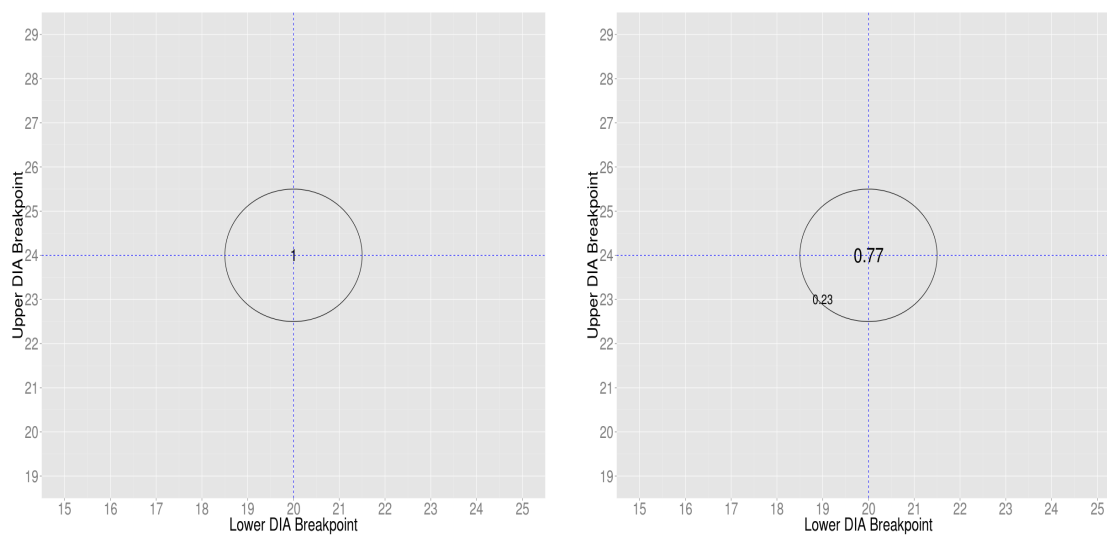
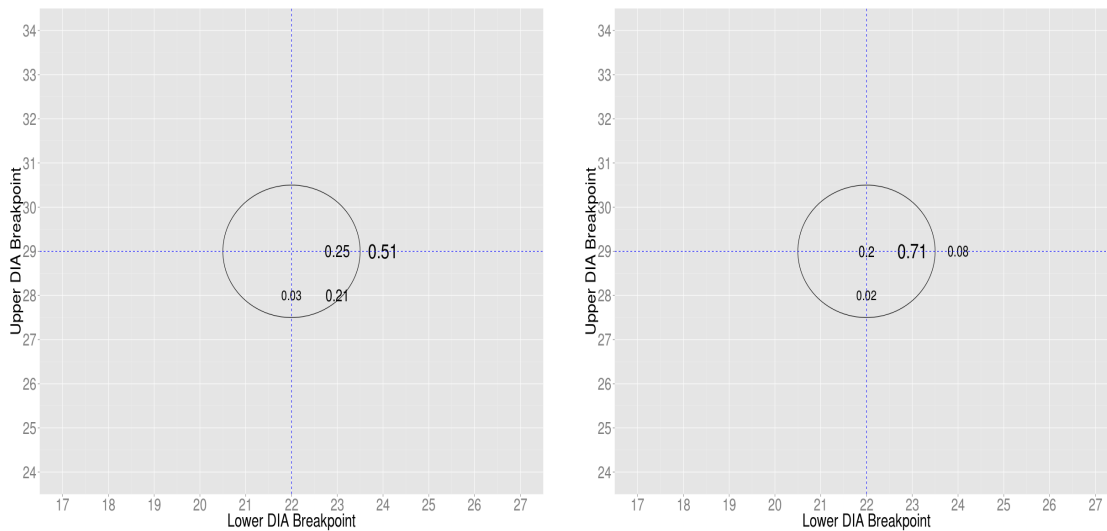Figure 3.16. ERB and BNP Breakpoints for MIC/DIA Relationship 1
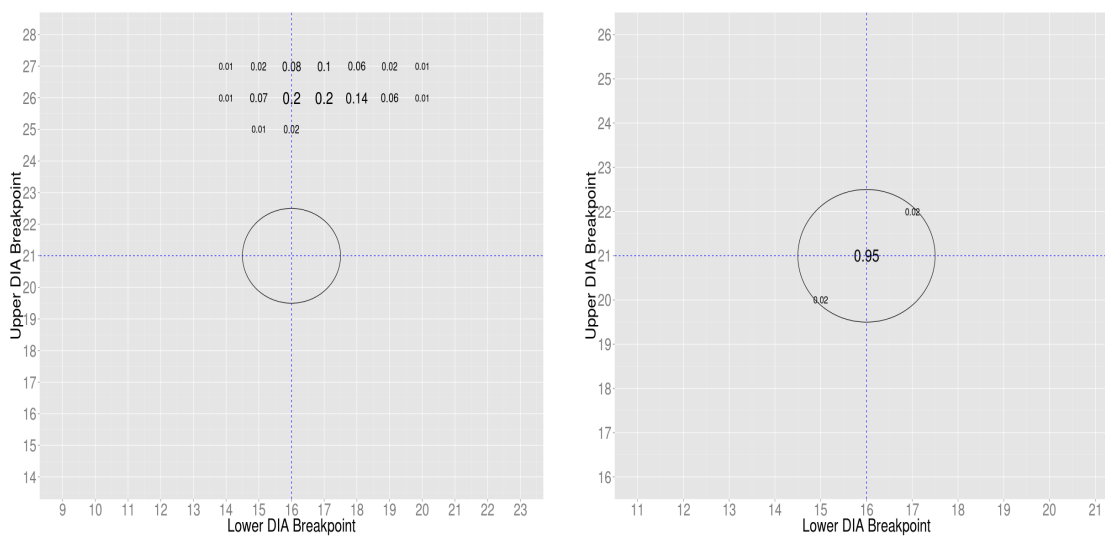
Figure 3.17. ERB and BNP Breakpoints for MIC/DIA Relationship 2
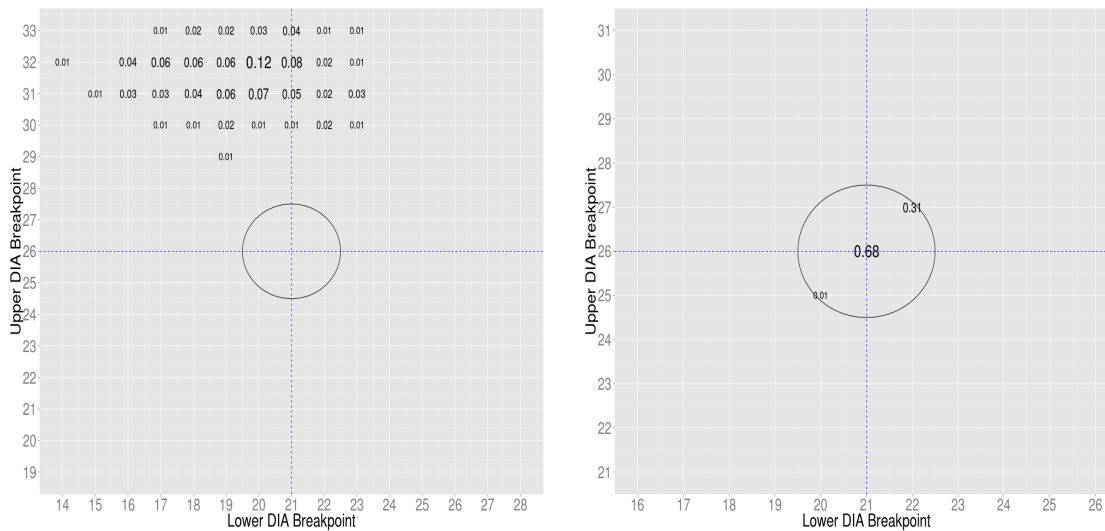
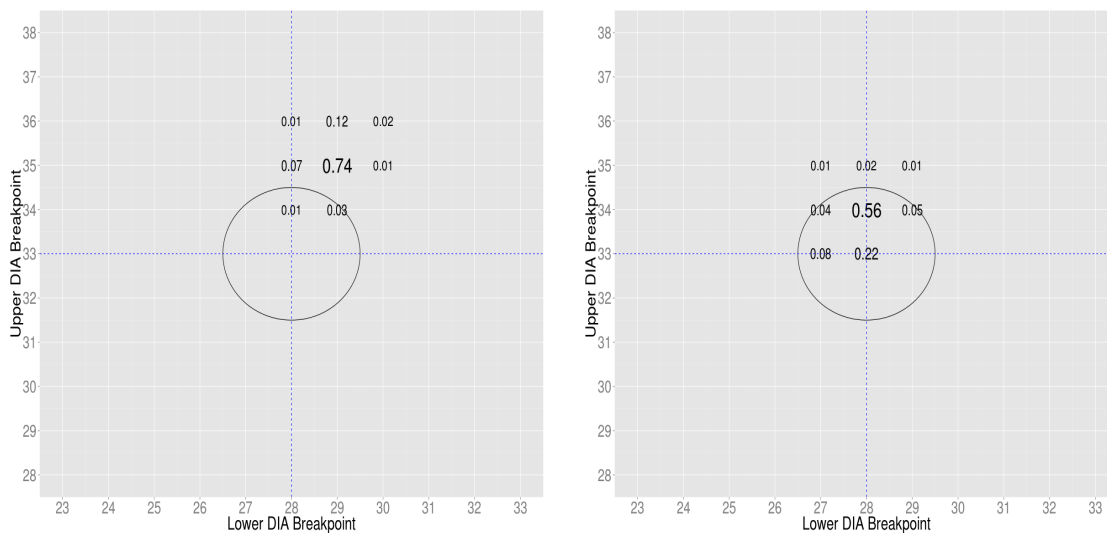Figure 3.18. ERB and BNP Breakpoints for MIC/DIA Relationship 3
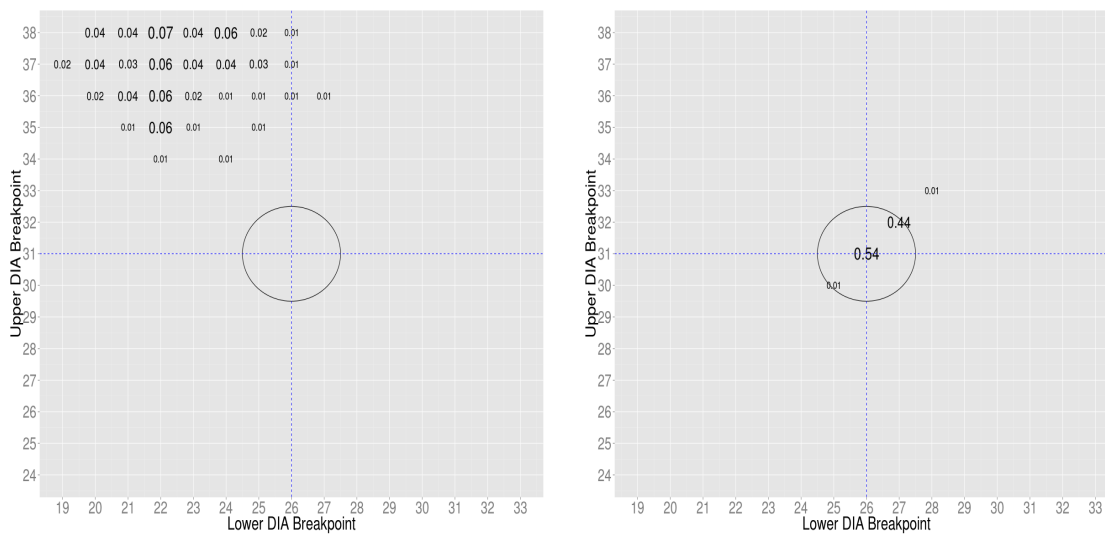
Figure 3.19. ERB and BNP Breakpoints for MIC/DIA Relationship 4
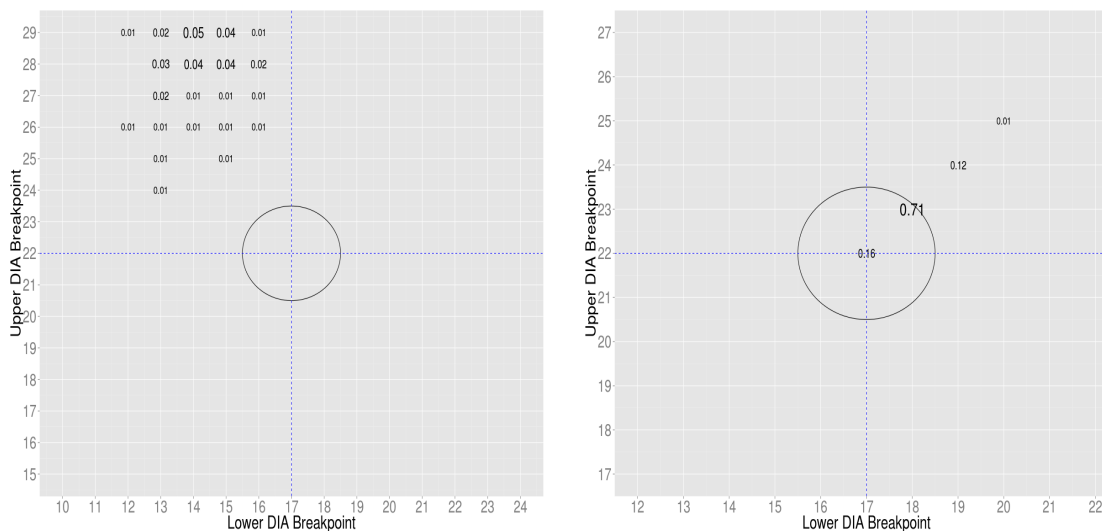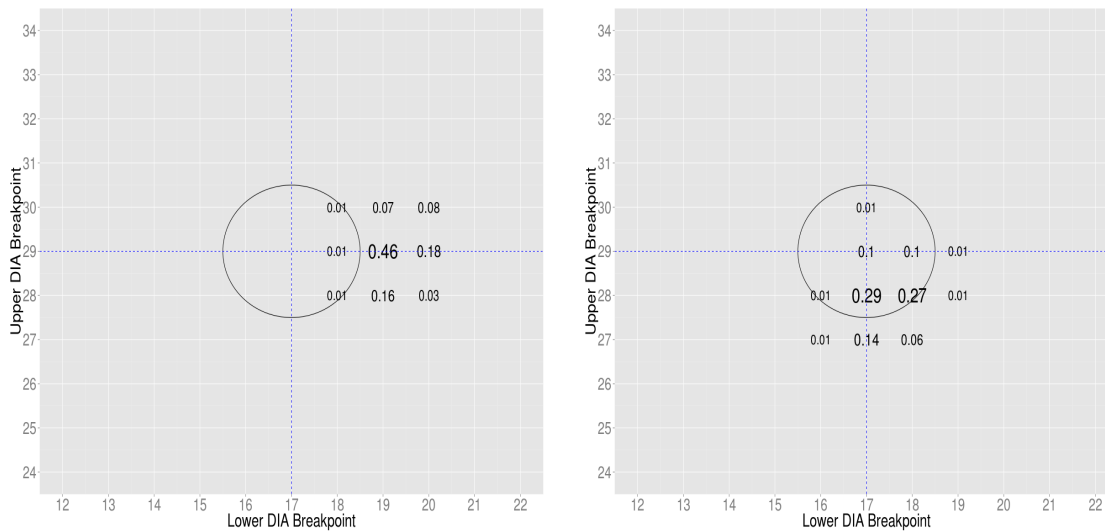
Figure 3.20. ERB and BNP Breakpoints for MIC/DIA Relationship 5

The ERB approach performed considerably worse when we increased the measurement error by 50%, however the BNP approach results remained relatively unchanged. This is a key result, since the BNP approach takes the measurement error into account it is much less likely to be affected by the increase in measurement error.

3.3.2 Comparison: BNP vs FNP

One of our key questions is whether the BNP approach, which models both the MIC density and MIC/DIA relationship simultaneously, does a better job fitting the data compared to the FNP approach, which does not consider the DIA data when estimating the MIC density. If it does not, there is little reason for the added computational time that the BNP approach requires.

To investigate this, we look at DIA breakpoint accuracy, MIC/DIA relationship accuracy, and MIC density accuracy for both approaches (condition I). We then increase the measurement error by 50% ($\sigma_m = 1.06$ and $\sigma_d = 3.18$) and assess the two approaches again (condition II).

Finally, we look at how both approaches estimate DIA breakpoint for different MIC breakpoints. In addition to the MIC breakpoints: (-1, 1), we investigate performance for the following MIC breakpoints: (-3, -1), (1, 3), (-4, -2), and (4, 2).

Condition I	
MIC Density	right skewed
MIC/DIA Relationship	all five
σ_m, σ_d	0.707, 2.121
MIC Breakpoints	-1, 1
Number of Pathogen Strains	1000
Assessment	DIA Breakpoint Accuracy MIC/DIA Relationship Accuracy MIC Density Accuracy

Table 3.4 Scenario Conditions and Assessments

Condition II - Increased Error	
MIC Density	three component symmetric, right skewed two component, skewed, right skewed
MIC/DIA Relationship	three-parameter logistic
σ_m, σ_d	1.06, 3.18
MIC Breakpoints	NA
Number of Pathogen Strains	1000
Assessment	DIA Breakpoint Accuracy MIC/DIA Relationship Accuracy MIC Density Accuracy

Table 3.5 Scenario Conditions and Assessments

Condition I
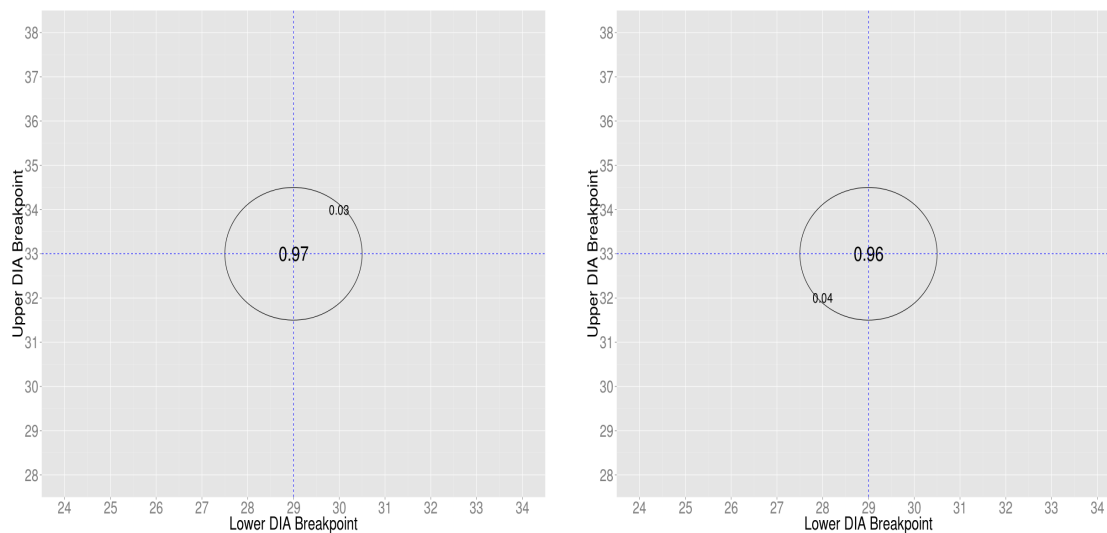
Figure 3.21. FNP and BNP Breakpoints for MIC/DIA Relationship 1
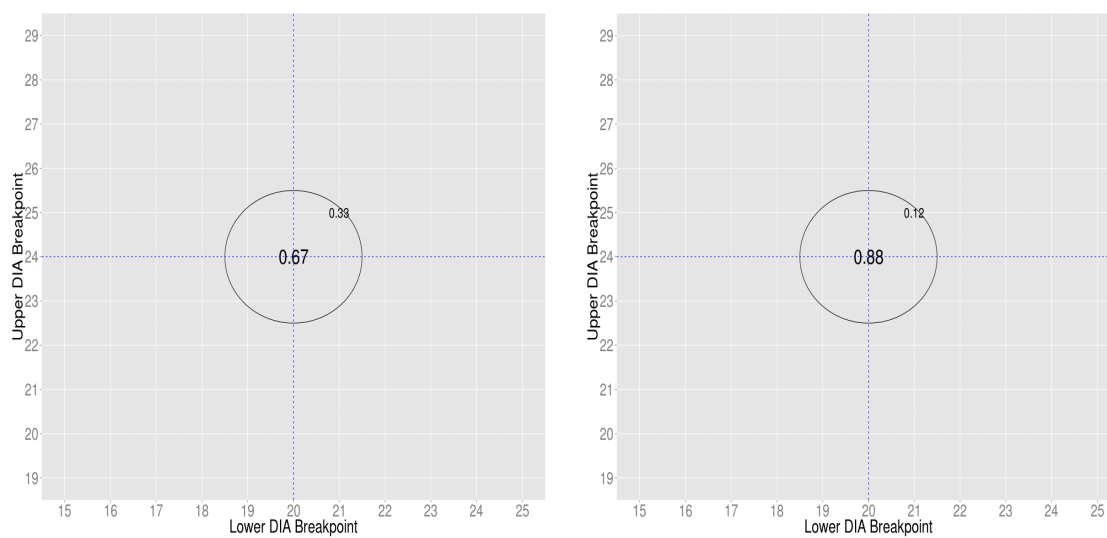
Figure 3.22. FNP and BNP Breakpoints for MIC/DIA Relationship 2
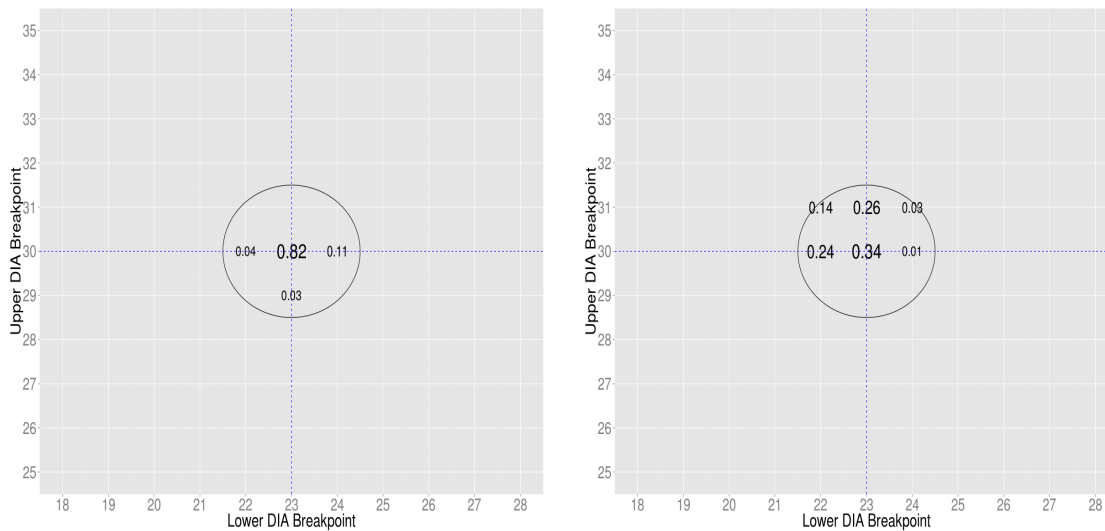
Figure 3.23. FNP and BNP Breakpoints for MIC/DIA Relationship 3
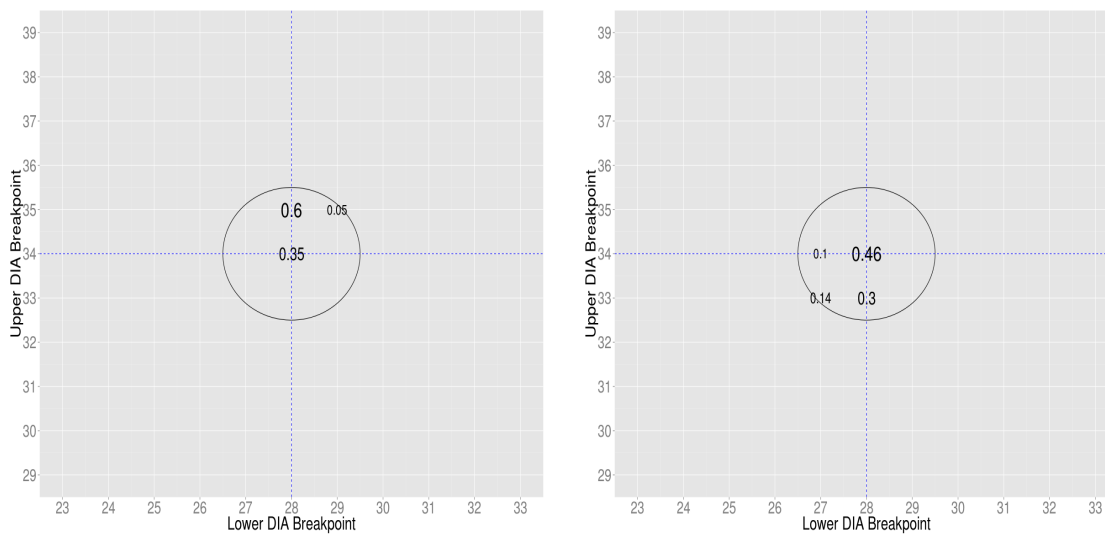
Figure 3.24. FNP and BNP Breakpoints for MIC/DIA Relationship 4
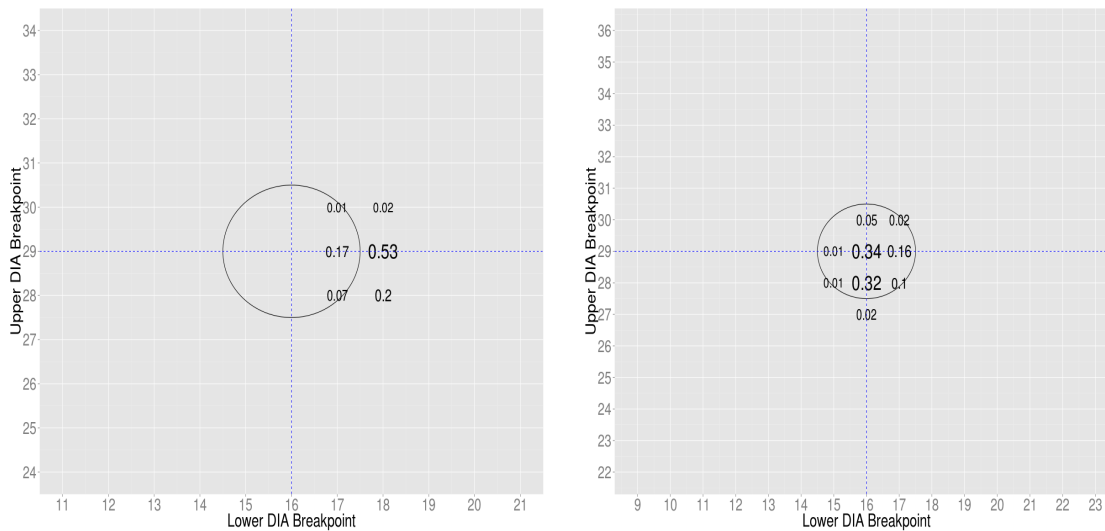
Figure 3.25. FNP and BNP Breakpoints for MIC/DIA Relationship 5
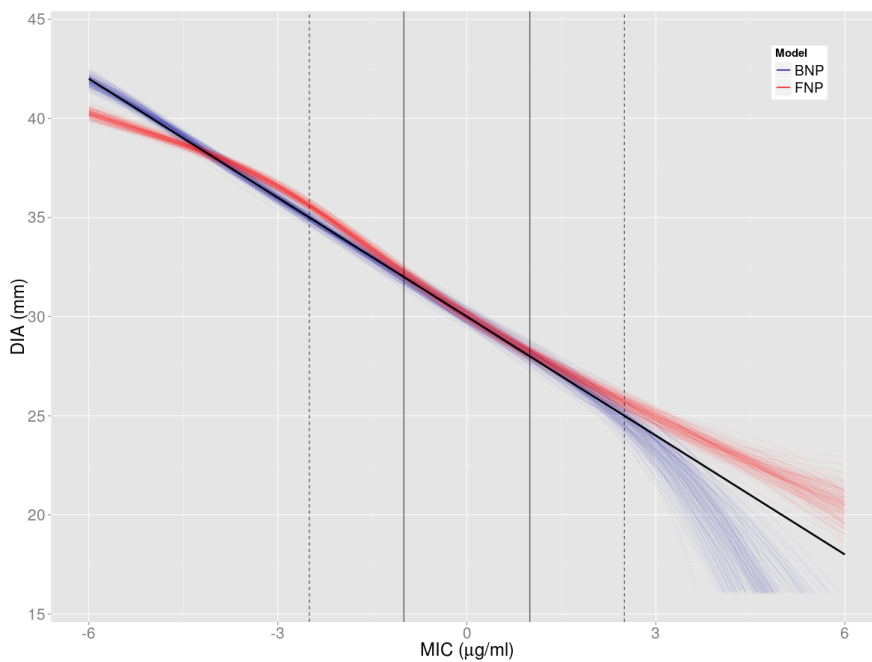
Figure 3.26. Median MIC/DIA Fits for Relationship 1
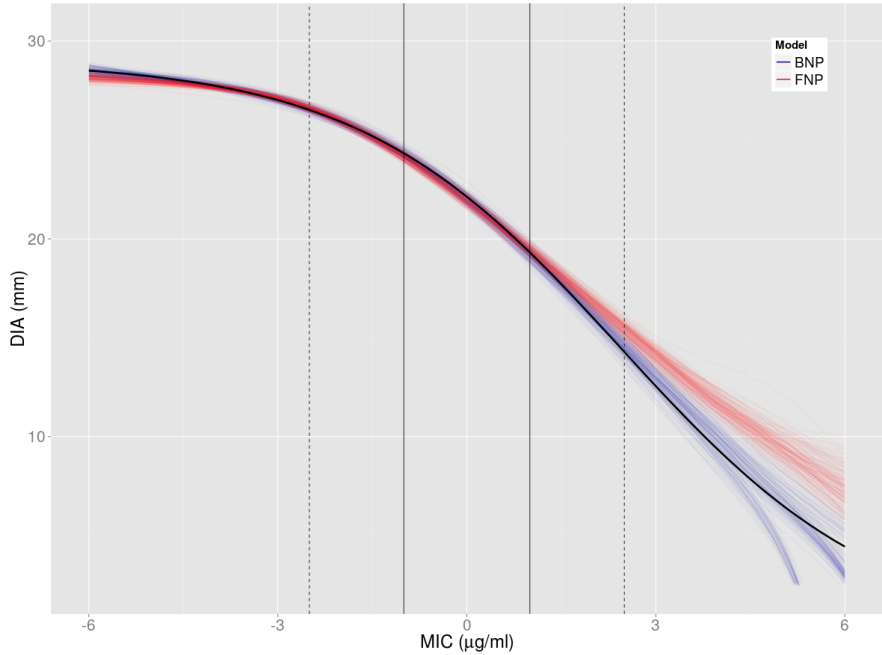
Figure 3.27. Median MIC/DIA Fits for Relationship 2
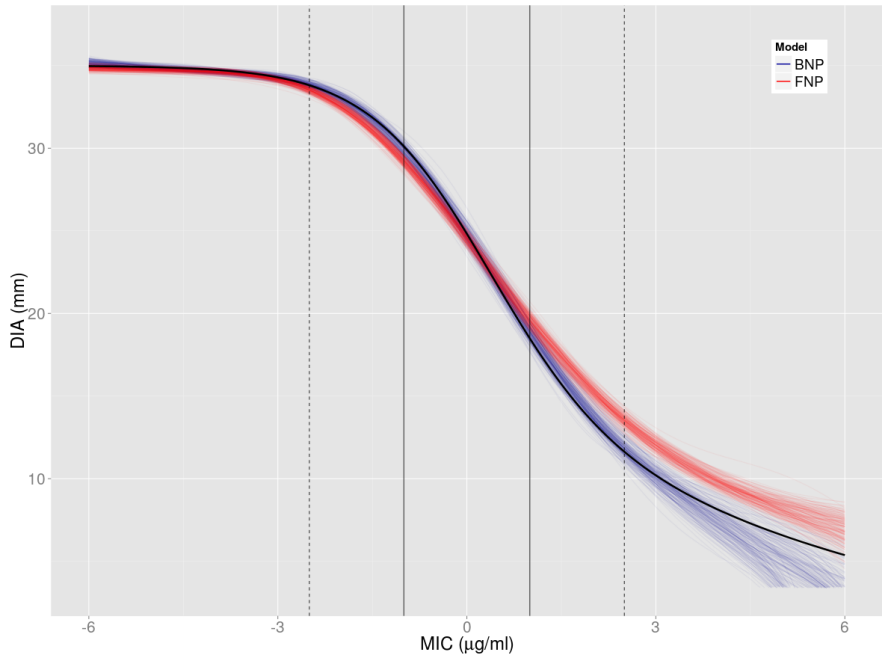
Figure 3.28. Median MIC/DIA Fits for Relationship 3
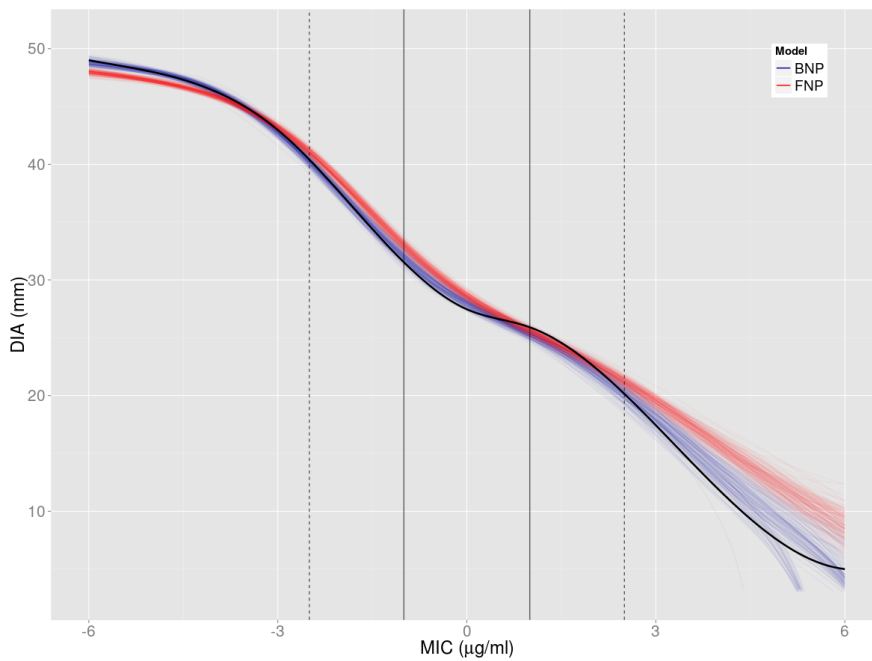
Figure 3.29. Median MIC/DIA Fits for Relationship 4
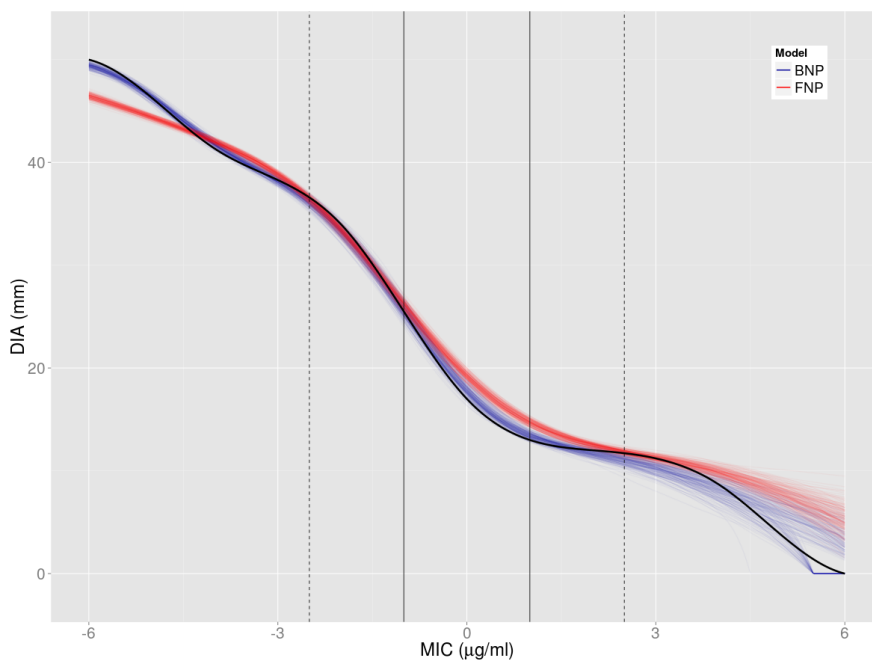
Figure 3.30. Median MIC/DIA Fits for Relationship 5

Relationship	Model	Mean	SD	Median
1	BNP	15.38	12.98	11.09
	FNP	33.98	15.71	31.21
2	BNP	14.10	13.09	9.31
	FNP	40.66	20.95	37.85
3	BNP	42.90	38.97	29.69
	FNP	104.2	43.56	95.72
4	BNP	66.36	45.67	52.98
	FNP	216.40	74.54	208.20
5	BNP	97.72	61.76	84.00
	FNP	391.10	138.40	373.30

Table 3.6 MIC/DIA relationship fit statistics for the BNP and FNP approaches when $\sigma_m = 0.707$. The BNP approach greatly outperforms the FNP approach for each of the MIC/DIA relationships.
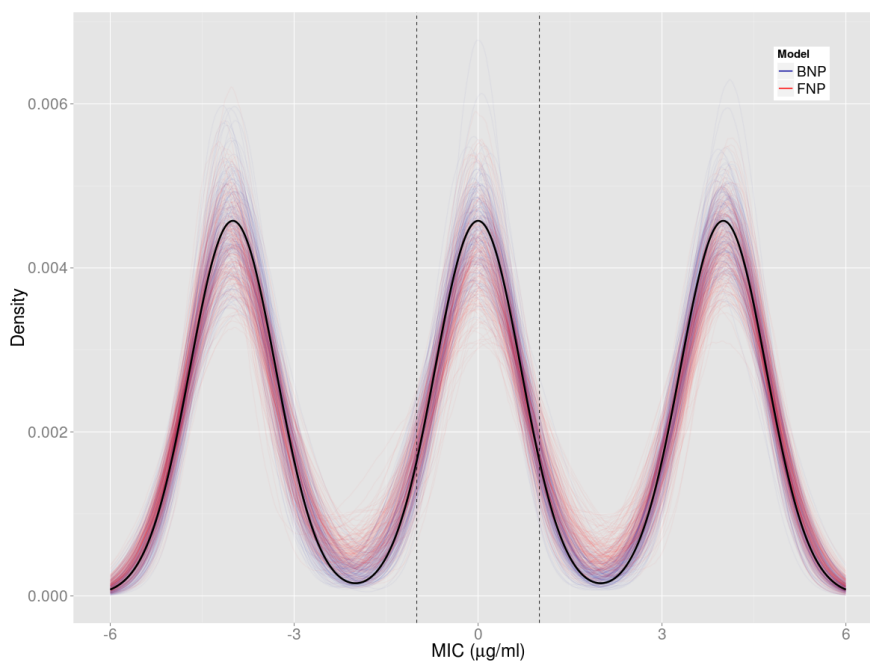
Figure 3.31. Median Density Fits for MIC Density 1
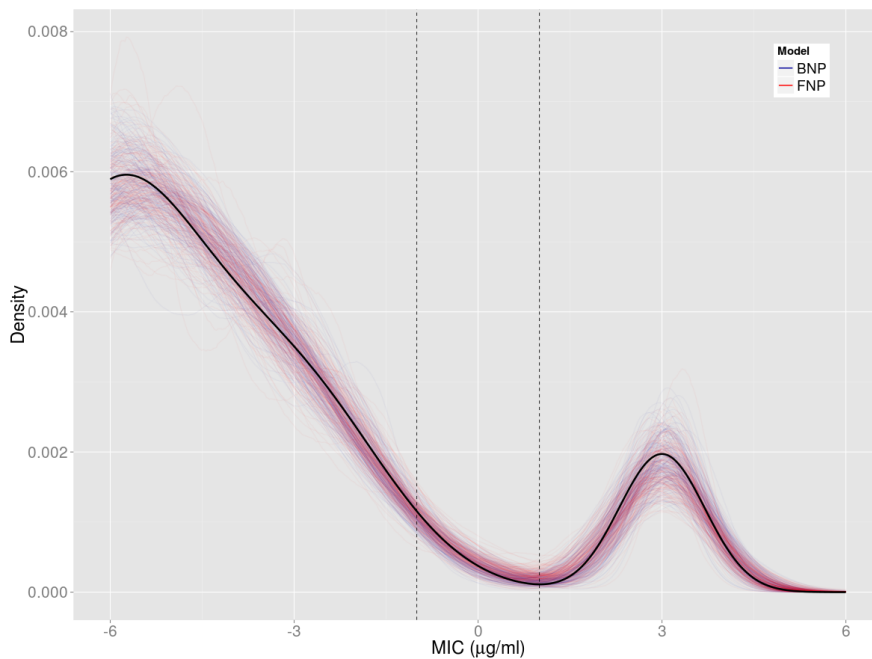
Figure 3.32. Median Density Fits for MIC Density 2
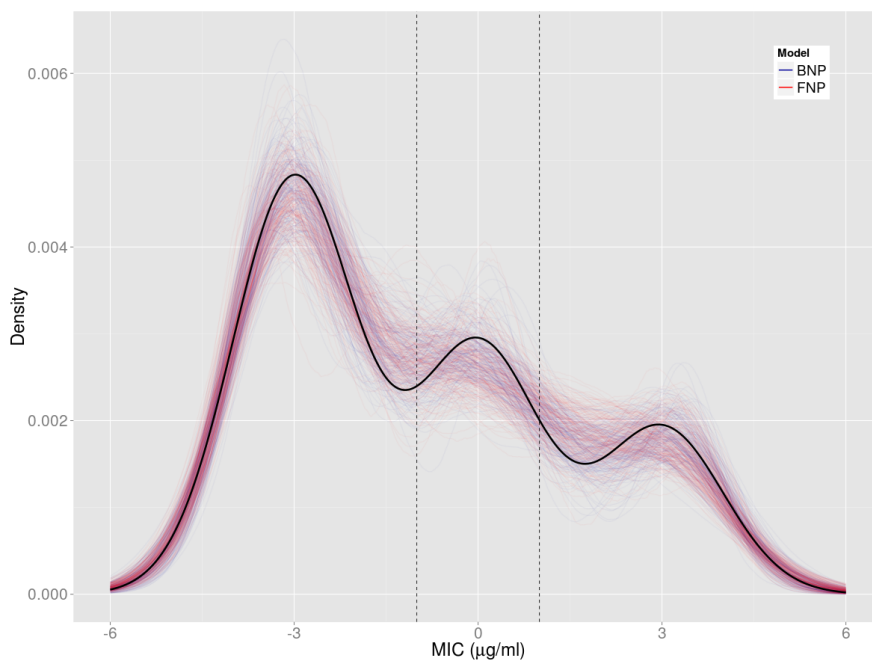
Figure 3.33. Median Density Fits for MIC Density 3
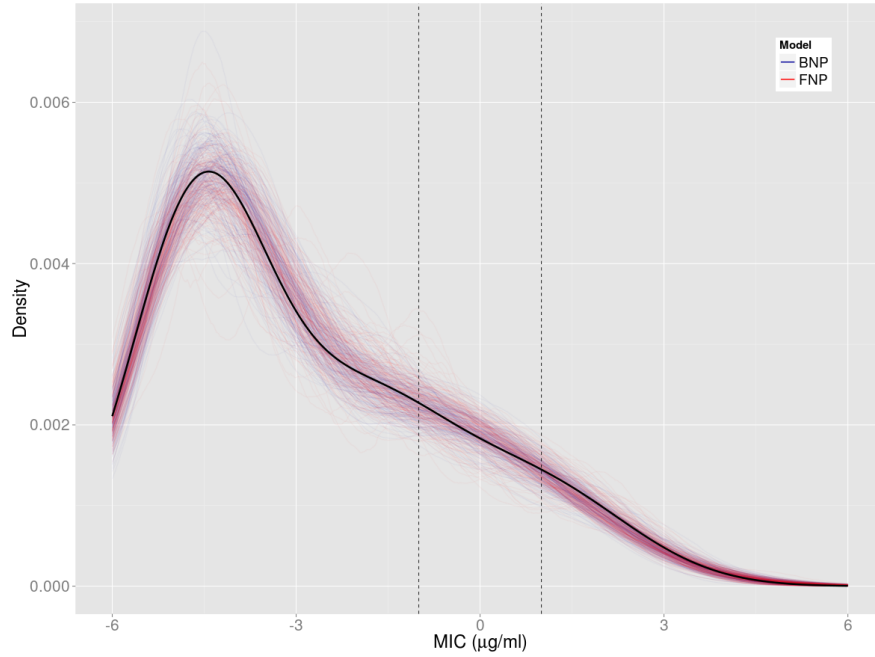
Figure 3.34. Median Density Fits for MIC Density 4

Relationship	Model	Mean	SD	Median
1	BNP	0.0048	0.0028	0.0043
	FNP	0.0080	0.0042	0.0073
2	BNP	0.0031	0.0021	0.0026
	FNP	0.0035	0.0028	0.0027
3	BNP	0.0042	0.0024	0.0038
	FNP	0.0052	0.0032	0.0042
4	BNP	0.0026	0.0021	0.0020
	FNP	0.0030	0.0026	0.0021

Table 3.7 Density fit statistics for the FNP and BNP approach when $\sigma_m = 0.707$. For each MIC density relationship, the BNP approach outperforms the FNP approach, however the difference is small.

Condition II - Increase Measurement Error 50%

Figure 3.35. FNP and BNP Breakpoints for MIC/DIA Relationship 1
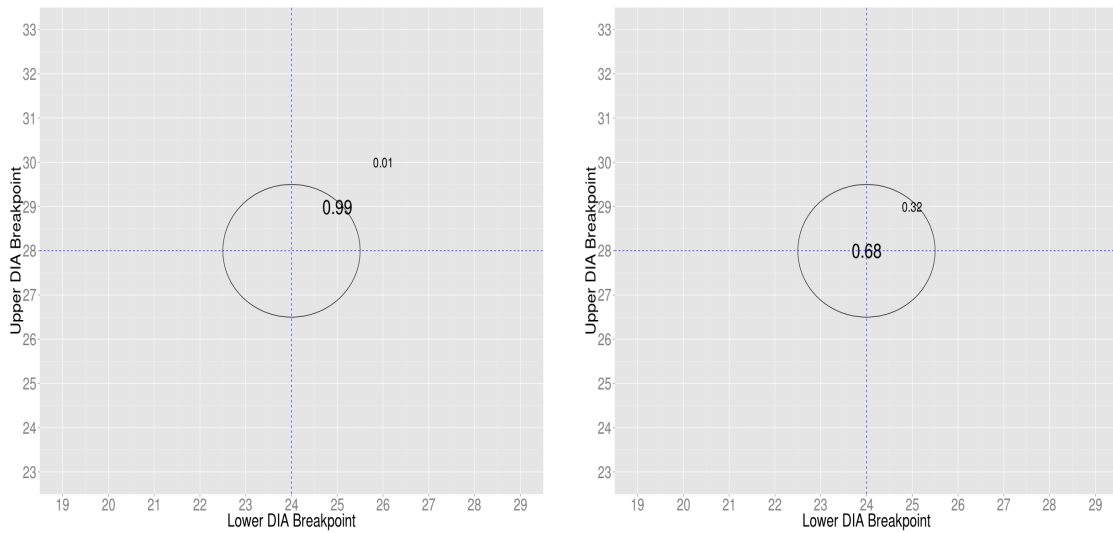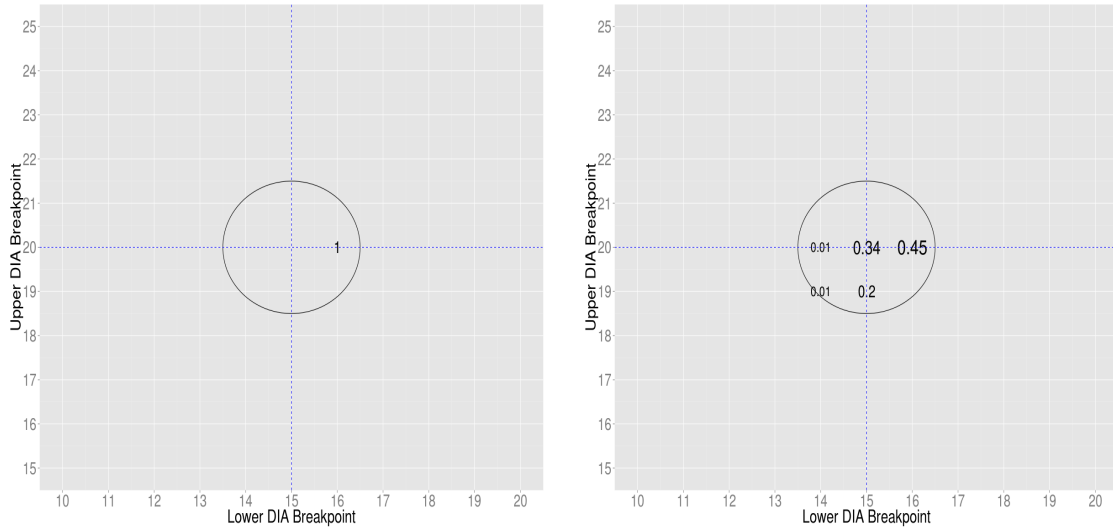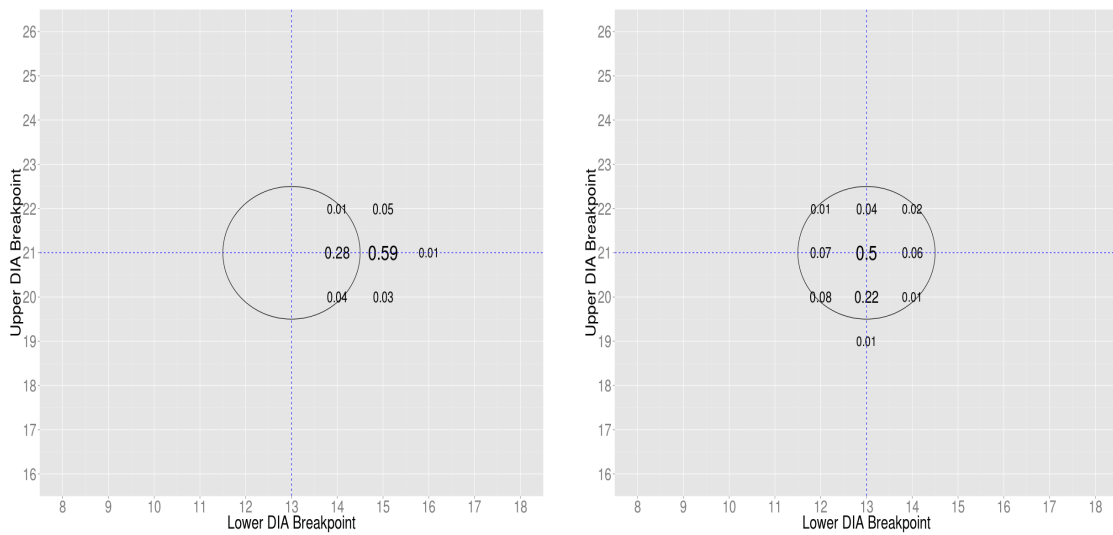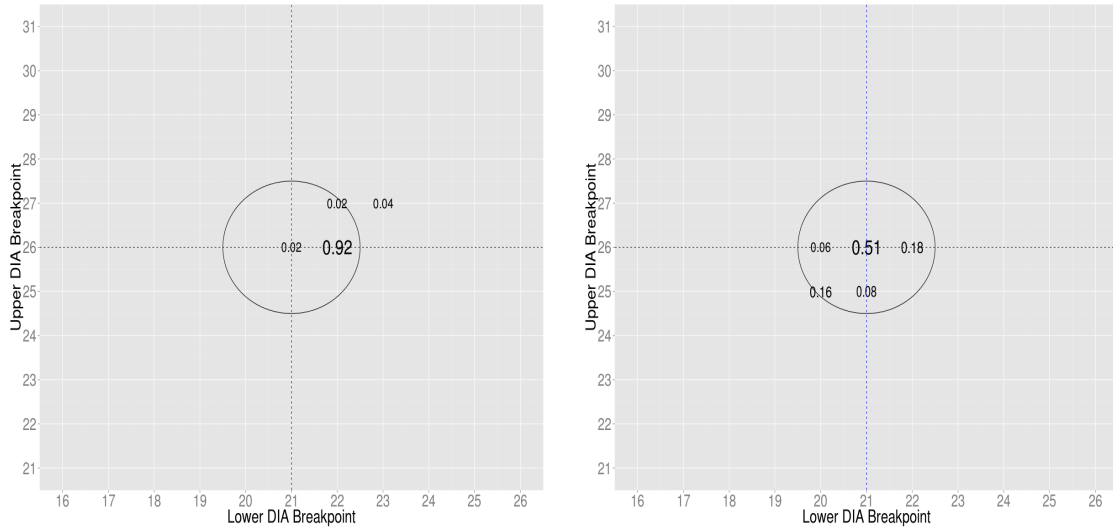

Figure 3.36. FNP and BNP Breakpoints for MIC/DIA Relationship 2

Figure 3.37. FNP and BNP Breakpoints for MIC/DIA Relationship 3

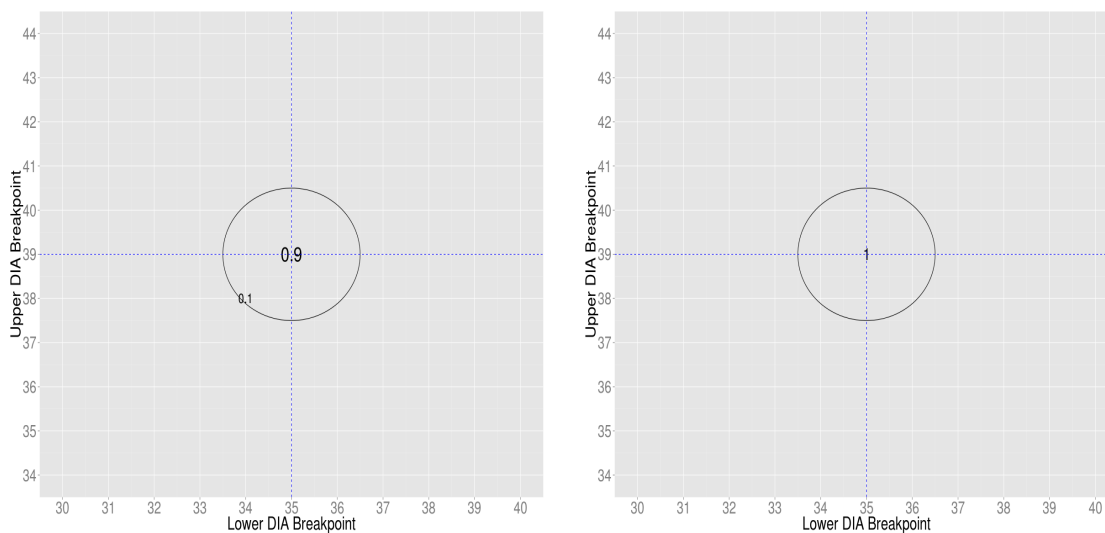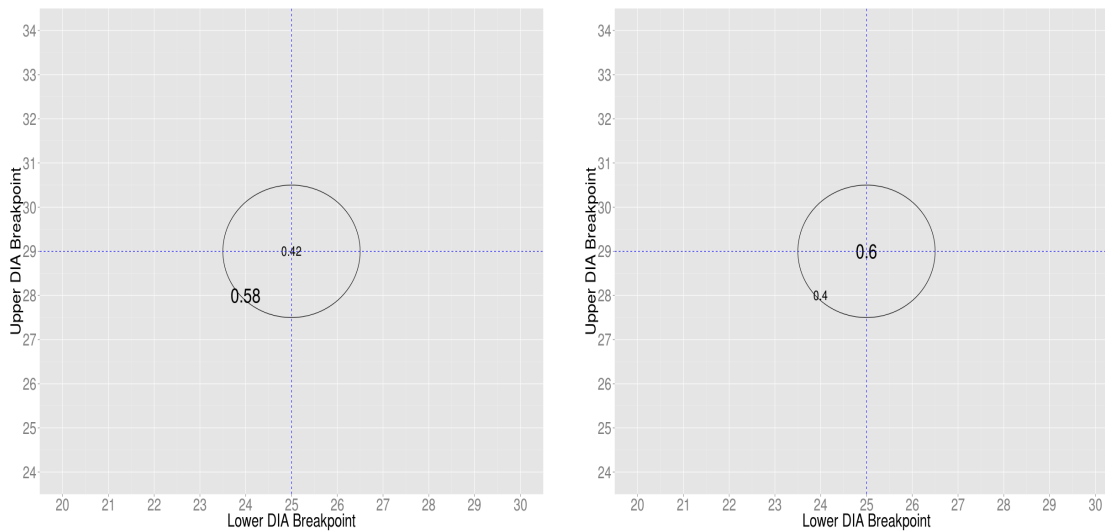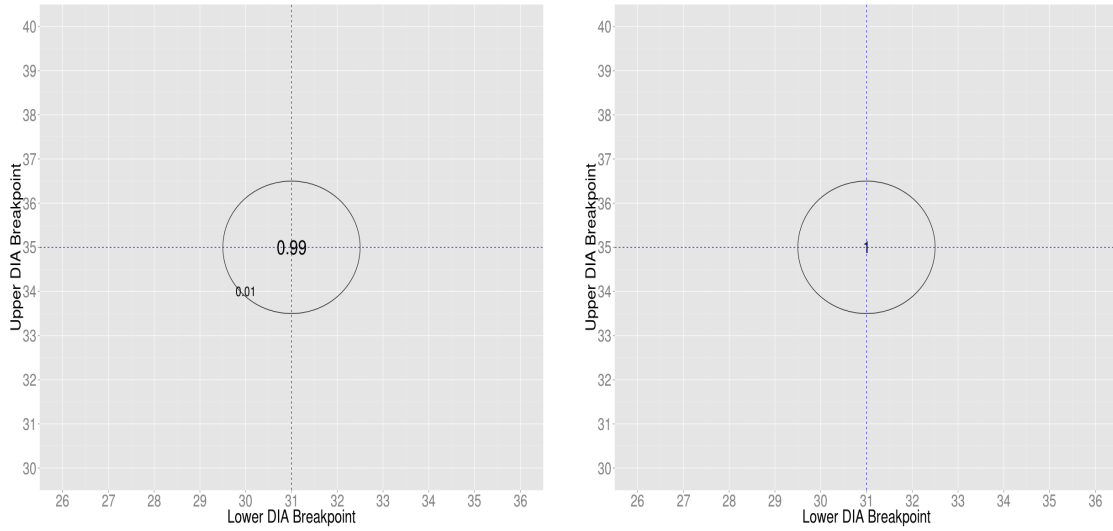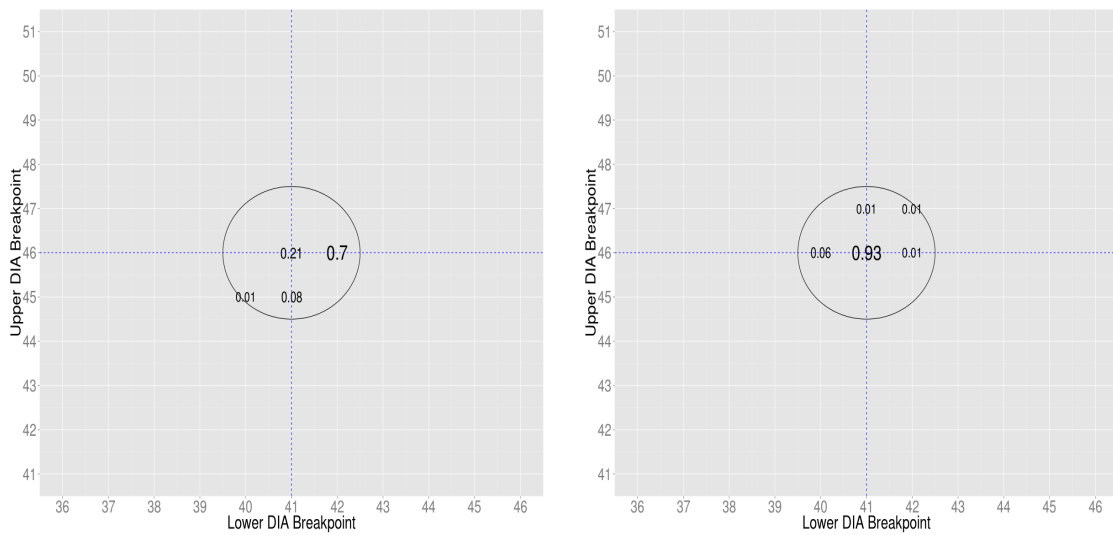Figure 3.38. FNP and BNP Breakpoints for MIC/DIA Relationship 4

Figure 3.39. FNP and BNP Breakpoints for MIC/DIA Relationship 5
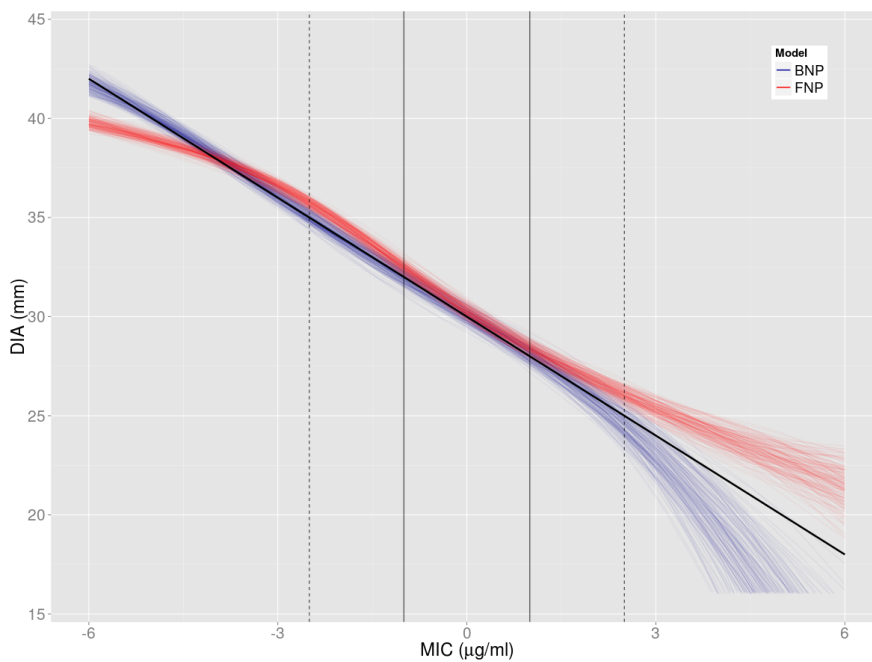
Figure 3.40. Median MIC/DIA Fits for Relationship 1
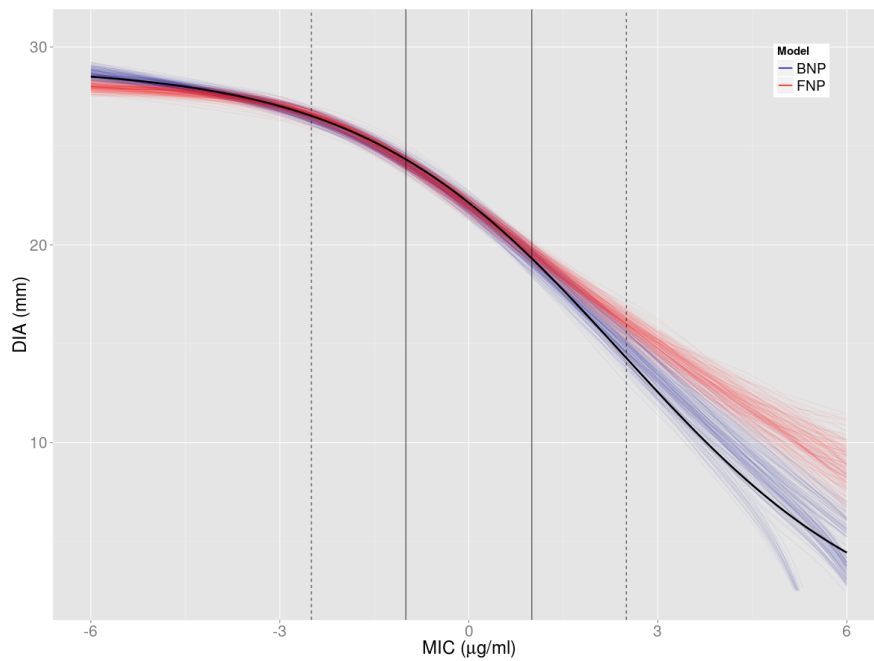
Figure 3.41. Median MIC/DIA Fits for Relationship 2
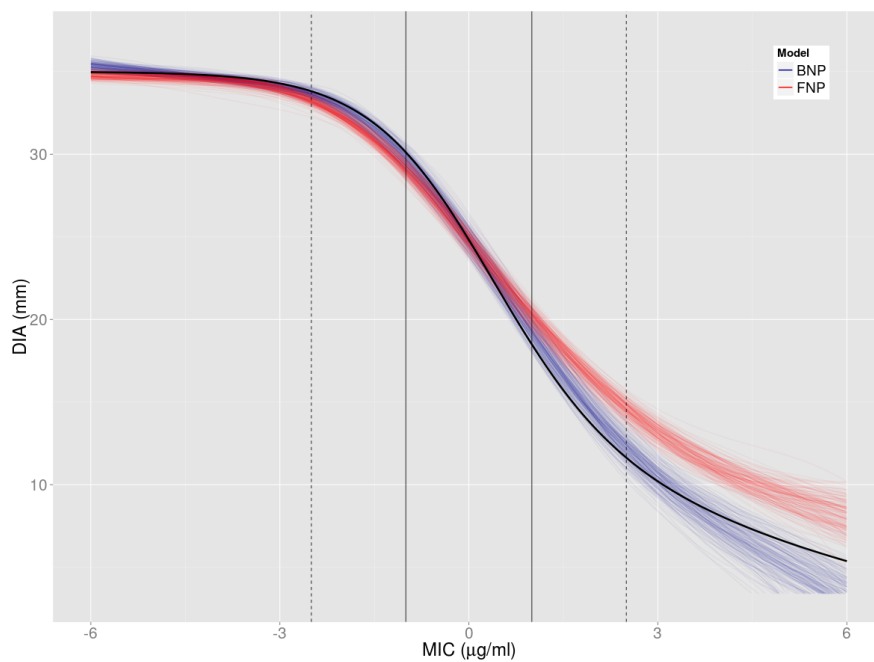
Figure 3.42. Median MIC/DIA Fits for Relationship 3
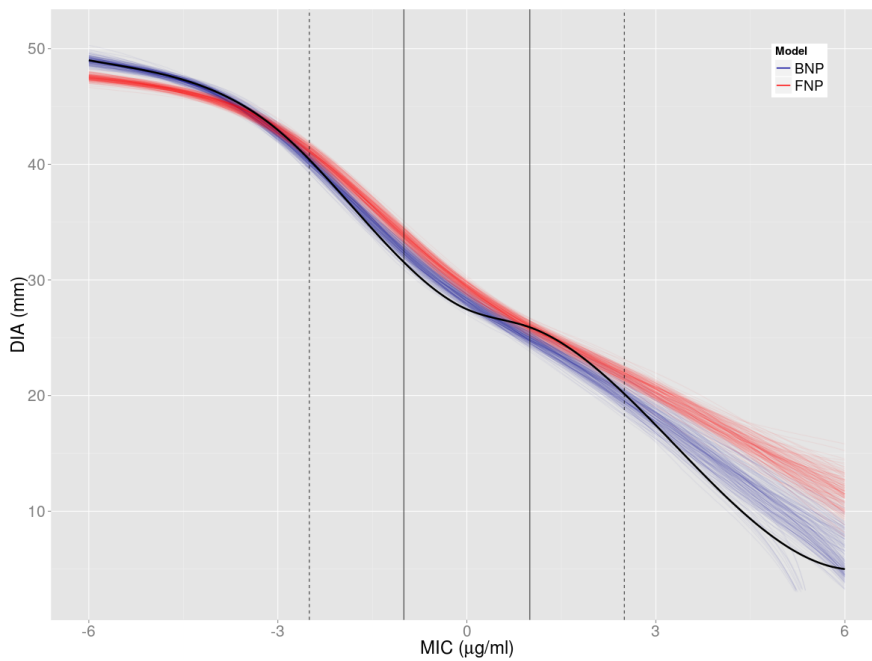
Figure 3.43. Median MIC/DIA Fits for Relationship 4
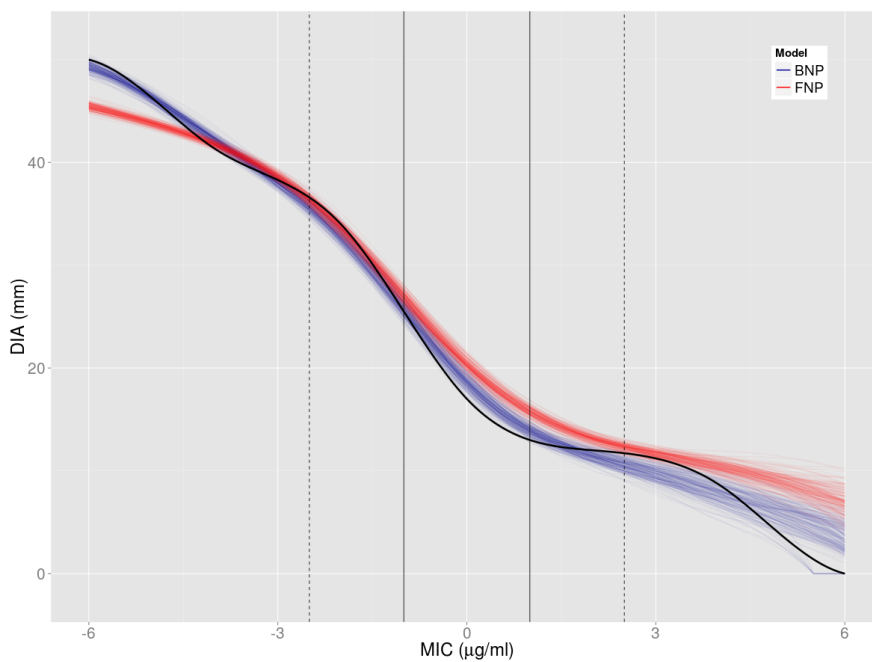
Figure 3.44. Median MIC/DIA Fits for Relationship 5

Relationship	Model	Mean	SD	Median
1	BNP	26.72	22.94	19.11
	FNP	85.24	39.08	77.33
2	BNP	34.71	27.35	25.71
	FNP	89.75	43.66	81.91
3	BNP	120.1	102.7	89.9
	FNP	563.1	147.4	560.8
4	BNP	175	92	162
	FNP	520	166	497
5	BNP	278	140	258
	FNP	981	278	949

Table 3.8 MIC/DIA relationship fit statistics for the BNP and FNP approaches when measurement error is increased by 50%. The BNP approach greatly outperforms the FNP approach for each of the MIC/DIA relationships.
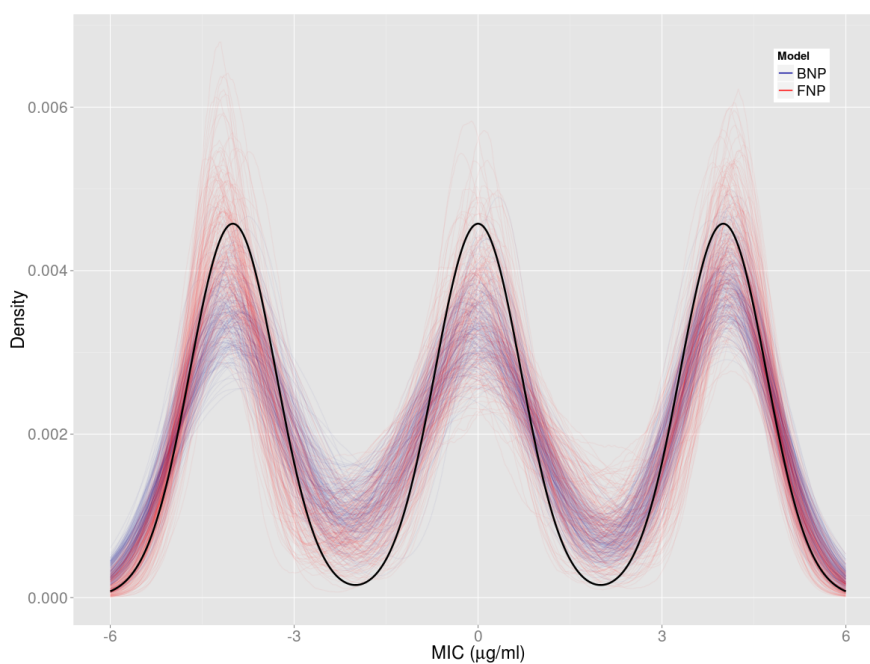
Figure 3.45. Median Density Fits for Relationship 1
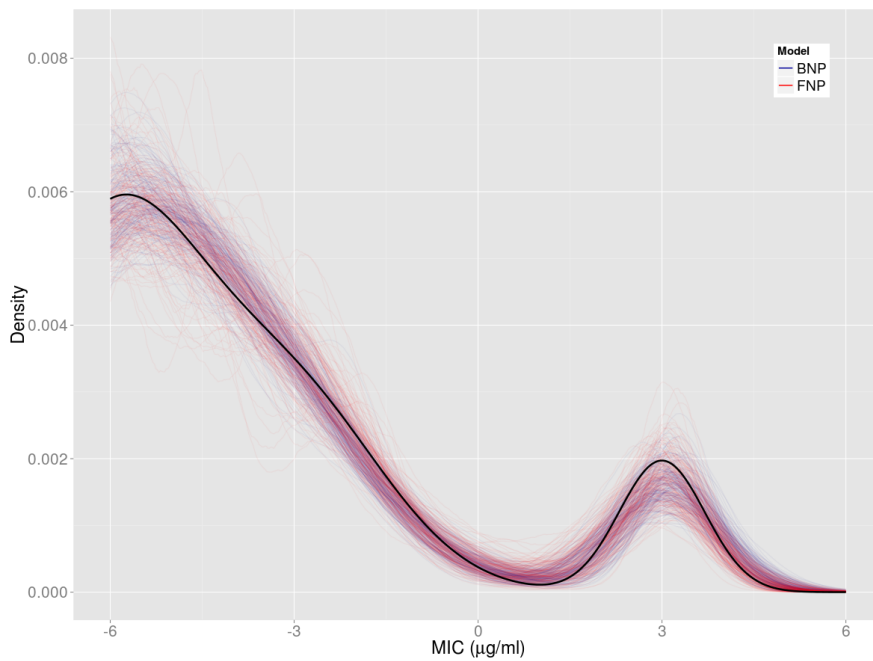
Figure 3.46. Median Density Fits for Relationship 2
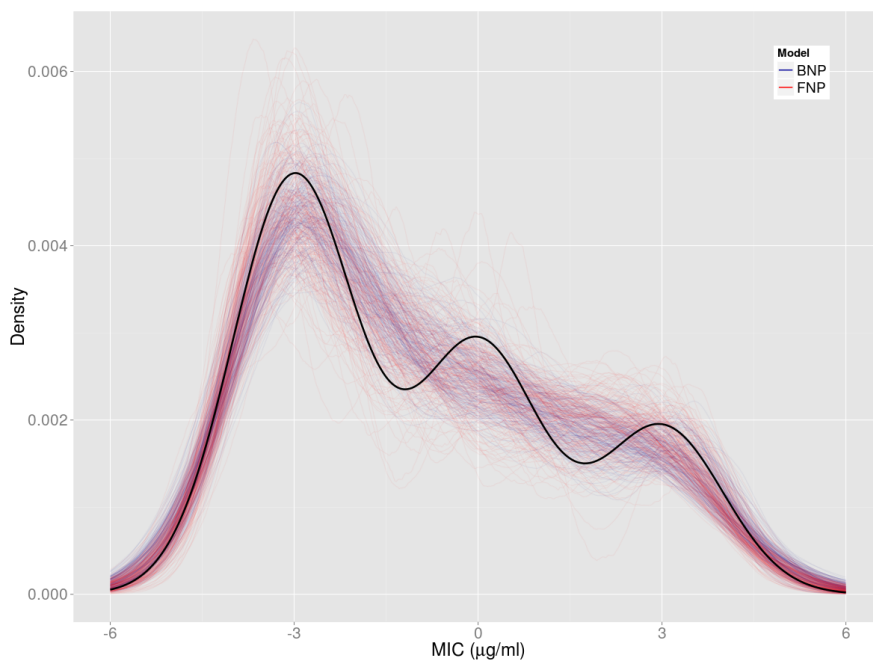
Figure 3.47. Median Density Fits for Relationship 3
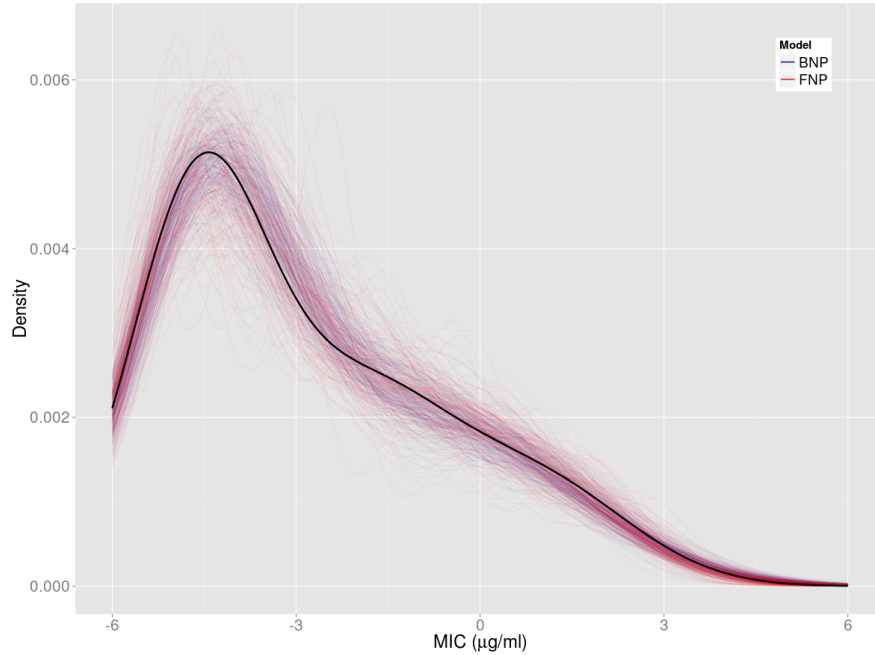
Figure 3.48. Median Density Fits for Relationship 4

Relationship	Model	Mean	SD	Median
1	BNP	0.0238	0.0078	0.0229
	FNP	0.0233	0.0117	0.0211
2	BNP	0.0043	0.0025	0.0037
	FNP	0.0062	0.0058	0.0046
3	BNP	0.0080	0.0039	0.0073
	FNP	0.0101	0.0062	0.0089
4	BNP	0.0034	0.0029	0.0025
	FNP	0.0056	0.0057	0.0037

Table 3.9 Density fit statistics for the FNP and BNP approach when measurement error is increased by 50%. For each MIC density relationship, the BNP approach outperforms the FNP approach. The difference in performance is much greater compared to when $\sigma_m = 0.707$.

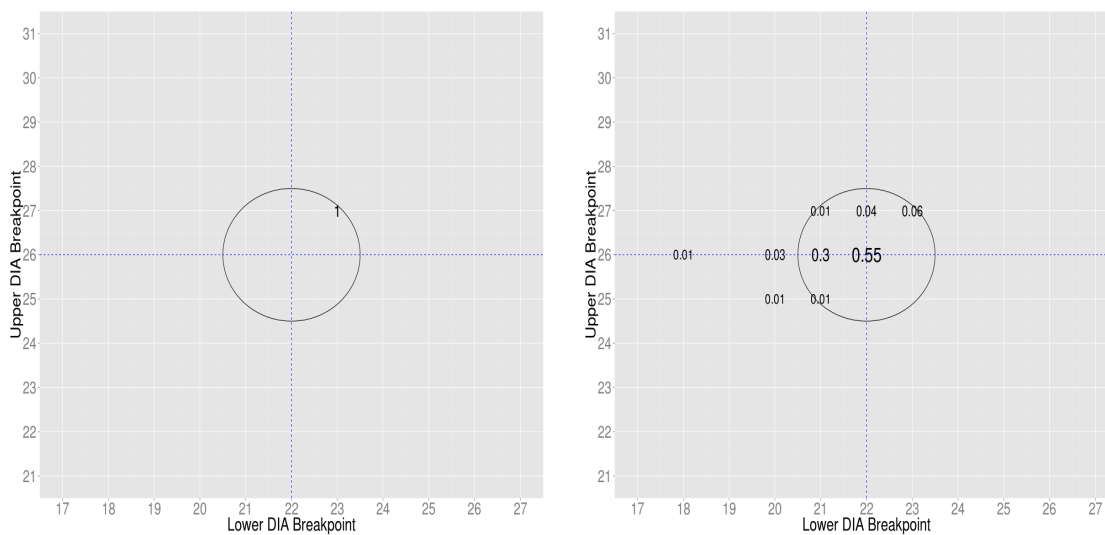**MIC Breakpoints: -3, -1**
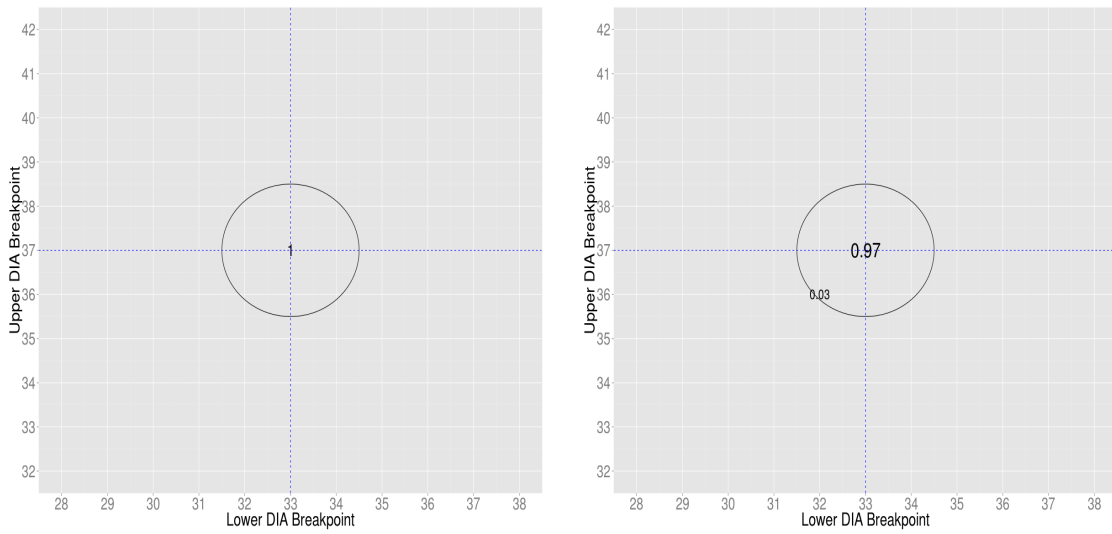
Figure 3.49. FNP and BNP Breakpoints for MIC/DIA Relationship 1 and MIC Breakpoints -3, -1
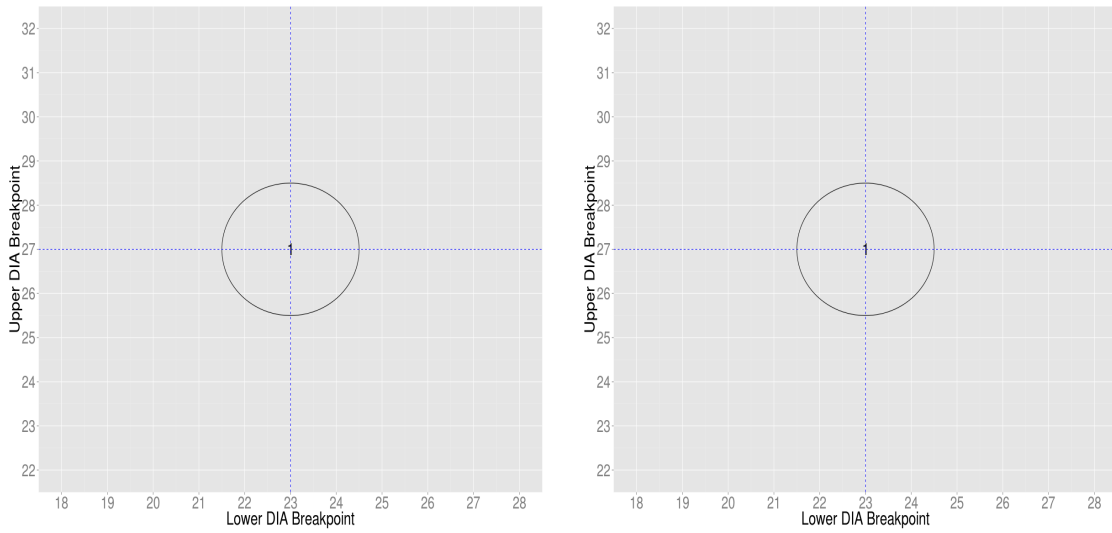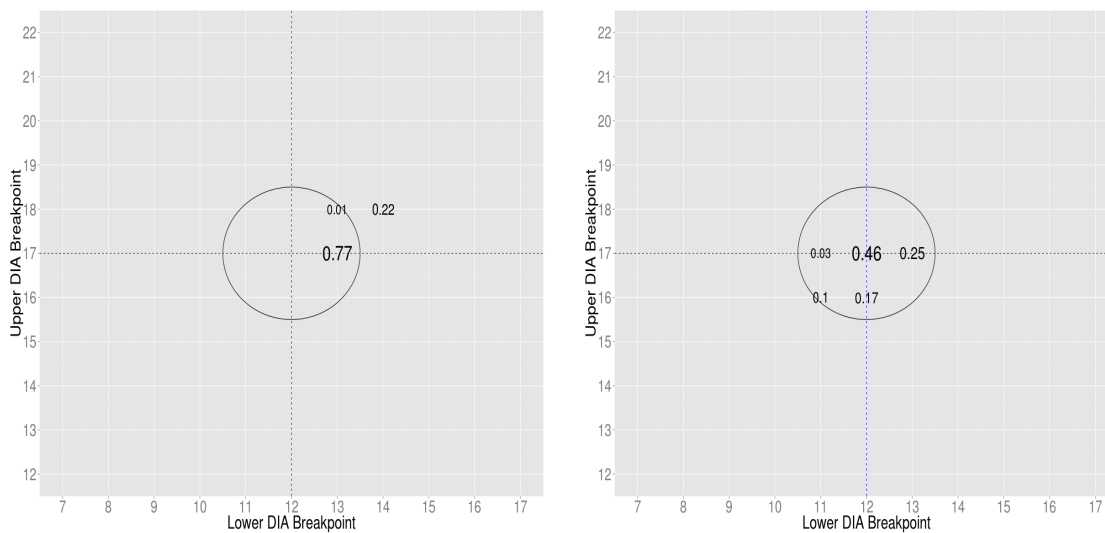

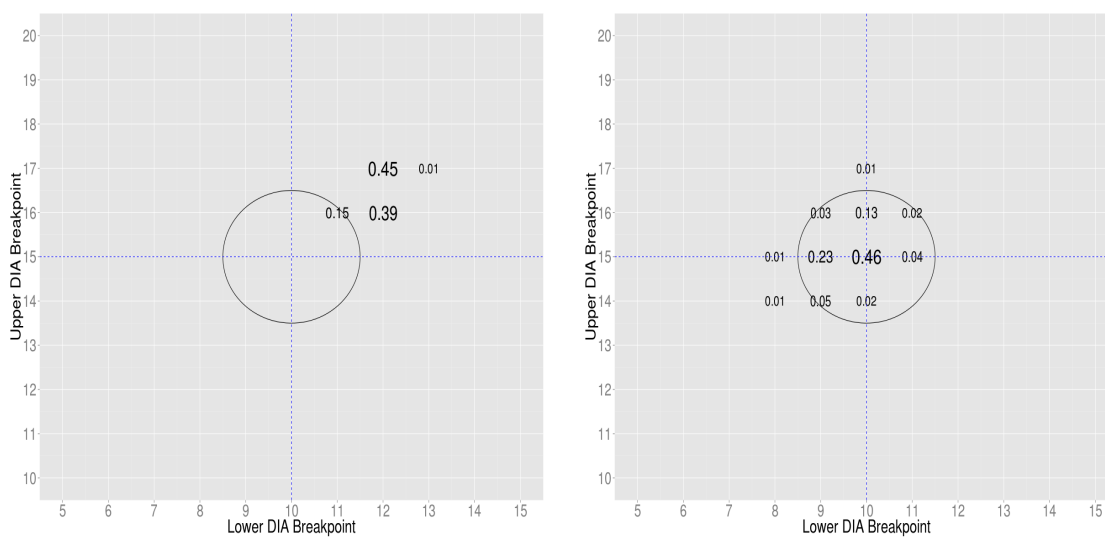Figure 3.50. FNP and BNP Breakpoints for MIC/DIA Relationship 2 and MIC Breakpoints -3, -1
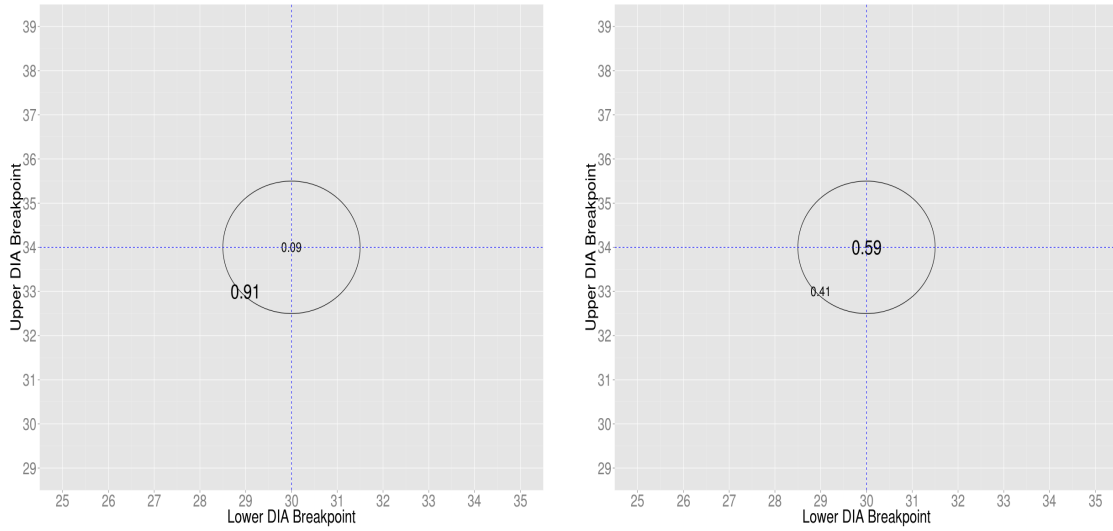
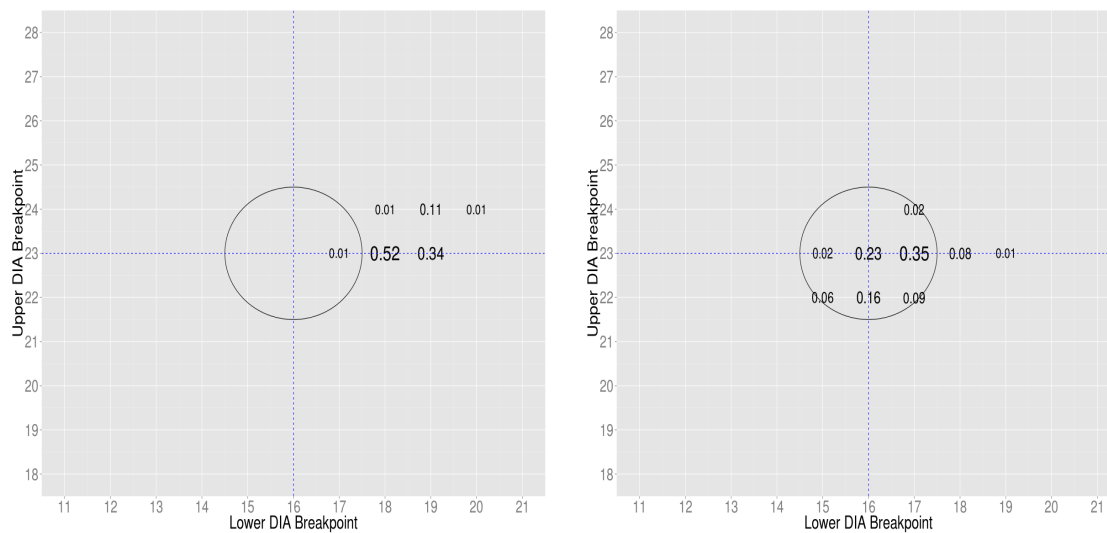Figure 3.51. FNP and BNP Breakpoints for MIC/DIA Relationship 3 and MIC Breakpoints -3, -1
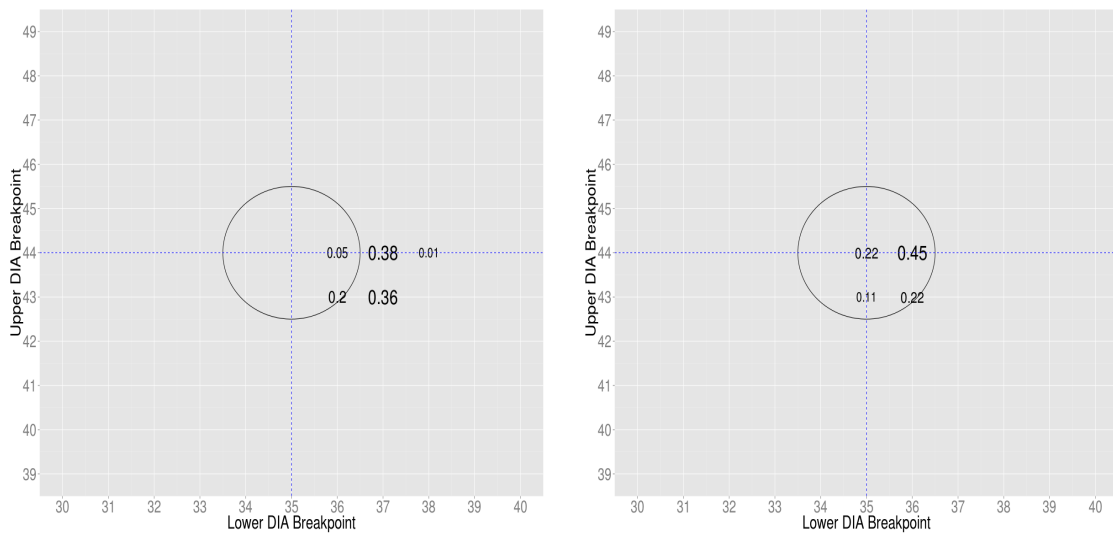
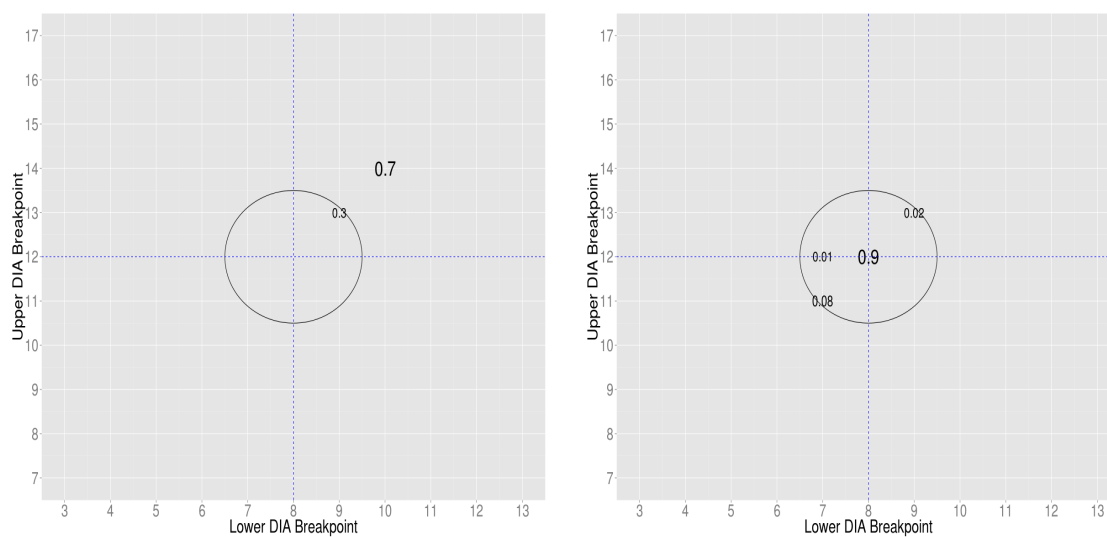Figure 3.52. FNP and BNP Breakpoints for MIC/DIA Relationship 4 and MIC Breakpoints -3, -1
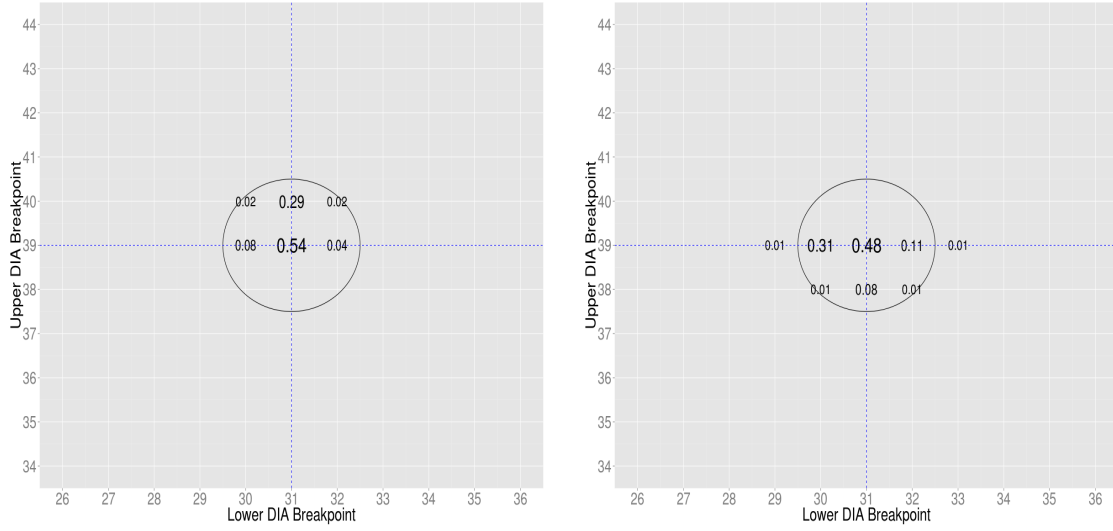
Figure 3.53. FNP and BNP Breakpoints for MIC/DIA Relationship 5 and MIC Breakpoints -3, -1

MIC Breakpoints: 1, 3

Figure 3.54. FNP and BNP Breakpoints for MIC/DIA Relationship 1 and MIC Breakpoints 1, 3

Figure 3.55. FNP and BNP Breakpoints for MIC/DIA Relationship 2 and MIC Breakpoints 1, 3

Figure 3.56. FNP and BNP Breakpoints for MIC/DIA Relationship 3 and MIC Breakpoints 1, 3

Figure 3.57. FNP and BNP Breakpoints for MIC/DIA Relationship 4 and MIC Breakpoints 1, 3
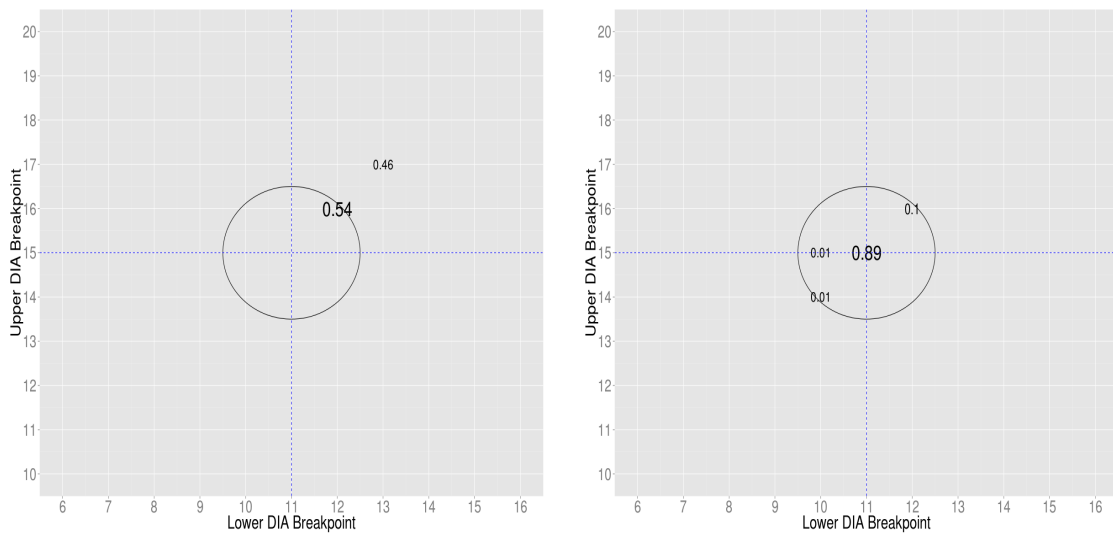
Figure 3.58. FNP and BNP Breakpoints for MIC/DIA Relationship 5 and MIC Breakpoints 1, 3

MIC Breakpoints: -4, -2

Figure 3.59. FNP and BNP Breakpoints for MIC/DIA Relationship 1 and MIC Breakpoints -4, -2

Figure 3.60. FNP and BNP Breakpoints for MIC/DIA Relationship 2 and MIC Breakpoints -4, -2

Figure 3.61.   FNP and BNP Breakpoints for MIC/DIA Relationship 3 and MIC Breakpoints -4, -2

Figure 3.62.   FNP and BNP Breakpoints for MIC/DIA Relationship 4 and MIC Breakpoints -4, -2
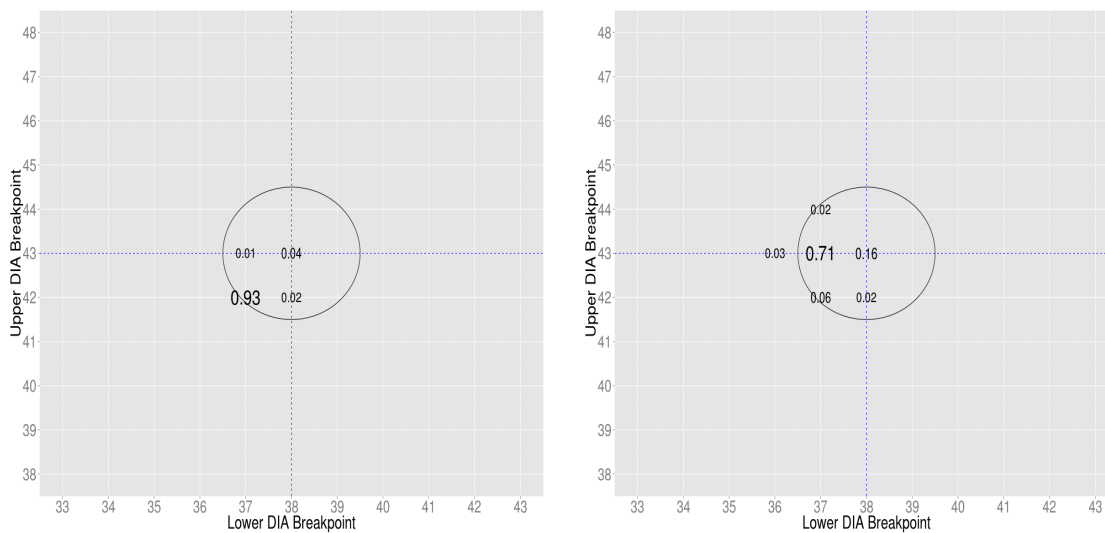
Figure 3.63. FNP and BNP Breakpoints for MIC/DIA Relationship 5 and MIC Breakpoints -4, -2

MIC Breakpoints: 2, 4

Figure 3.64. FNP and BNP Breakpoints for MIC/DIA Relationship 1 and MIC Breakpoints 2, 4

Figure 3.65. FNP and BNP Breakpoints for MIC/DIA Relationship 2 and MIC Breakpoints 2, 4

Figure 3.66. FNP and BNP Breakpoints for MIC/DIA Relationship 3 and MIC Breakpoints 2, 4

Figure 3.67. FNP and BNP Breakpoints for MIC/DIA Relationship 4 and MIC Breakpoints 2, 4

Figure 3.68. FNP and BNP Breakpoints for MIC/DIA Relationship 5 and MIC Breakpoints 2, 4

Summary

Interestingly, for both approaches there is no clear winner in terms of breakpoint performance when the MIC breakpoints are set at (-1, 1). Breakpoint performance was roughly similar for each MIC/DIA relationship in both measurement error cases. However, the MIC/DIA relationship statistics were much better for the BNP approach for both measurement error cases. For the MIC density estimates, for the first condition the BNP approach performed slightly better for all four cases. However when the measurement error increased the BNP approach performed much better in estimating the underlying MIC density.

Since the BNP approach did a better job of fitting the true MIC/DIA relationship, it is somewhat surprising the breakpoint performance was similar between the two approaches for MIC breakpoint (-1, 1). When the MIC breakpoints were shifted however we do see a performance difference for some of the MIC breakpoints.

For MIC breakpoints (-3, -1) performance was again roughly similar. For MIC breakpoints (1, 3) the BNP approach performed better for MIC/DIA relationships 3, 4, and 5. For MIC breakpoints (-4, -2) DIA breakpoint performance was roughly similar betweem the two approaches. For MIC breakpoints (2, 4) the BNP approach exceeded the FNP approach for all MIC/DIA relationships.

Give these results, the BNP approach does a better job of estimating the underlying MIC/DIA relationship and MIC density, but breakpoint performance tends to be roughly comparable. However, when the MIC breakpoints are moved toward the right side of the data, as in (1, 3) and (2, 4), the BNP approach exceeds the FNP approach.

3.3.3   BNP vs L4P Comparison

Condition	
MIC Density	right skewed
MIC/DIA Relationship	all five
σ_m, σ_d	0.707, 2.121
MIC Breakpoints	-1, 1
Number of Pathogen Strains	1000
Assessment	DIA Breakpoint Accuracy
	MIC Density Accuracy
	MIC/DIA Relationship Accuracy

Table 3.10 Scenario Conditions and Assessments

For this simulation study we compare the BNP approach to the L4P approach. We look at how well the approaches perform in terms of DIA breakpoint performance, estimating the underlying MIC/DIA relationship, and estimating the MIC density.
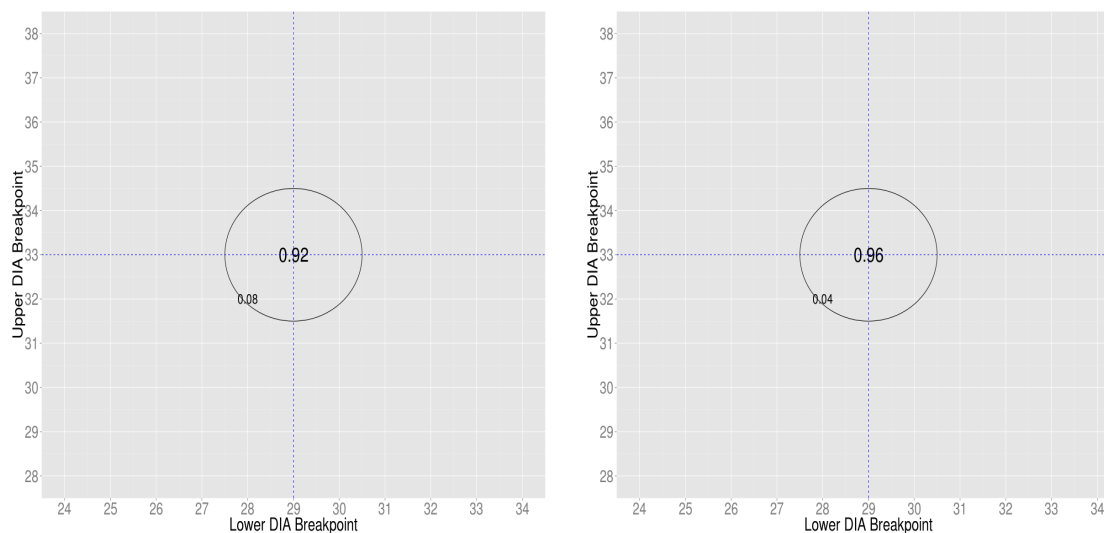
Figure 3.69. L4P and BNP Breakpoints for MIC/DIA Relationship 1
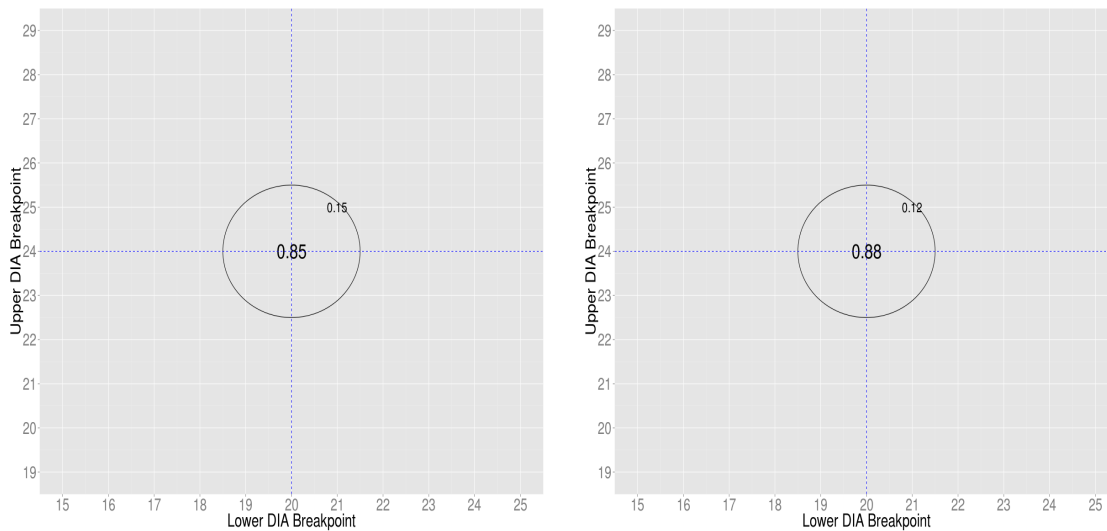
Figure 3.70. L4P and BNP Breakpoints for MIC/DIA Relationship 2
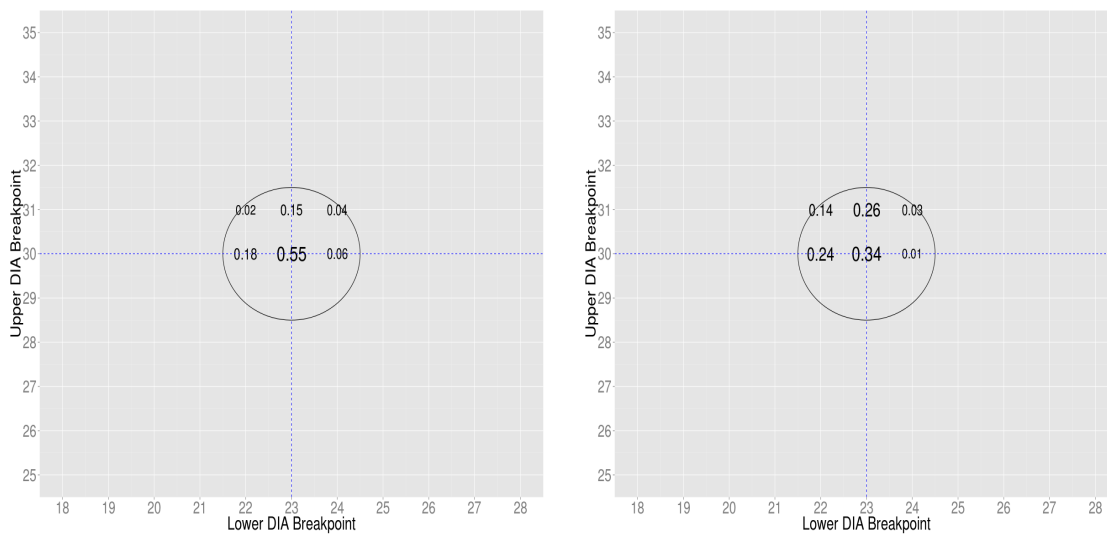
Figure 3.71. L4P and BNP Breakpoints for MIC/DIA Relationship 3
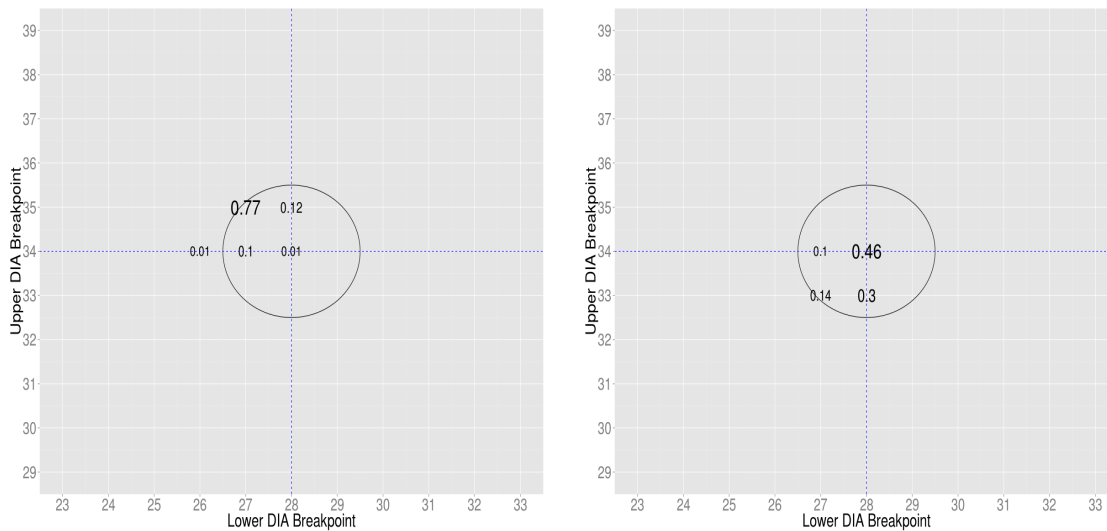
Figure 3.72. L4P and BNP Breakpoints for MIC/DIA Relationship 4
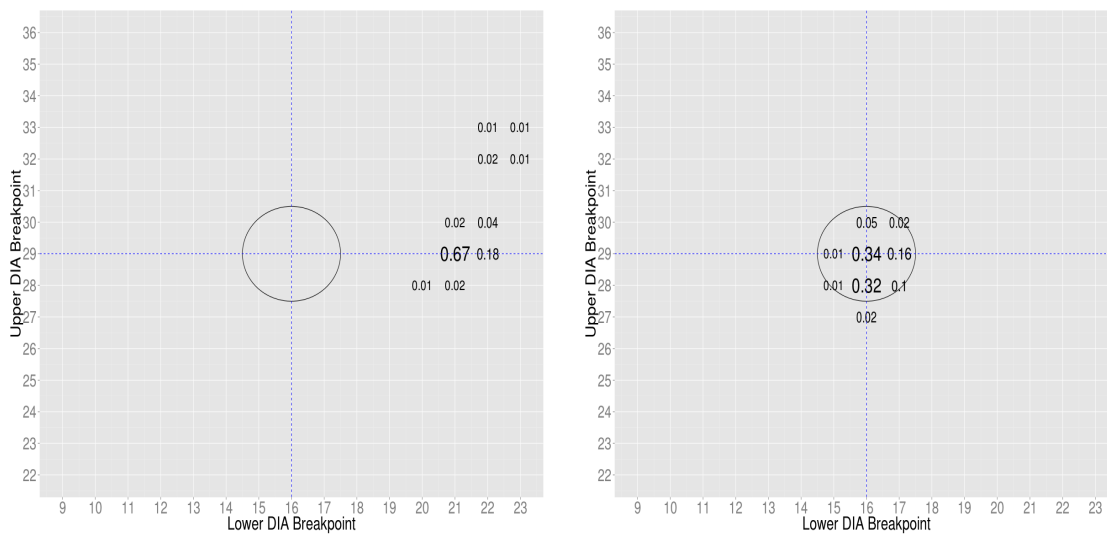
Figure 3.73. L4P and BNP Breakpoints for MIC/DIA Relationship 5
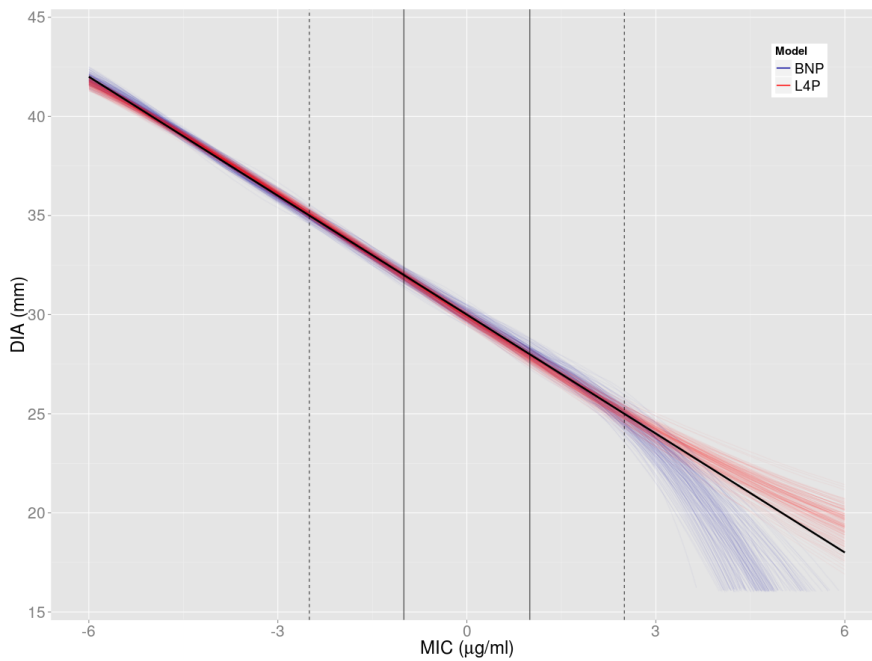
Figure 3.74. Median MIC/DIA Fits for Relationship 1
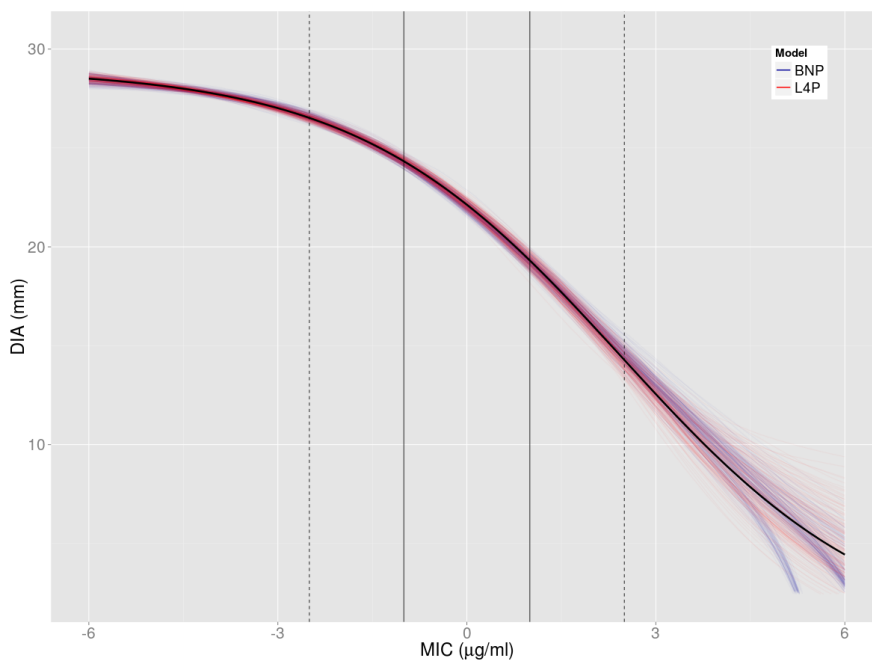
Figure 3.75. Median MIC/DIA Fits for Relationship 2
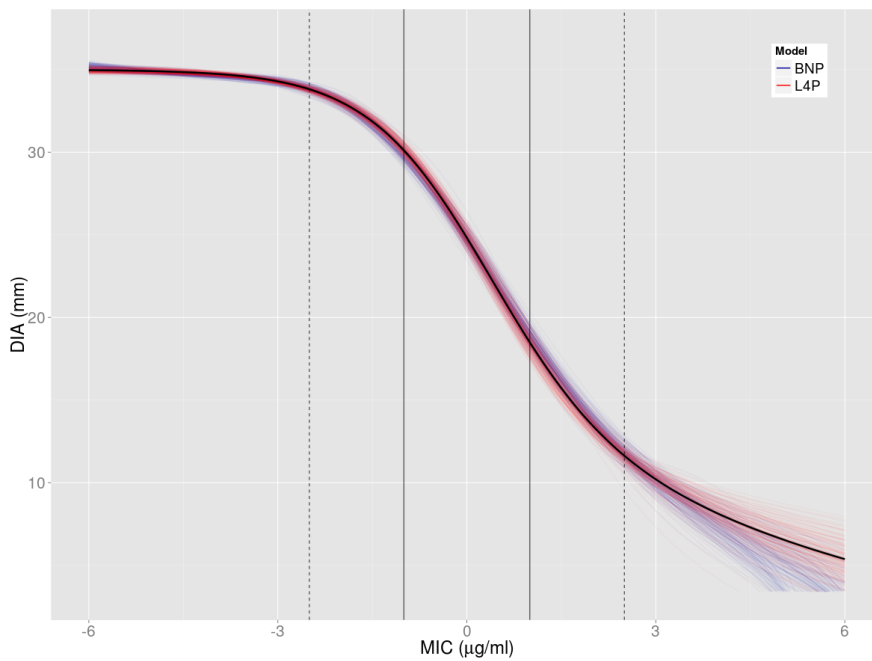
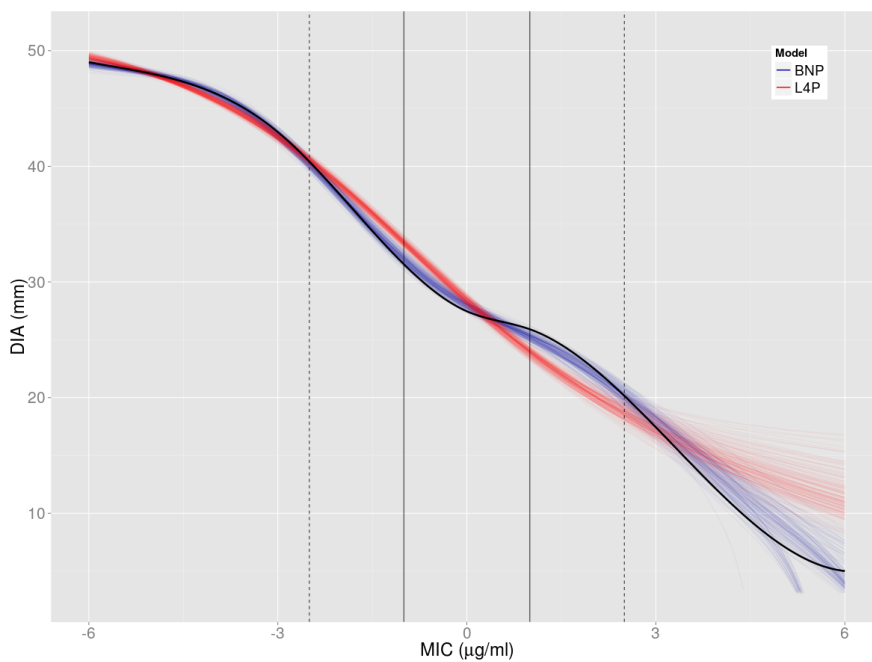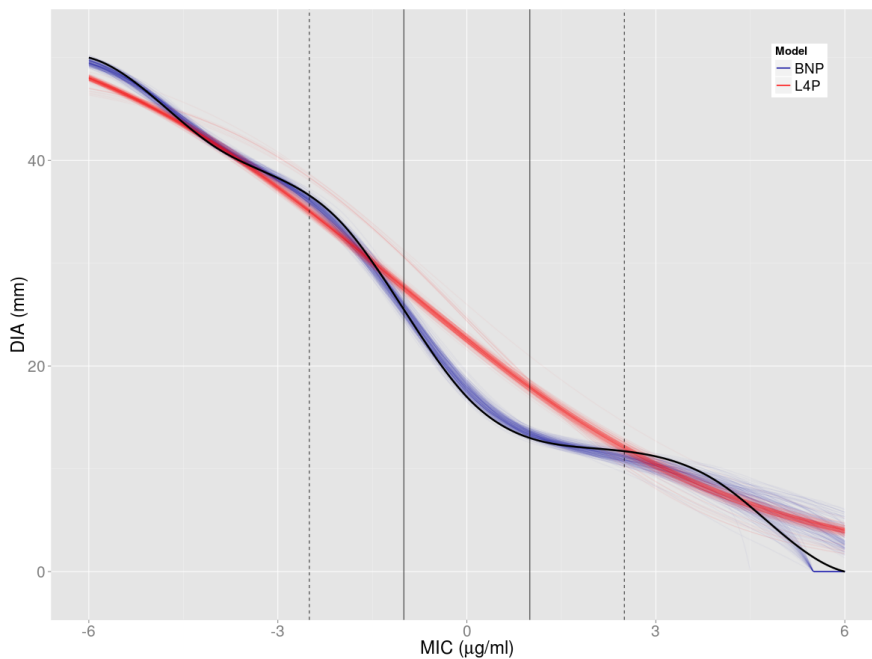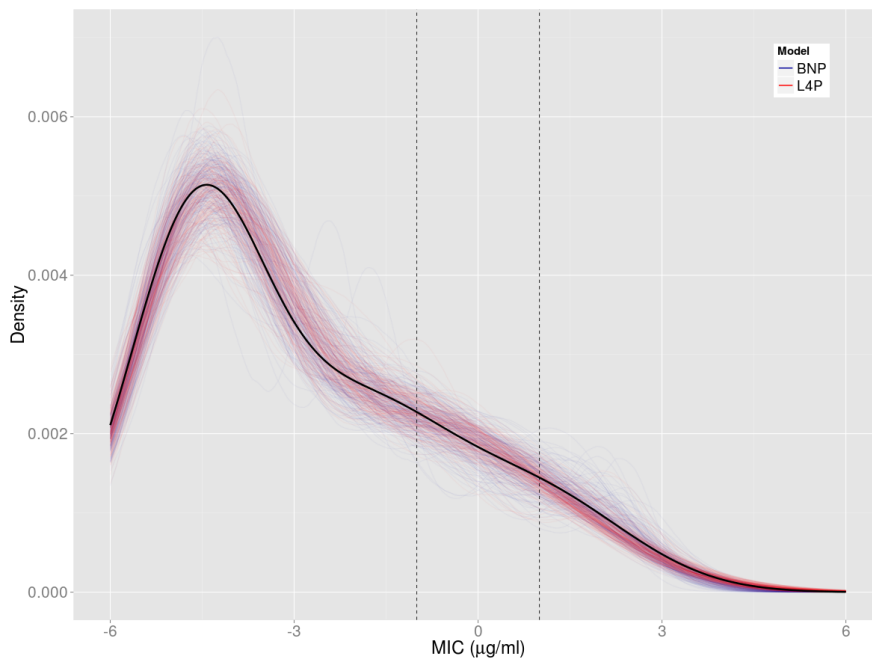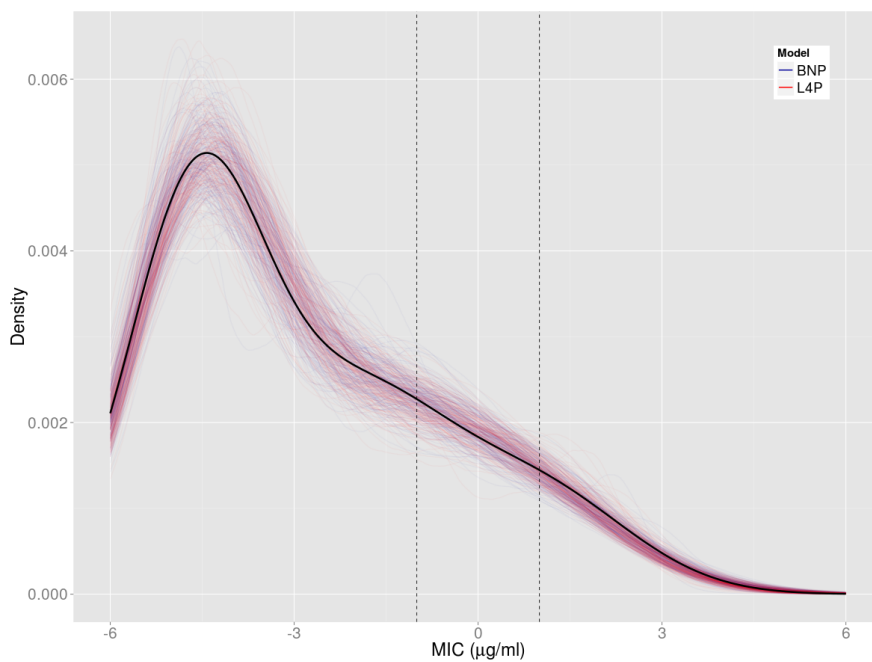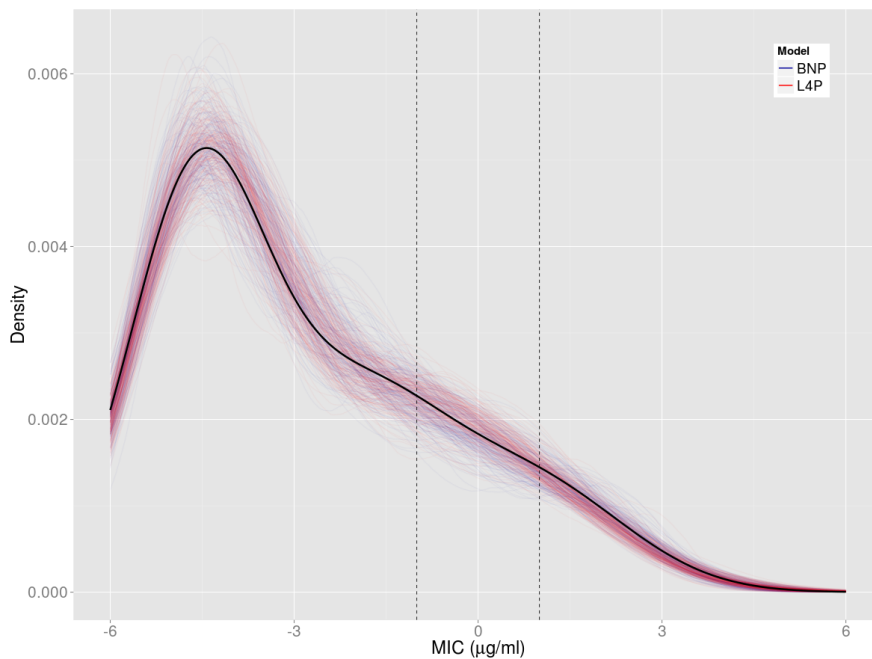Figure 3.76. Median MIC/DIA Fits for Relationship 3

Figure 3.77. Median MIC/DIA Fits for Relationship 4

Figure 3.78. Median MIC/DIA Fits for Relationship 5

Relationship	Model	Mean	SD	Median
1	BNP	15.38	12.98	11.09
	L4P	8.94	10.21	5.58
2	BNP	14.10	13.09	9.31
	L4P	13.82	13.92	9.59
3	BNP	42.90	38.97	29.69
	L4P	30.12	29.31	19.91
4	BNP	66.36	45.67	52.98
	L4P	607.10	131.92	599.20
5	BNP	97.72	61.76	84.00
	L4P	2850.00	714.40	2720.00

Table 3.11 MIC/DIA Relationship Fit Statistics

Figure 3.79. Median MIC Density Fits for Relationship 1

Figure 3.80. Median MIC Density Fits for Relationship 2

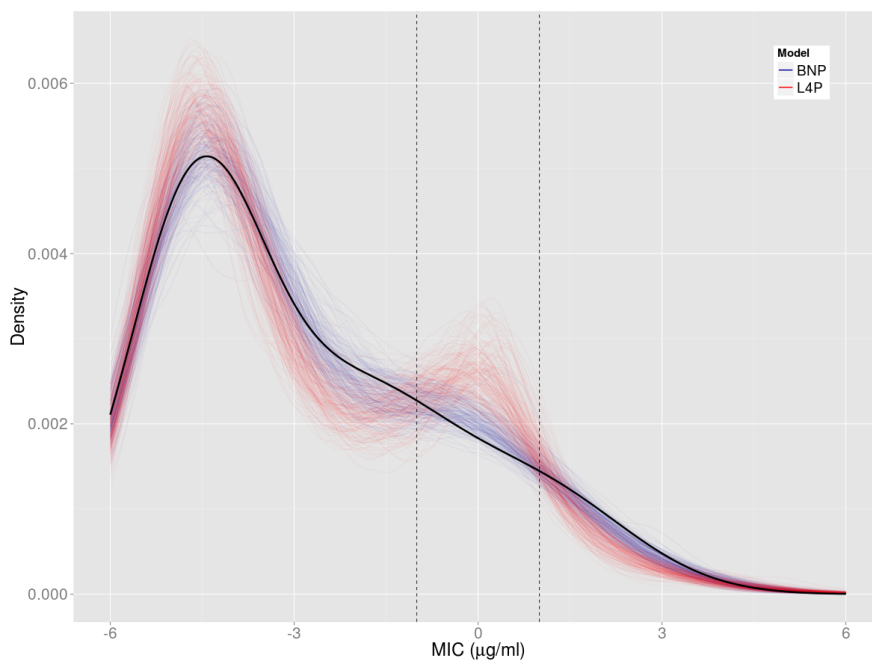Figure 3.81. Median MIC Density Fits for Relationship 3

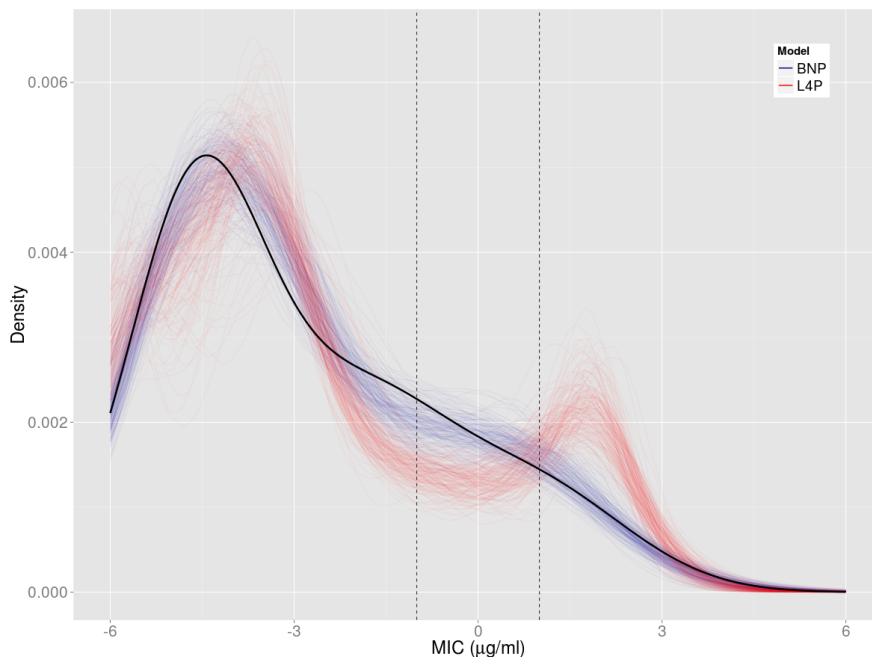Figure 3.82. Median MIC Density Fits for Relationship 4

Figure 3.83. Median MIC Density Fits for Relationship 5

Performance is similar between the BNP and L4P approaches for relationships 1, 2, and 3. However for relationships 4 and 5, where the L4P approach does not estimate the MIC/DIA relationship well (Figures 3.77 and 3.78), the BNP approach greatly outperforms the L4P approach in terms DIA breakpoint estimation, the MIC/DIA relationship, and the underlying MIC/DIA relationship. Interestingly, the biased fit of the MIC/DIA relationship for the L4P approach in relationships 4 and 5 also biases the underlying MIC density estimate (Figures 3.82 and 3.83).

Relationship	Model	Percentage Exact	Percentage ±1
1	BNP	0.96	1
	L4P	0.92	1
2	BNP	0.88	1
	L4P	0.85	1
3	BNP	0.34	1
	L4P	0.55	1
4	BNP	0.43	0.99
	L4P	0.01	0.99
5	BNP	0.34	0.97
	L4P	0	0

Table 3.12 DIA breakpoint performance for the BNP and L4P approaches. Performance is similar for the first three relationships when the L4P approach fits the underlying MIC/DIA relationship well. However in relationships 4 and 5, where the L4P approach does not fit the relationship well, the BNP approach greatly exceeds it.

3.3.4 Comparison: RJBNP vs BNP

Condition	
MIC Density	right skewed
MIC/DIA Relationship	all five
σ_m, σ_d	0.707, 2.121
MIC Breakpoints	-1, 1
Number of Pathogen Strains	1000
Assessment	DIA Breakpoint Accuracy
	MIC Density Accuracy
	MIC/DIA Relationship Accuracy

Table 3.13 Scenario Conditions and Assessments

In this scenario we investigate the performance of our two approaches to knot selection, BNP and RJBNP. We visually display DIA breakpoint performance, the

underlying MIC/DIA relationship fits, and the underlying MIC density fits from the 200 generated scatterplots.
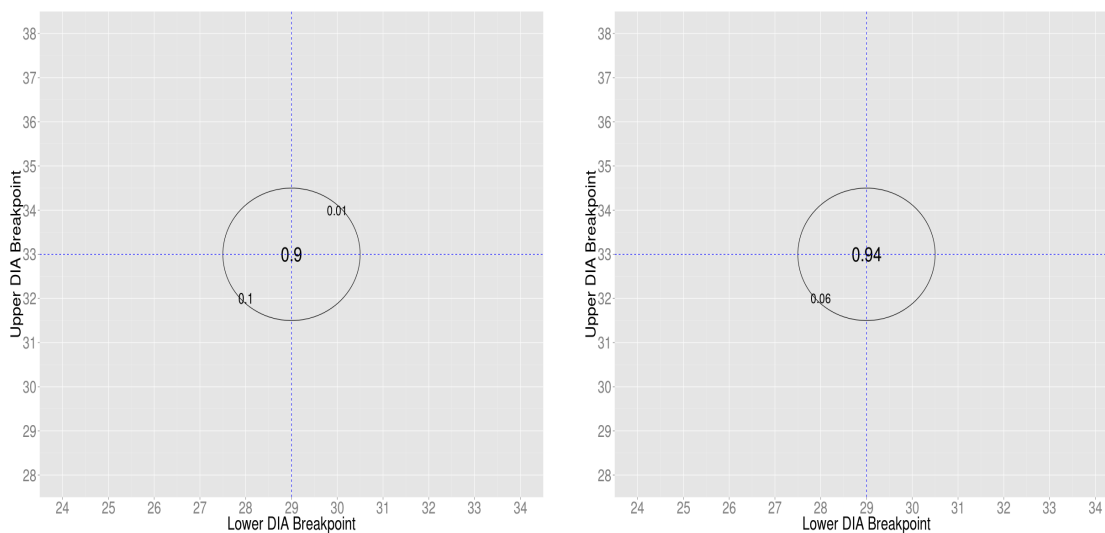

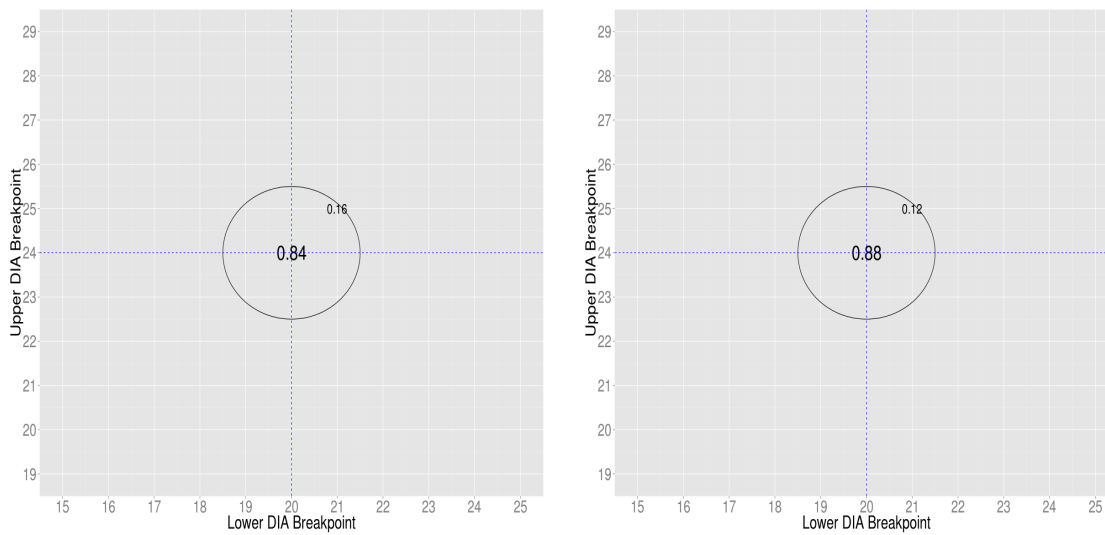Figure 3.84. RJBNP and BNP Breakpoints for MIC/DIA Relationship 1

Figure 3.85. RJBNP and BNP Breakpoints for MIC/DIA Relationship 2
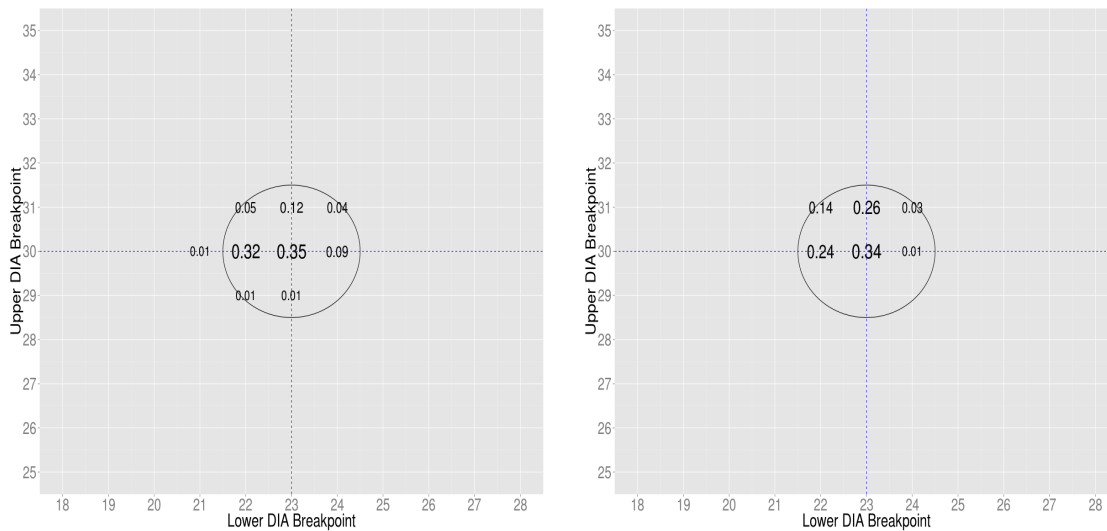
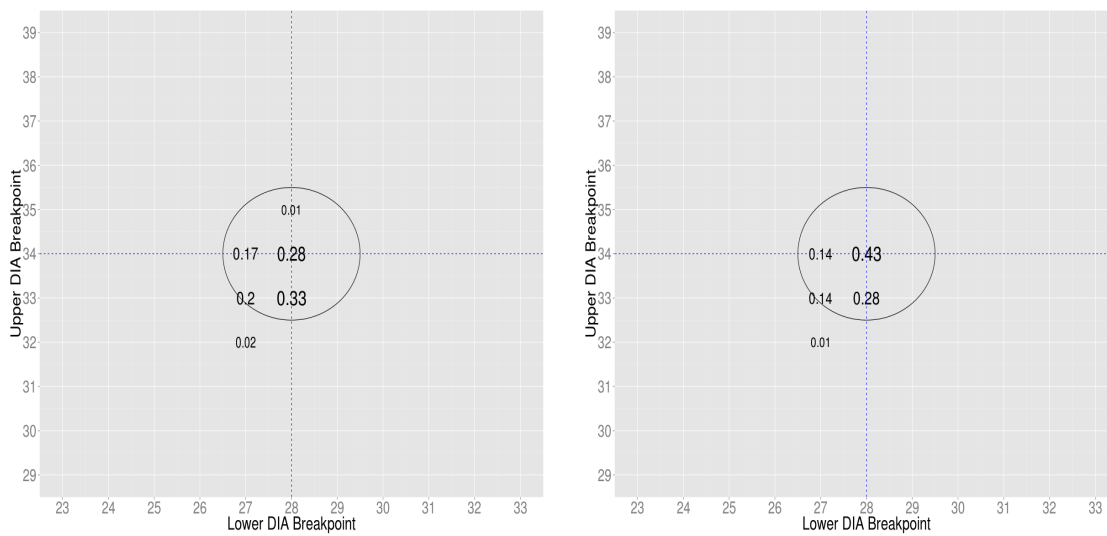Figure 3.86. RJBNP and BNP Breakpoints for MIC/DIA Relationship 3

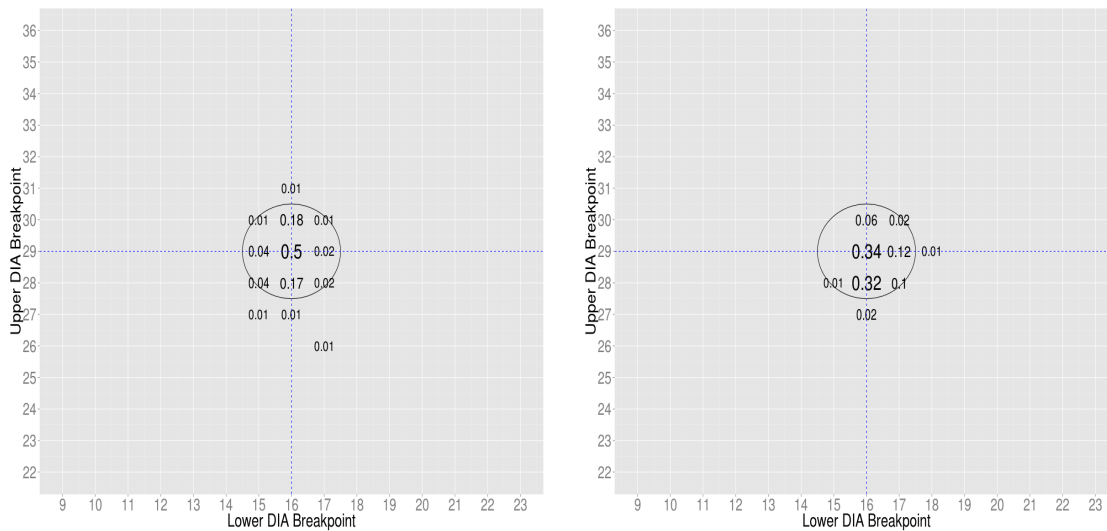Figure 3.87. RJBNP and BNP Breakpoints for MIC/DIA Relationship 4

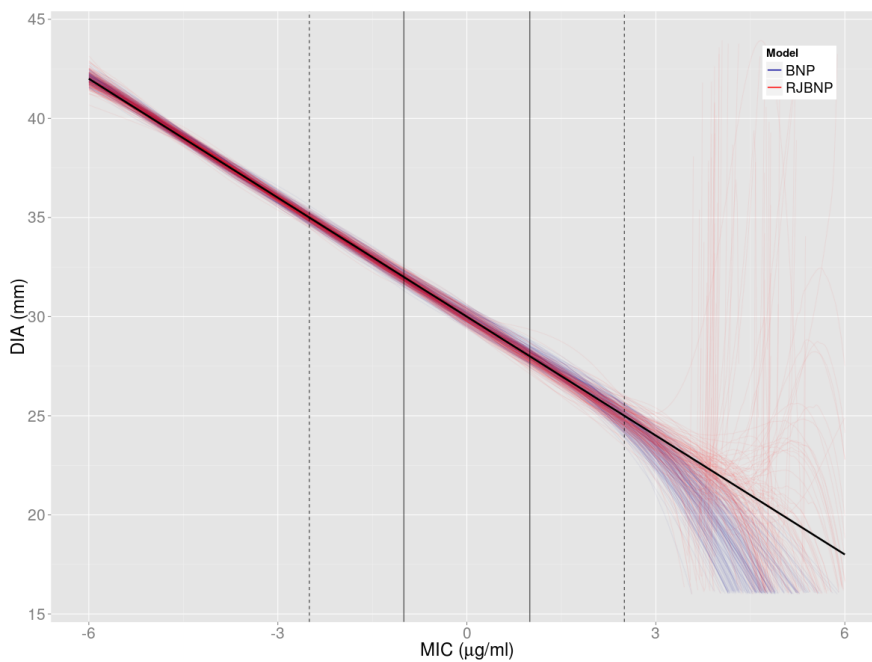Figure 3.88. RJBNP and BNP Breakpoints for MIC/DIA Relationship 5

Figure 3.89. Median MIC/DIA Fits for Relationship 1
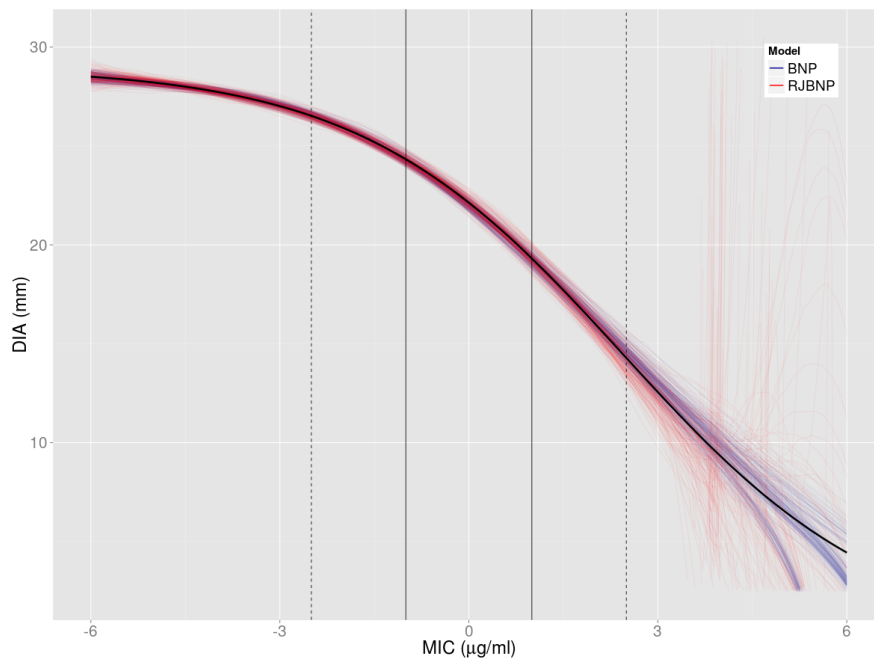
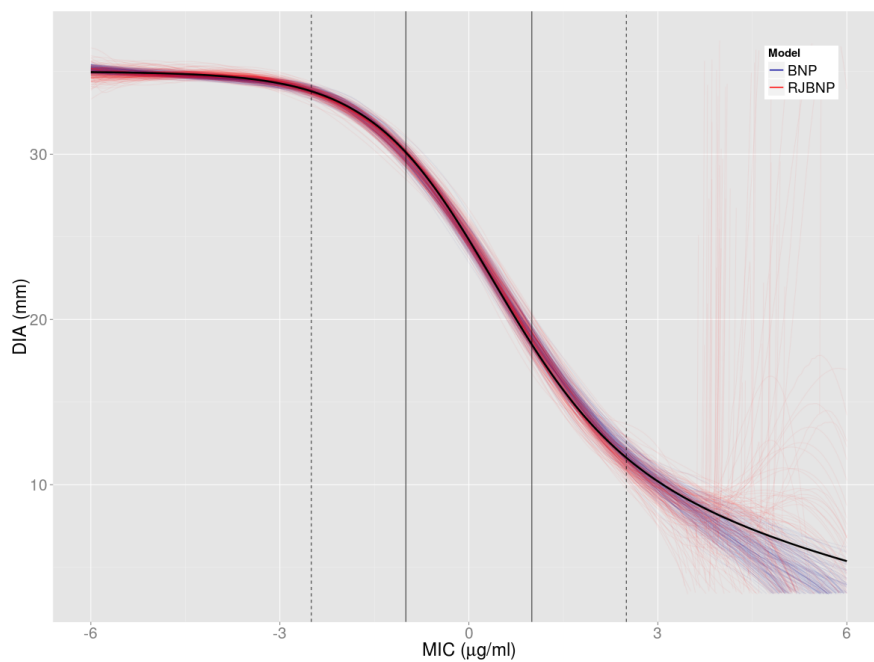Figure 3.90. Median MIC/DIA Fits for Relationship 2

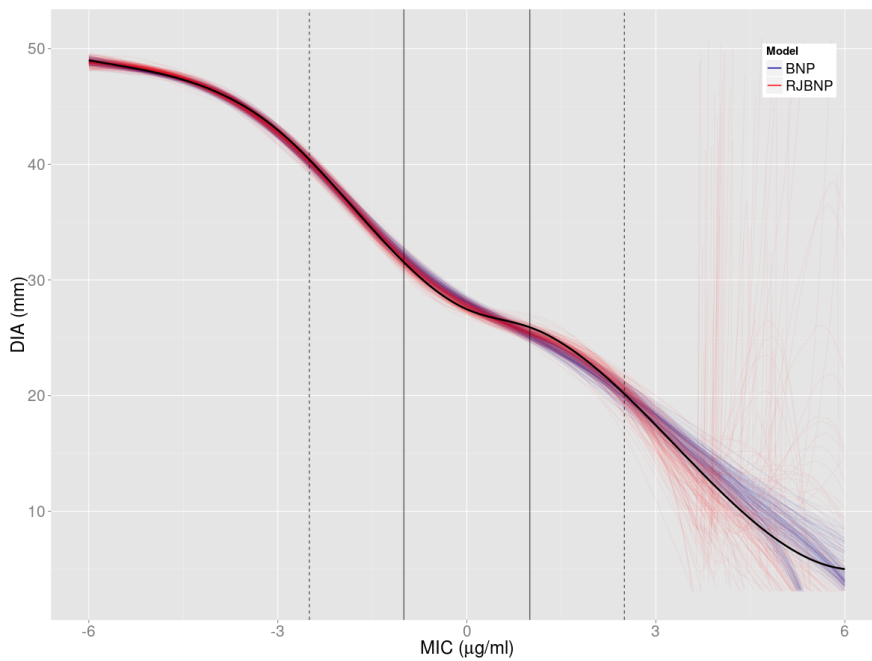Figure 3.91. Median MIC/DIA Fits for Relationship 3

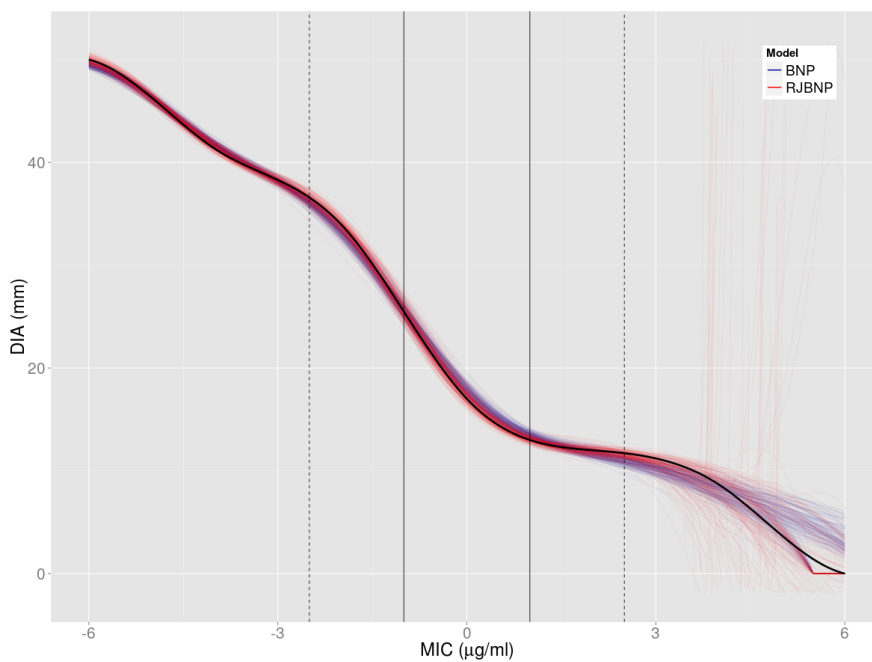Figure 3.92. Median MIC/DIA Fits for Relationship 4

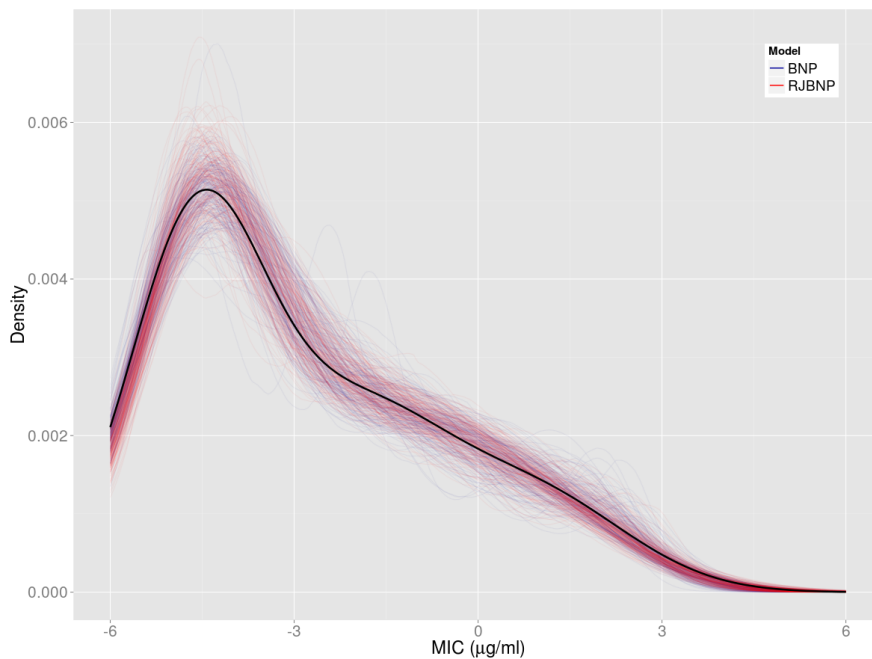Figure 3.93. Median MIC/DIA Fits for Relationship 5

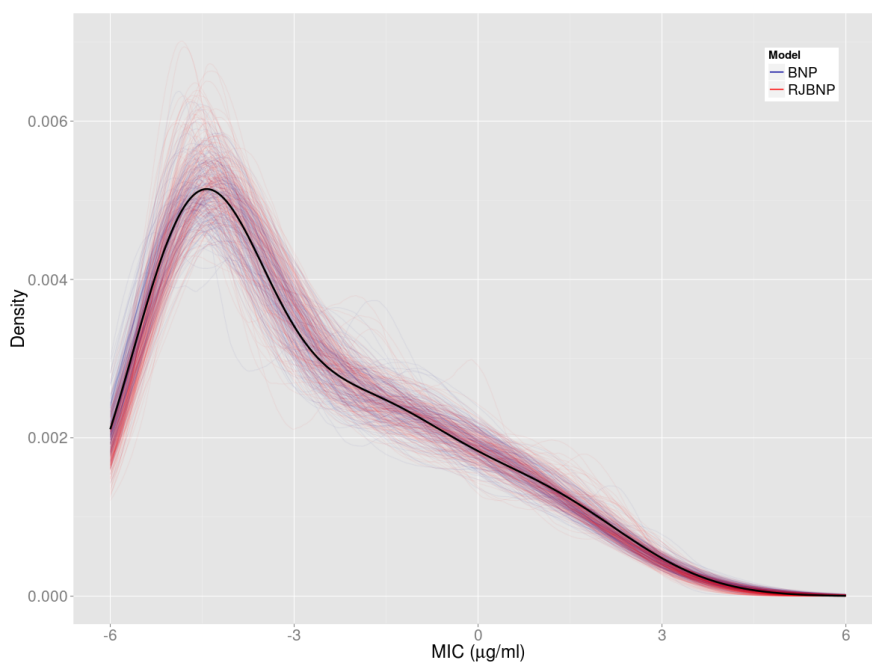Figure 3.94. Median MIC Density Fits for Relationship 1

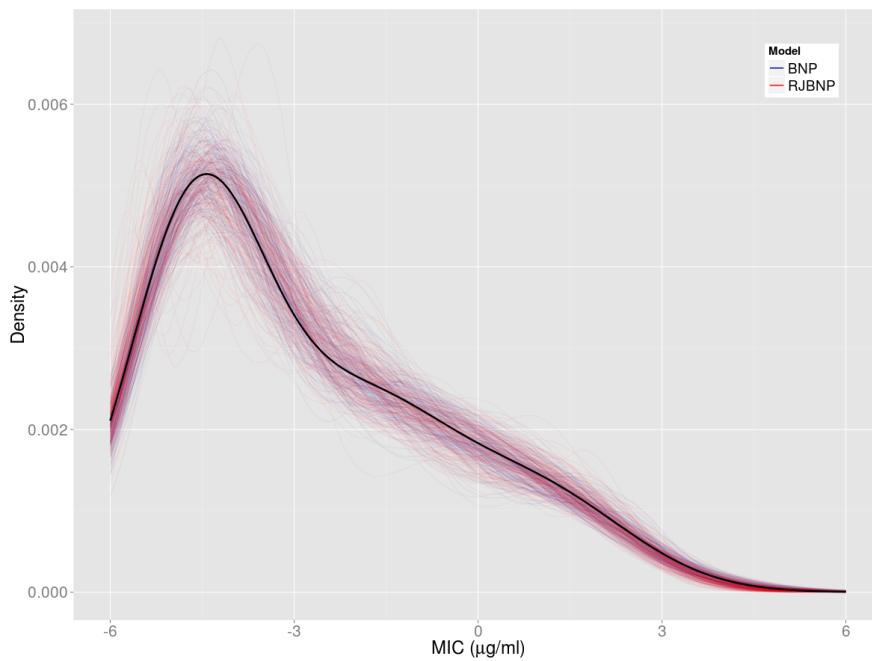Figure 3.95. Median MIC Density Fits for Relationship 2

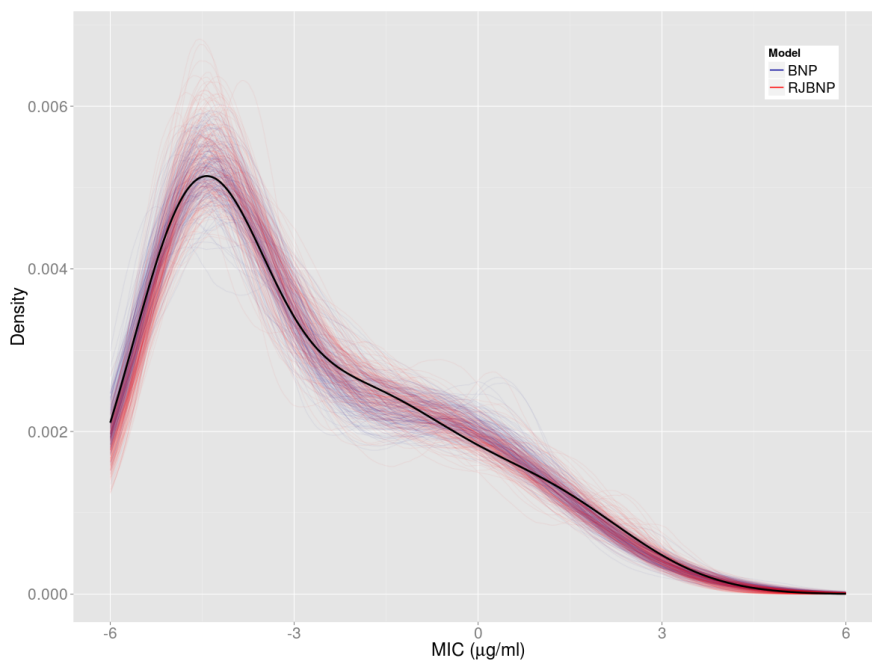Figure 3.96. Median MIC Density Fits for Relationship 3

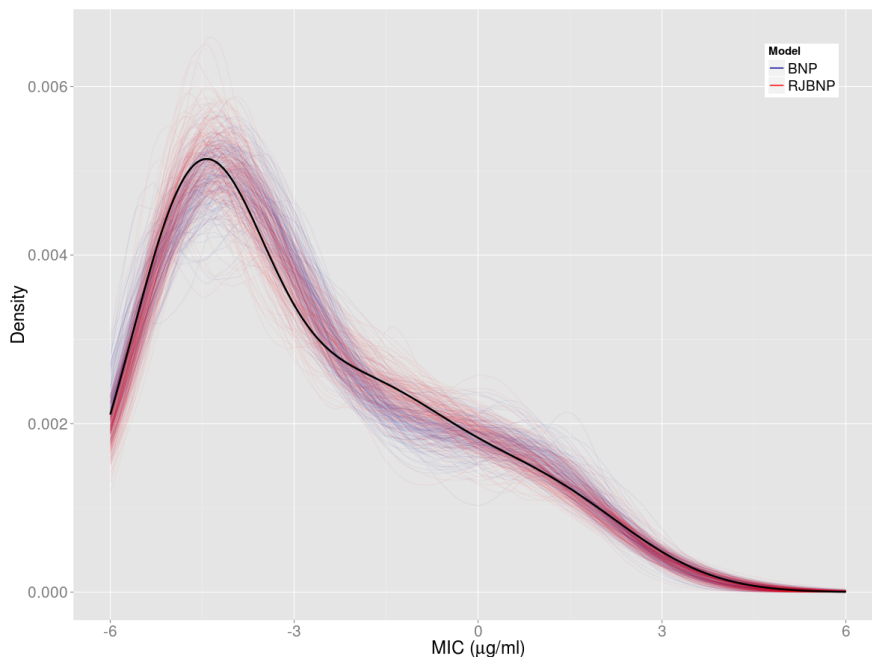Figure 3.97. Median MIC Density Fits for Relationship 4

Figure 3.98. Median MIC Density Fits for Relationship 5

We see the BNP and BNPRJ approaches are very comparable in terms of DIA breakpoint performance for all five MIC/DIA relationships. The estimates of the underlying MIC density are also very similar between the two approaches. The estimated fits for the underlying MIC/DIA relationship are also similar except near the right endpoint where the RJBNP approach can vary wildly.

For the RJBNP approach, the coefficients are not restricted to be positive to ensure montonicity. Instead, the resulting fit is checked after updates to make the curve is monotonic except near the endpoints. This allows more flexibility in the coefficients and better mixing for the reversible jump step. However, for some scatterplots, the fit spikes way up towards the right side of the data. More investigation is needed to prevent such large swings in the estimated fit.

3.3.5 BNP - Few Isolates Near the Indeterminant Range

Condition	Value
MIC Density	two component
MIC/DIA Relationship	all five
σ_m, σ_d	0.707, 2.121
MIC Breakpoints	-1, 1
Number of Pathogen Strains	1000
Assessment	DIA Breakpoint Accuracy
	MIC Density Accuracy
	MIC/DIA Relationship Accuracy

Table 3.14 Scenario Conditions and Assessments

Some real data sets have shown few isolates near the indeterminant range, with a majority of isolates in the susceptible range. In this scenario we investigate the performance of the BNP approach when this is the case. We look at DIA breakpoint performance, MIC density accuracy, and MIC/DIA relationship accuracy.
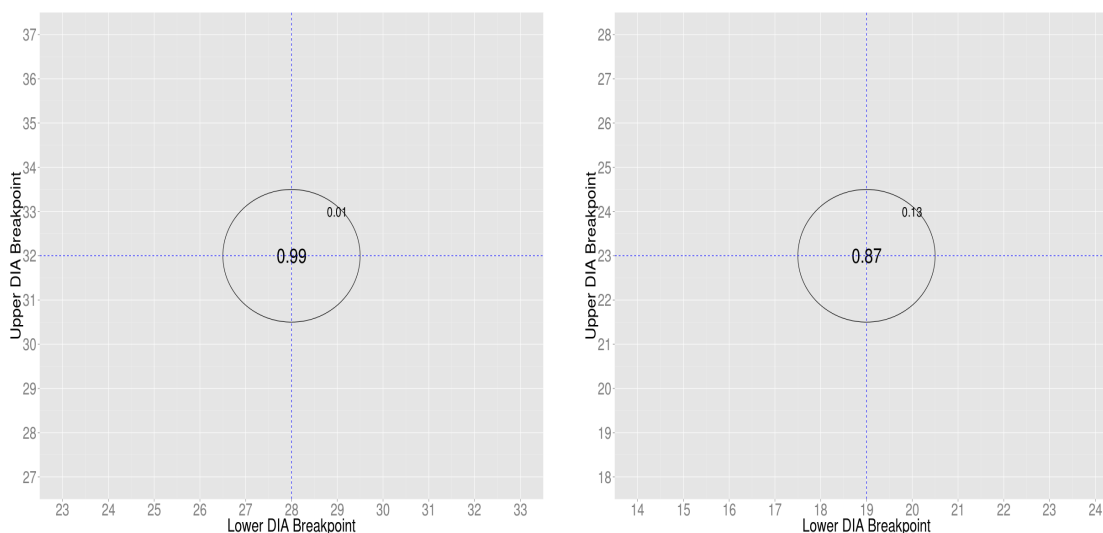
Figure 3.99. BNP Breakpoints for MIC/DIA Relationships 1 and 2
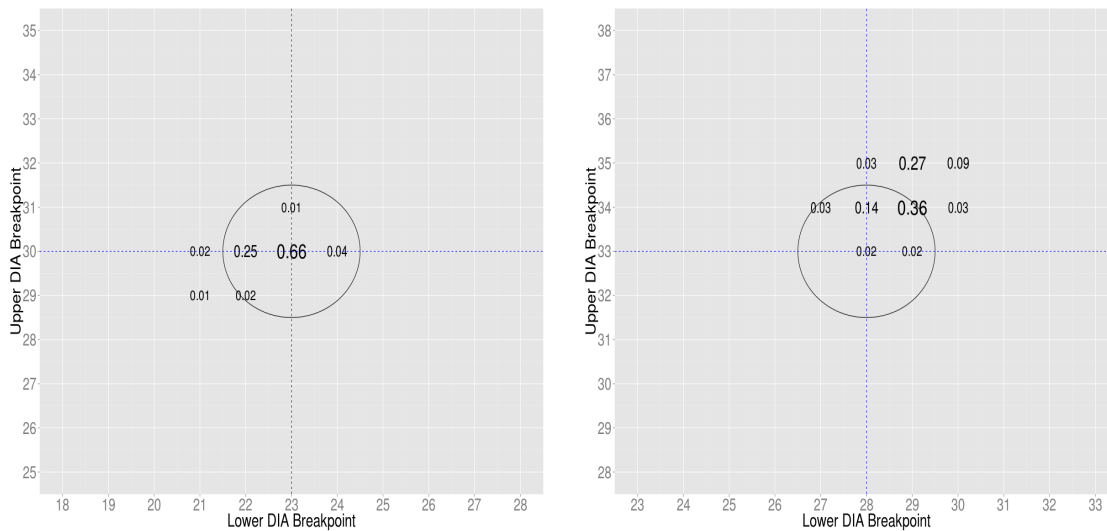
Figure 3.100. BNP Breakpoints for MIC/DIA Relationships 3 and 4
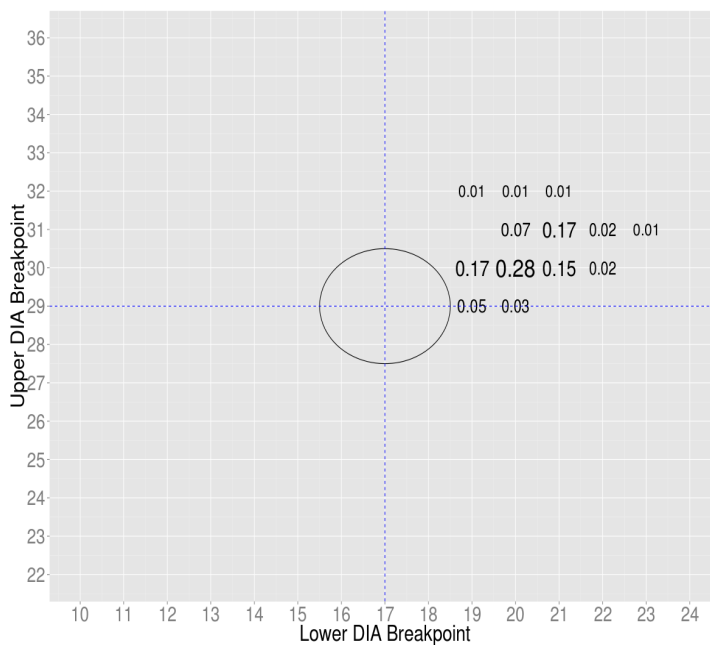
Figure 3.101. BNP Breakpoints for MIC/DIA Relationship 5

Figure 3.102. Median MIC/DIA Fits for Relationship 1

Figure 3.103. Median MIC/DIA Fits for Relationship 2

Figure 3.104. Median MIC/DIA Fits for Relationship 3

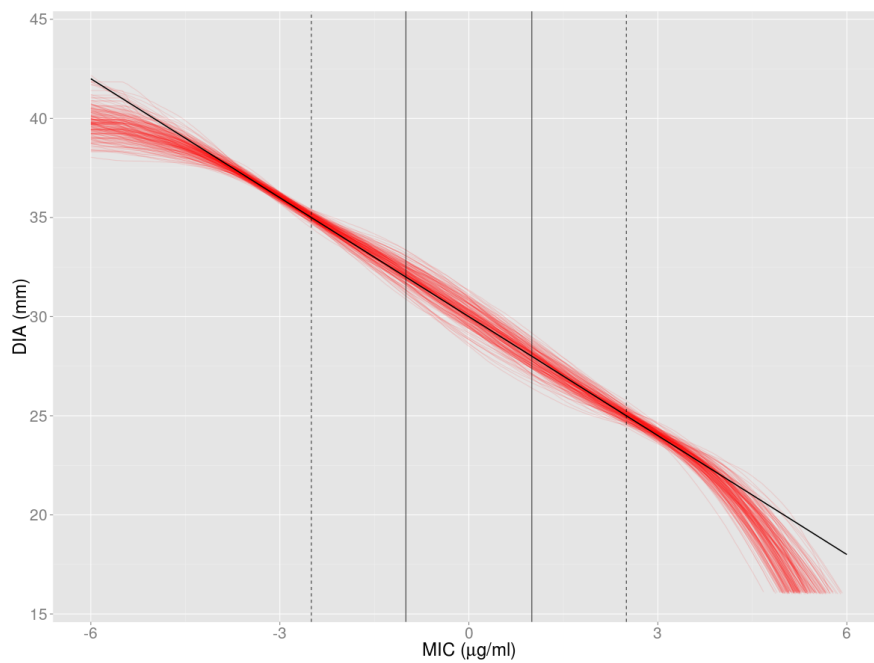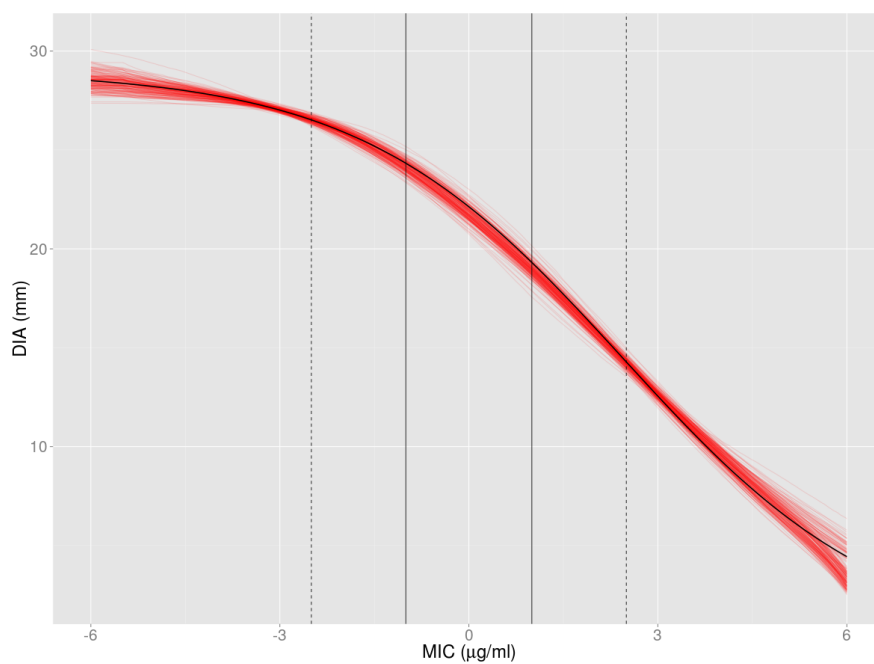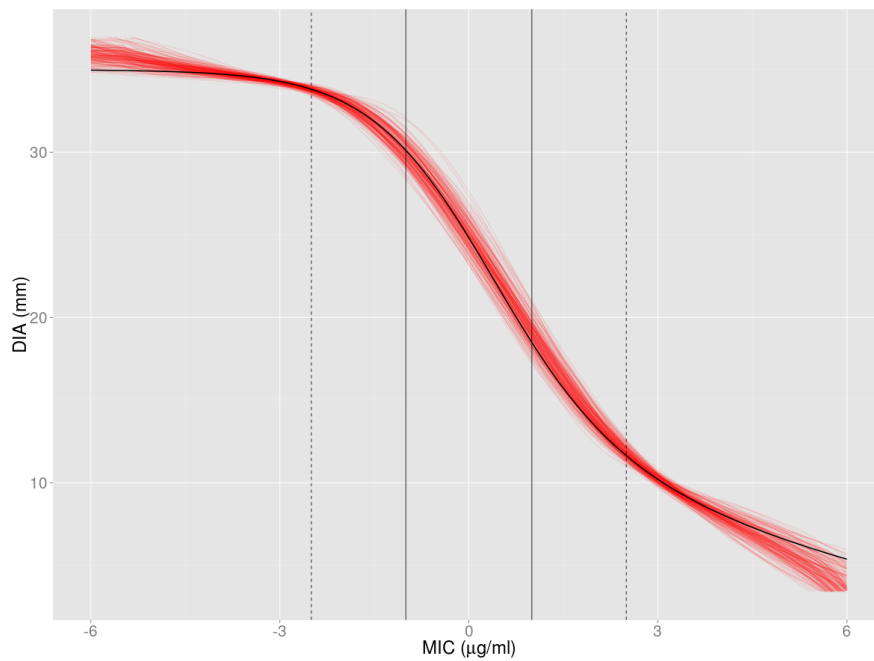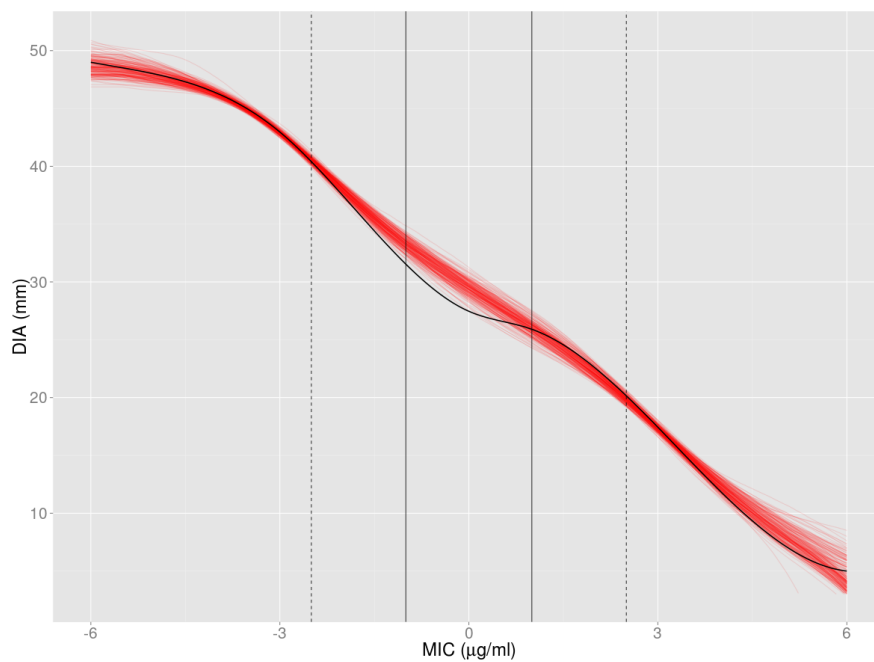Figure 3.105. Median MIC/DIA Fits for Relationship 4
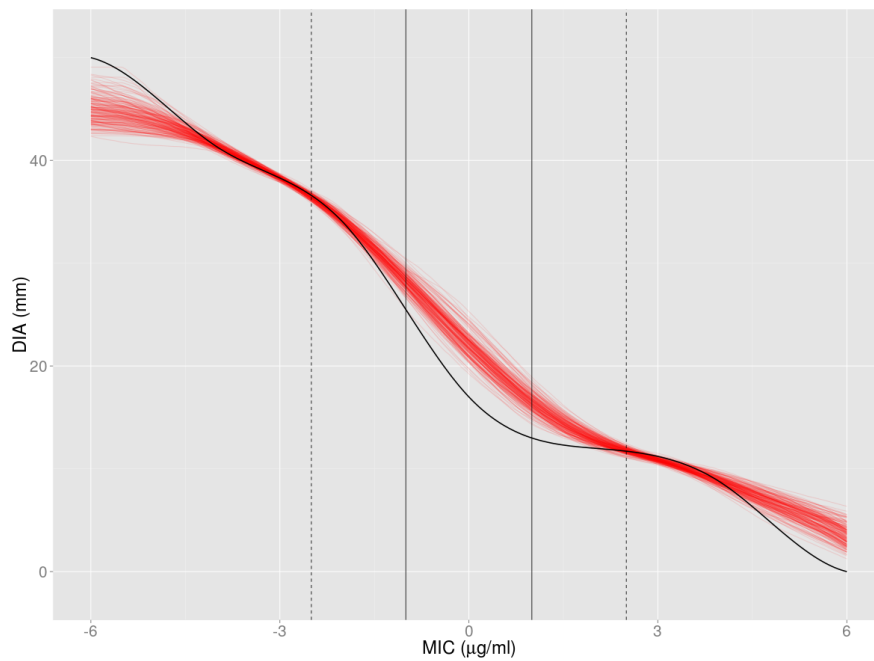
Figure 3.106. Median MIC/DIA Fits for Relationship 5
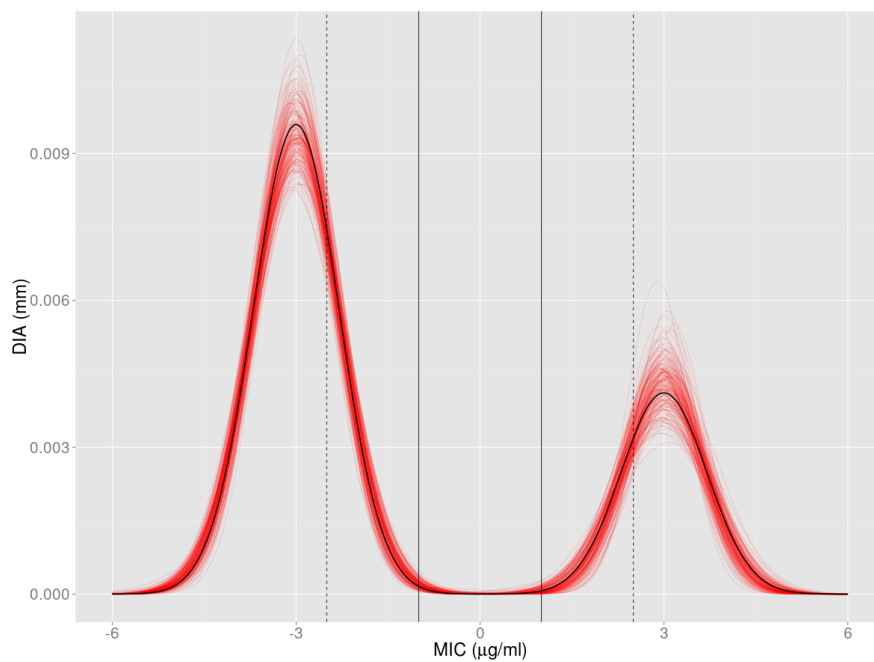
Figure 3.107. Median MIC Density Fits for Relationship 1
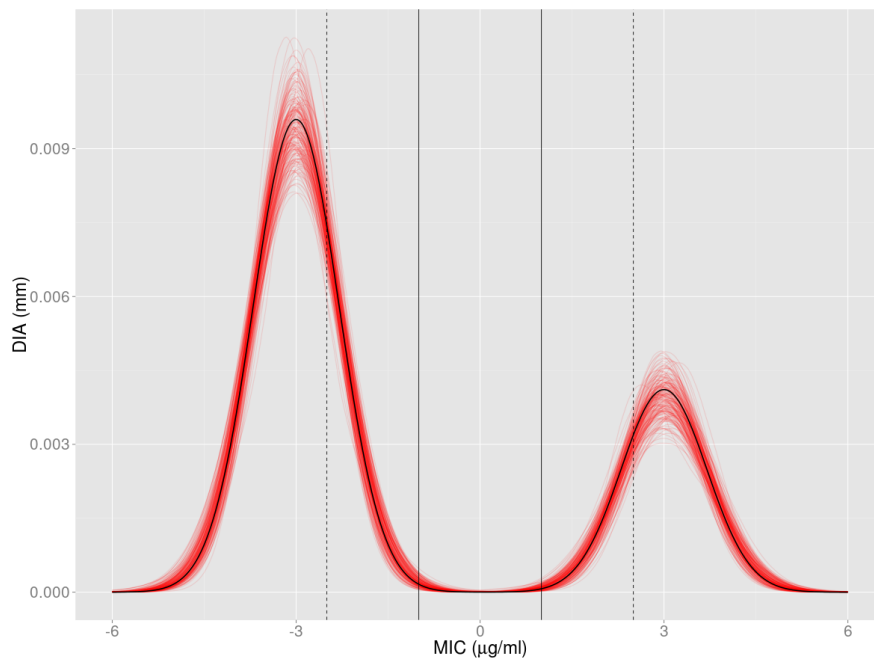
Figure 3.108. Median MIC Density Fits for Relationship 2
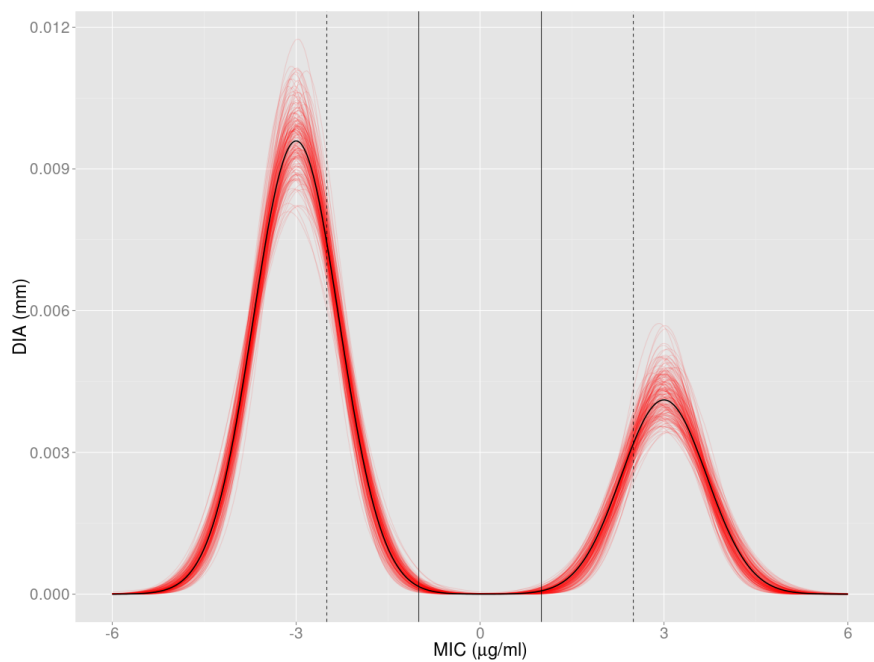
Figure 3.109. Median MIC Density Fits for Relationship 3
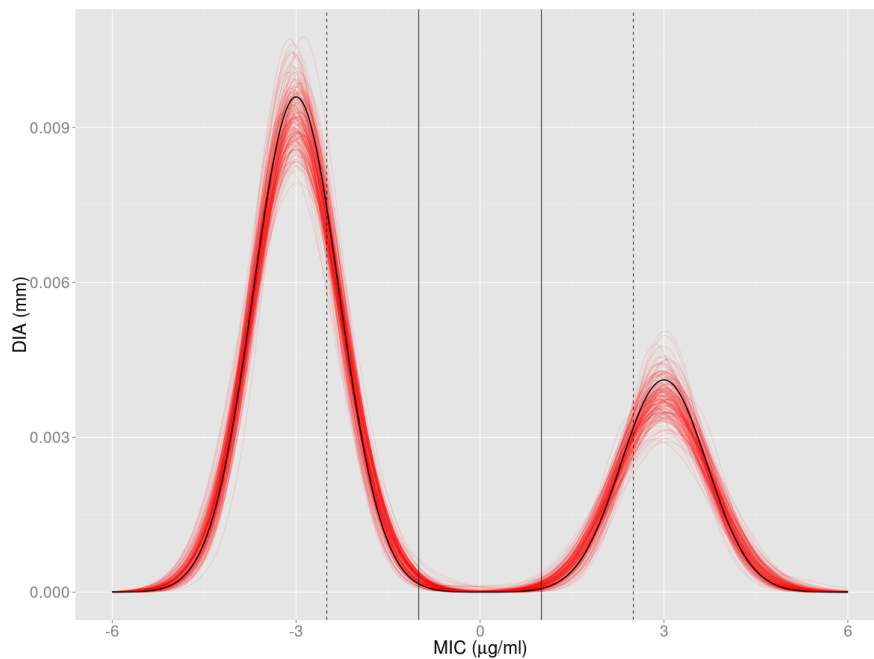
Figure 3.110. Median MIC Density Fits for Relationship 4
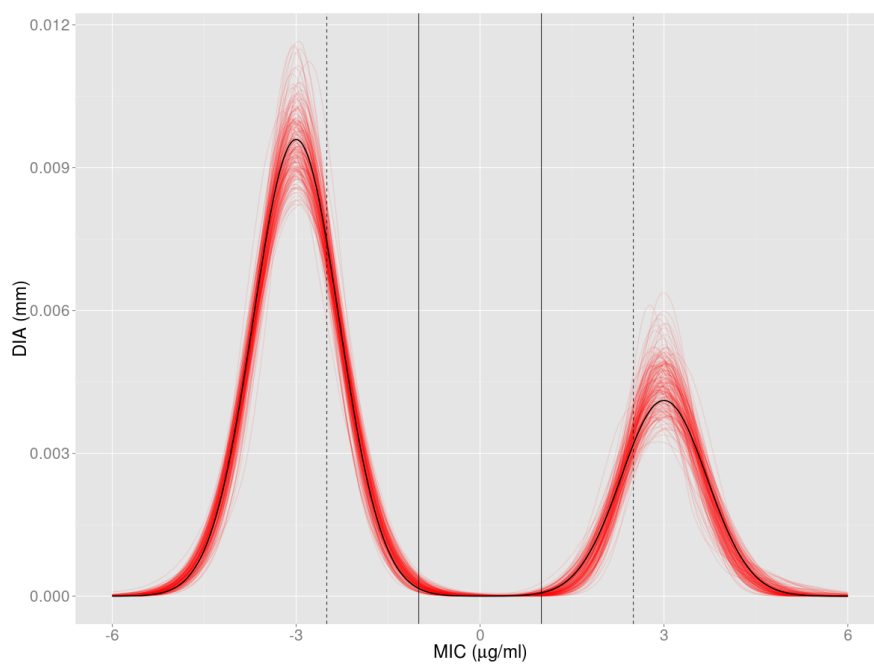
Figure 3.111. Median MIC Density Fits for Relationship 5

Relationship	Density	Mean	SD	Median
1	5	38.26	43.86	21.20
2	5	31.35	38.59	19.37
3	5	98.86	110.39	60.45
4	5	369.60	183.29	338.60
5	5	2023.00	851.33	1945.00
1	4	15.38	12.98	11.09
2	4	14.10	13.09	9.31
3	4	42.90	38.97	29.69
4	4	66.36	45.67	52.98
5	4	97.72	61.76	84.00

Table 3.15 MIC/DIA Fit statistics for the BNP approach when there are few isolates near the indeterminant range (density 5) compared to when the density is right skewed unimodal (density 4).

The BNP approach does a good job of correctly estimating DIA breakpoints when there are few isolates near the indeterminant range. However, when the true relationship differs significantly in this area, as in relationships 4 and 5, the DIA breakpoint estimates are off. This is likely due to the fact there is simply no data in this area, and thus no information, to accurately estimate the underlying MIC/DIA relationship. In situations like these, we recommend setting the DIA breakpoints as conservatively as possible.

3.3.6 BNP Approach for Smaller Sample Size

Condition	
MIC Density	skew right bimodal
MIC/DIA Relationship	all five
σ_m, σ_d	0.707, 2.121
MIC Breakpoints	-1, 1
Number of Pathogen Strains	500
Assessment	DIA Breakpoint Accuracy
	MIC/DIA Relationship Accuracy

Table 3.16 Scenario Conditions and Assessments

Previous simulation scenarios have considered a sample size of 1000 isolates. CLSI guidelines recommend a minimum of 500 isolates should be used in a typical susceptibility experiment. We assess the performance using a sample size of the minimum recommended size, 500, using the BNP, FNP, and L4P approaches. This would be of interest to practitioners since running an experiment with a smaller number of isolates saves money. We show the resulting DIA breakpoint accuracy for all three approaches. We then show the resulting MIC/DIA relationship estimates and the MIC density estimates for the BNP approach.
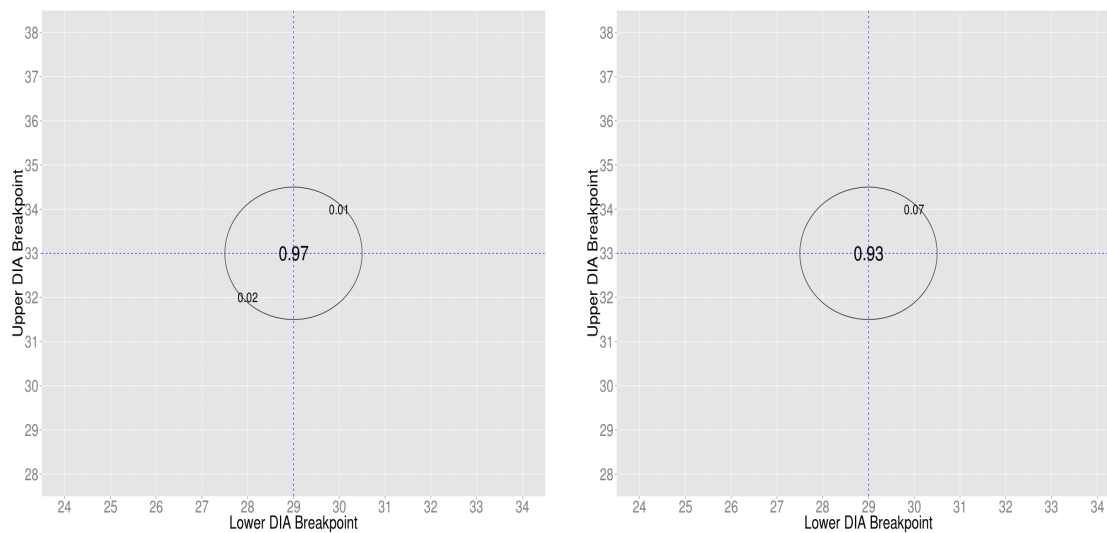
Figure 3.112. L4P (left) and FNP (right) Breakpoints for MIC/DIA Relationship 1
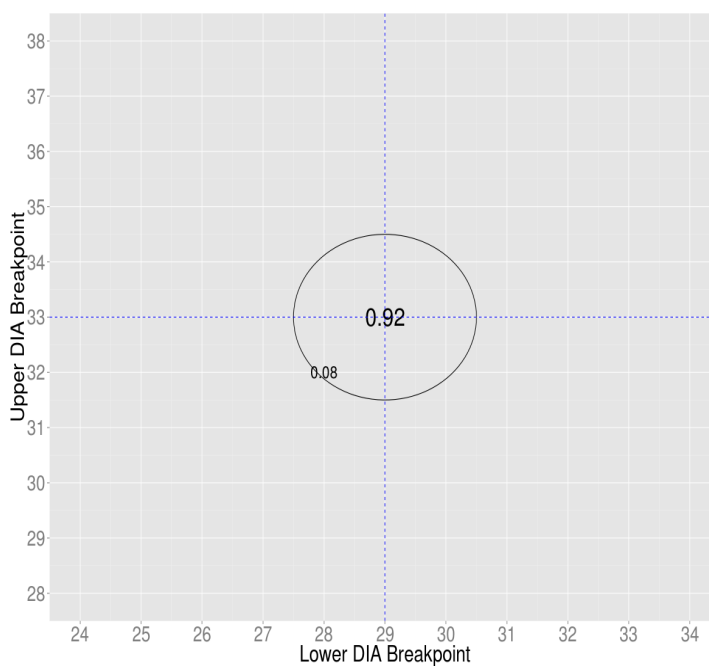
Figure 3.113. BNP Breakpoints for MIC/DIA Relationship 1
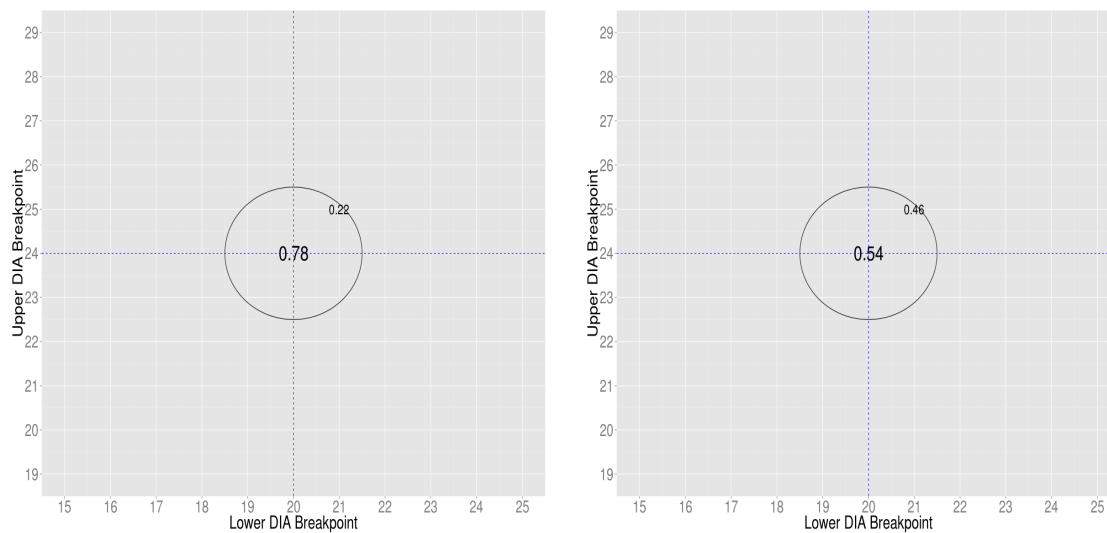
Figure 3.114. L4P (left) and FNP (right) Breakpoints for MIC/DIA Relationship 2
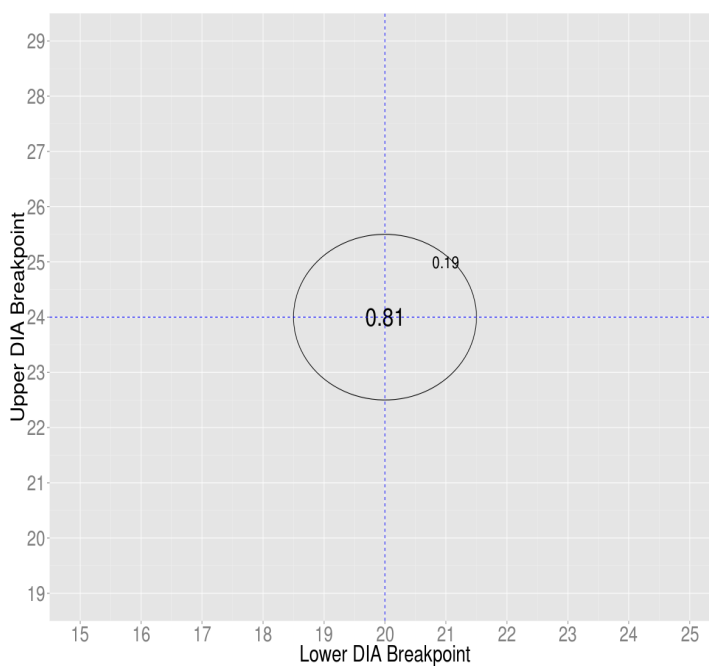
Figure 3.115. BNP Breakpoints for MIC/DIA Relationship 2
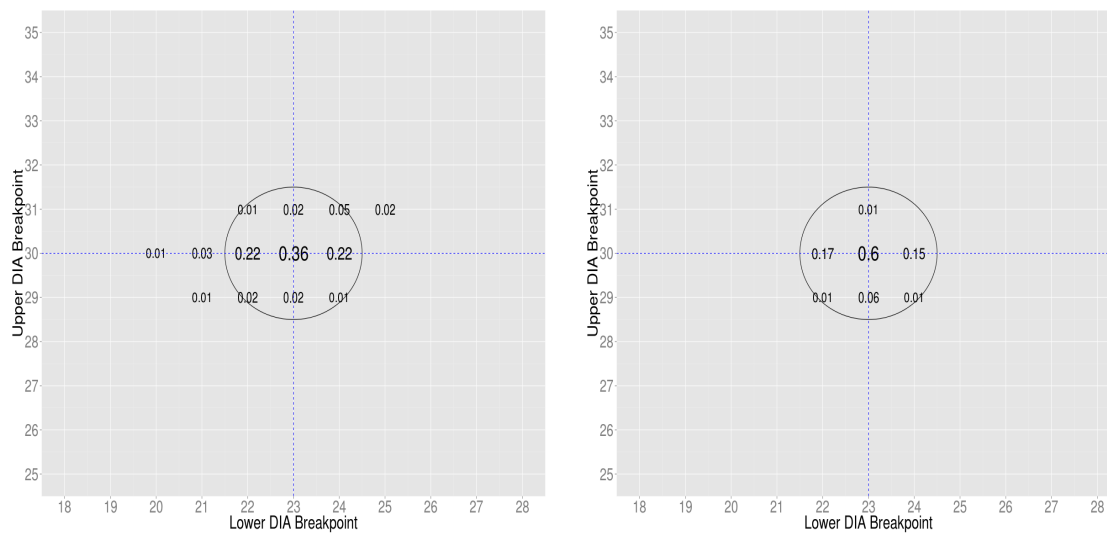
Figure 3.116. L4P (left) and FNP (right) Breakpoints for MIC/DIA Relationship 3
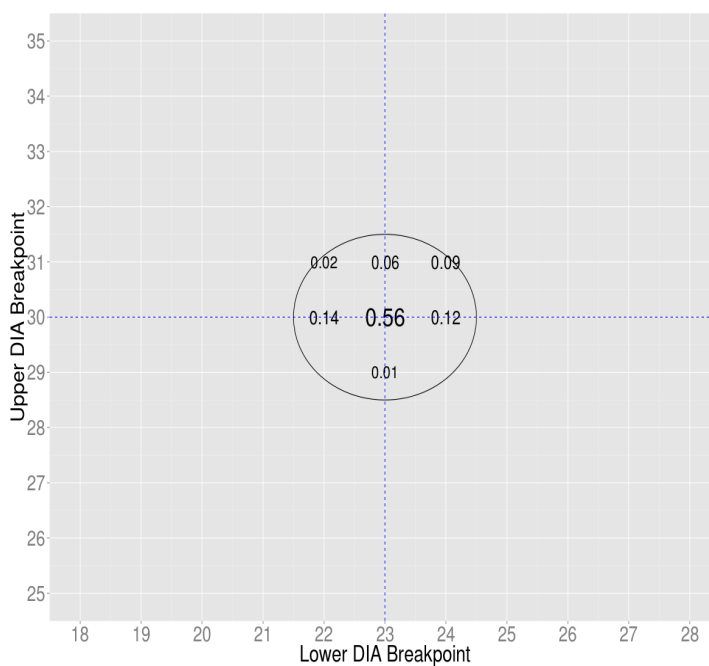
Figure 3.117. BNP Breakpoints for MIC/DIA Relationship 3
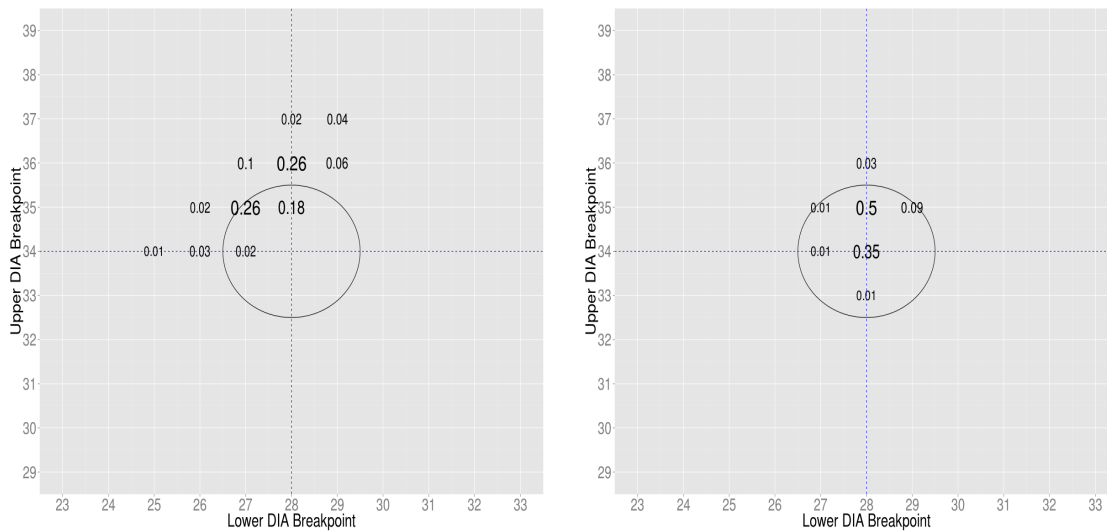
Figure 3.118. L4P (left) and FNP (right) Breakpoints for MIC/DIA Relationship 4
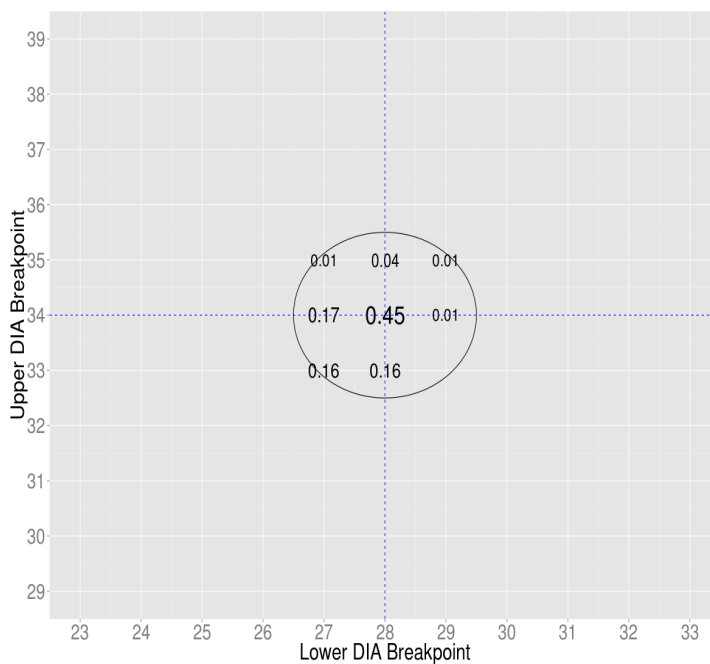
Figure 3.119. BNP Breakpoints for MIC/DIA Relationship 4
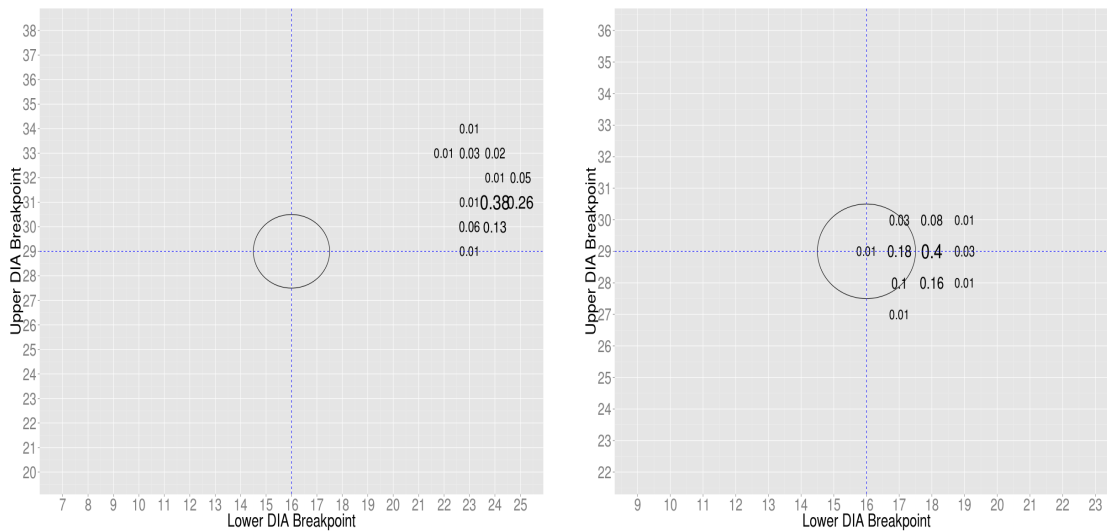
Figure 3.120. L4P (left) and FNP (right) Breakpoints for MIC/DIA Relationship 5
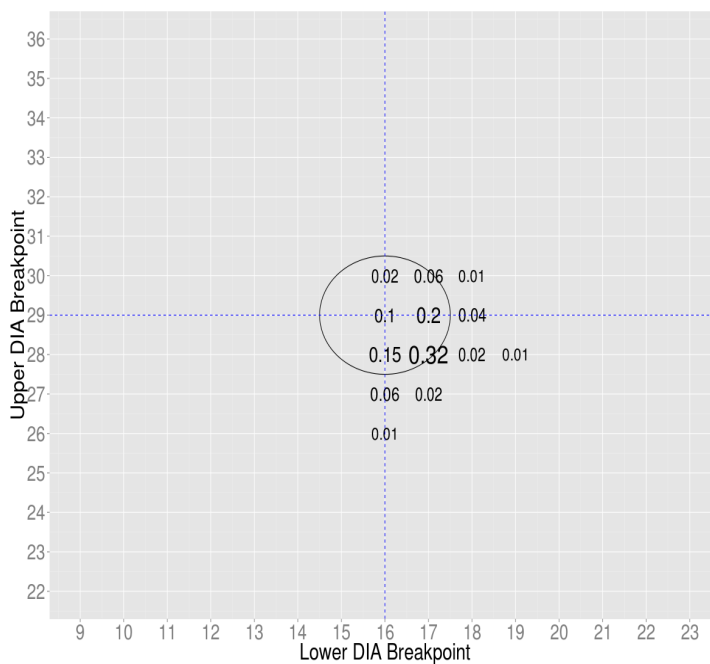
Figure 3.121. BNP Breakpoints for MIC/DIA Relationship 5
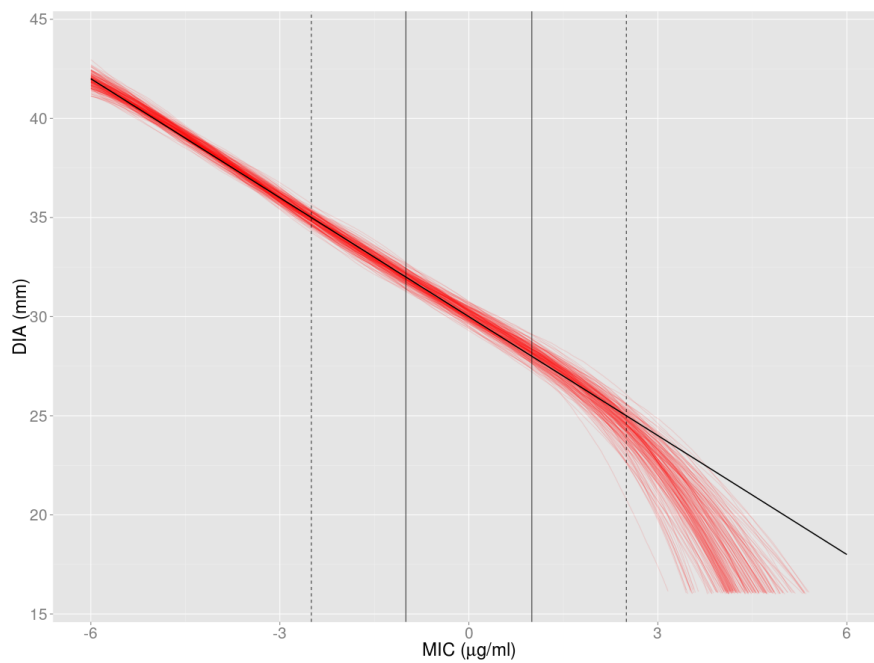
Figure 3.122. Median MIC/DIA Fits (BNP) for Relationship 1
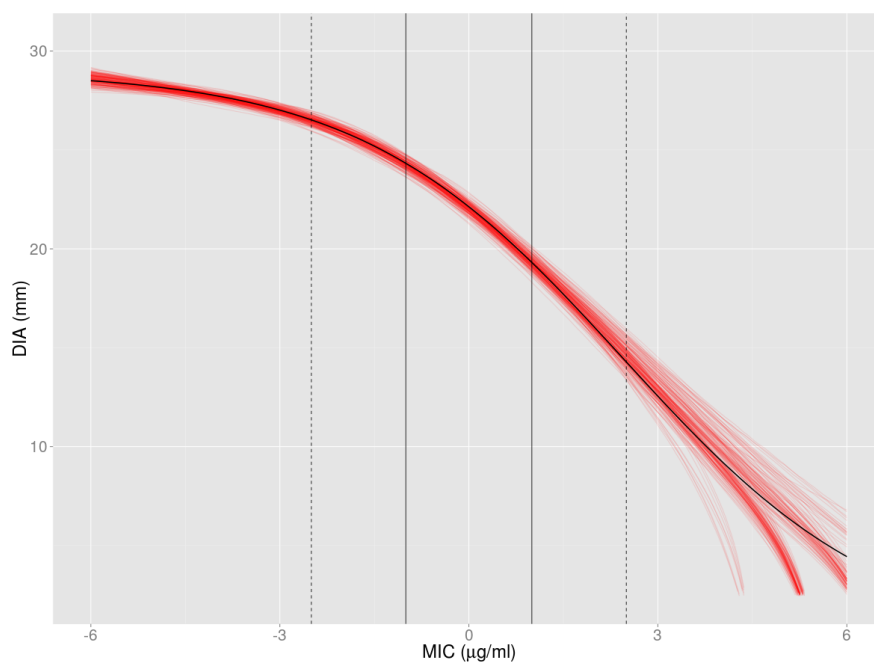
Figure 3.123. Median MIC/DIA Fits (BNP) for Relationship 2
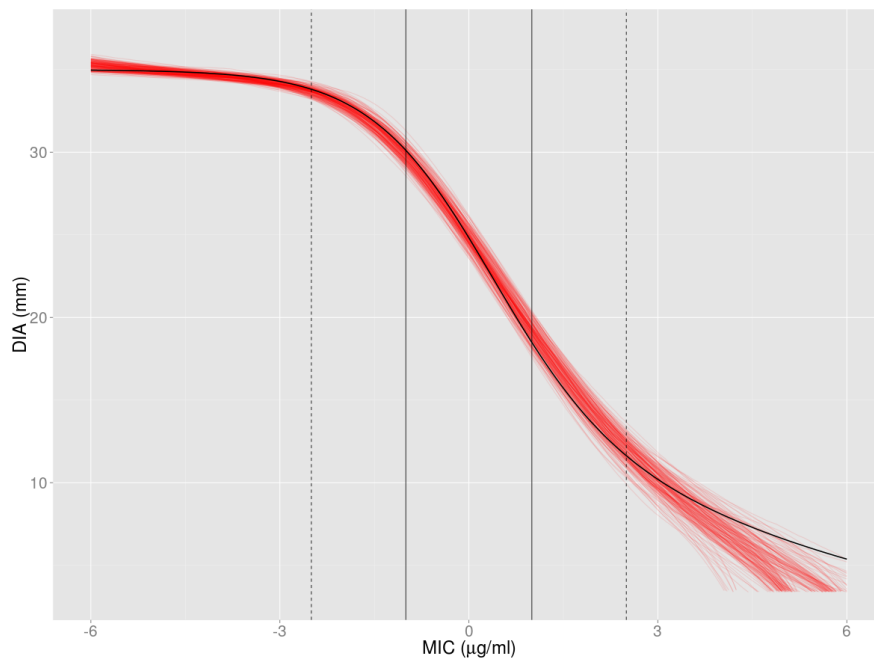
Figure 3.124. Median MIC/DIA Fits (BNP) for Relationship 3
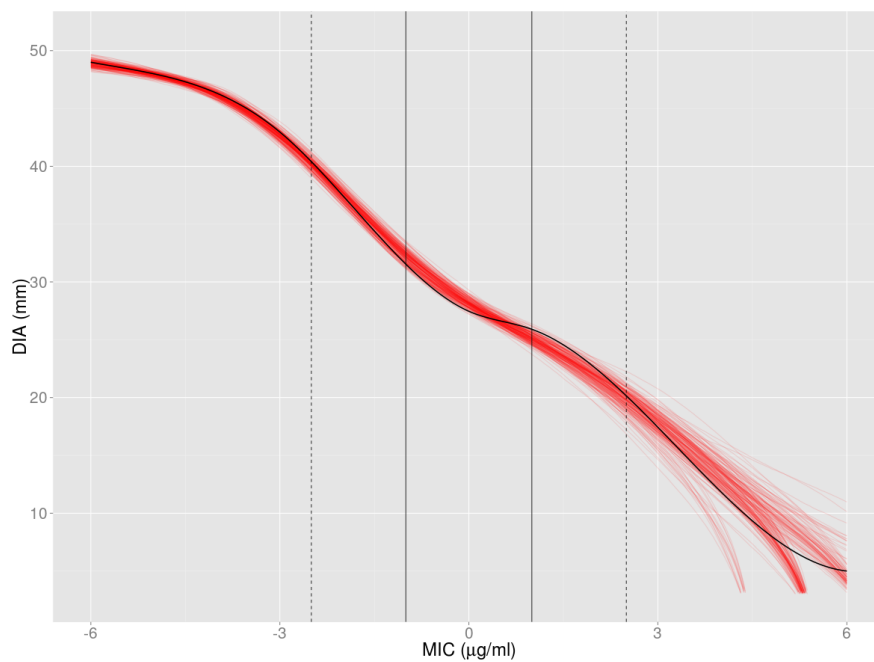
Figure 3.125. Median MIC/DIA Fits (BNP) for Relationship 4
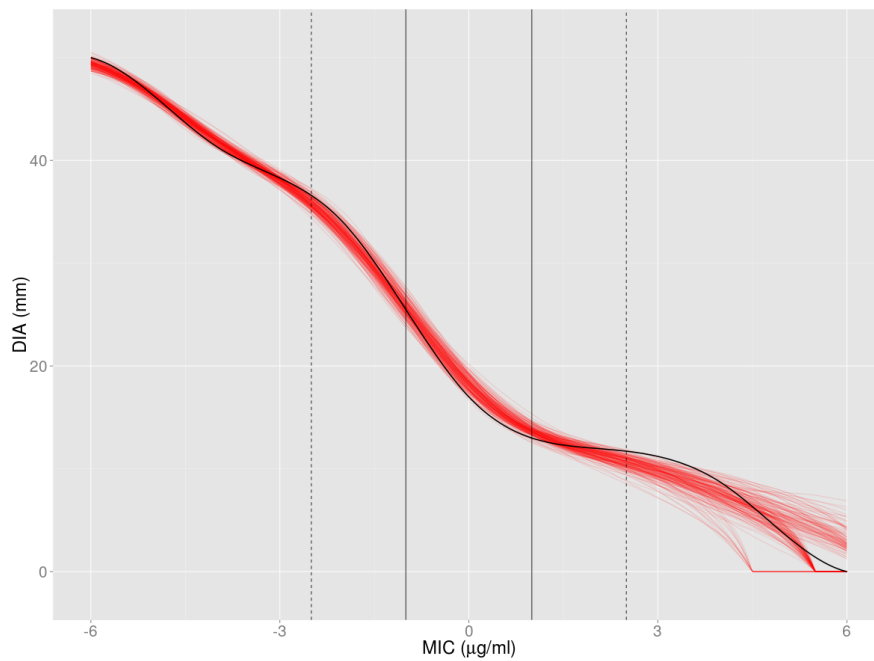
Figure 3.126. Median MIC/DIA Fits (BNP) for Relationship 5
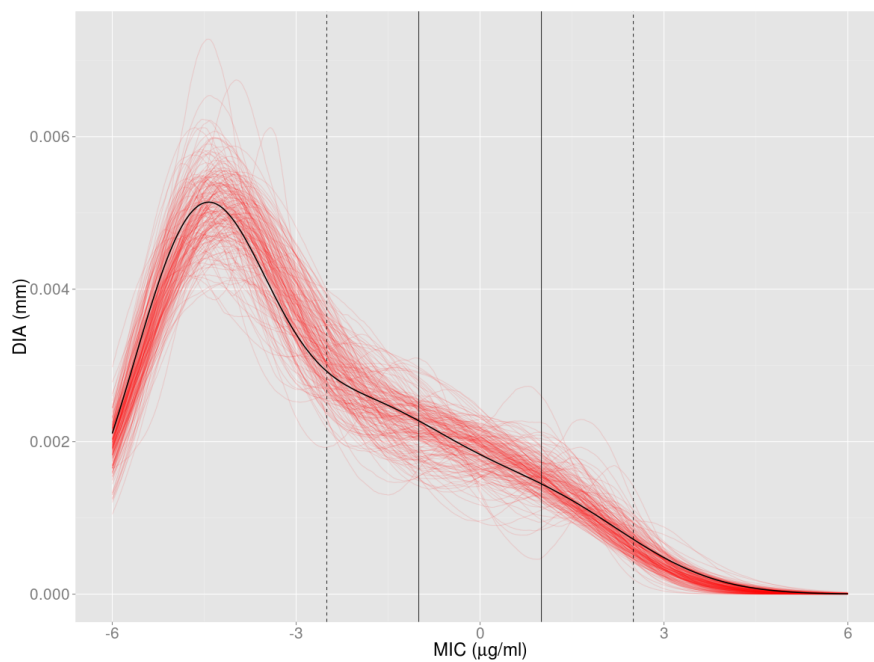
Figure 3.127. Median MIC Density Fits (BNP) for Relationship 1
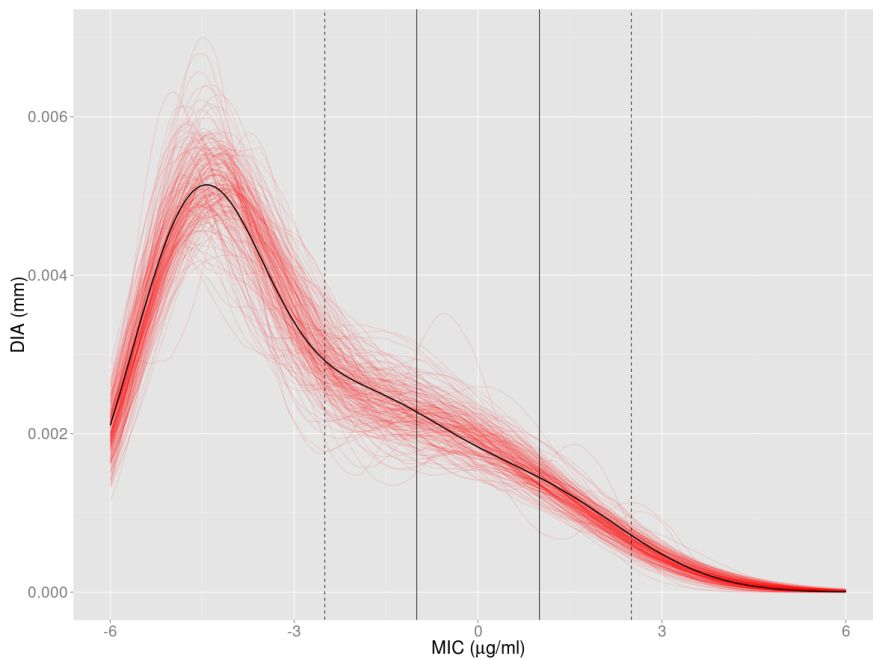
Figure 3.128. Median MIC Density Fits (BNP) for Relationship 2
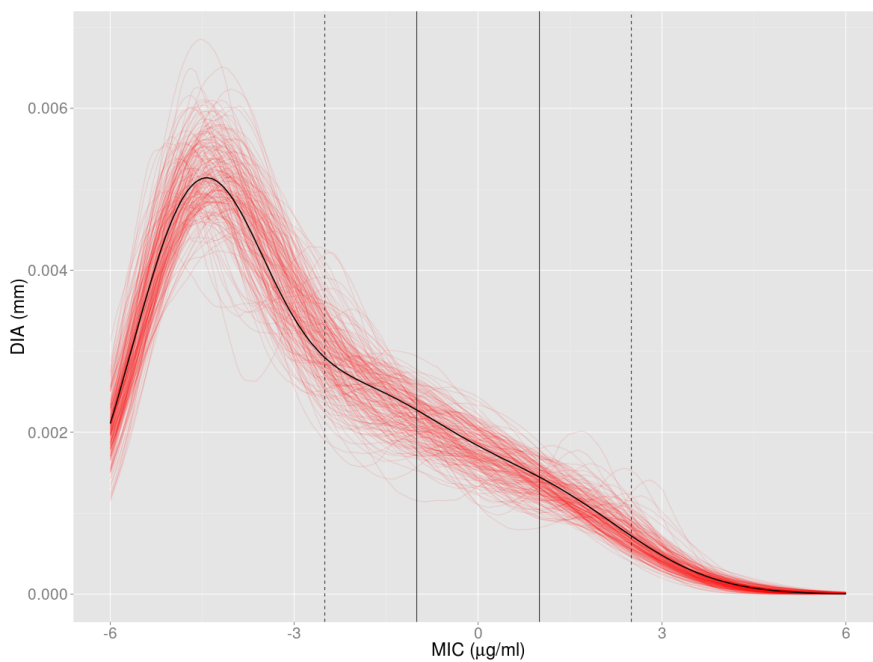
Figure 3.129. Median MIC Density Fits (BNP) for Relationship 3
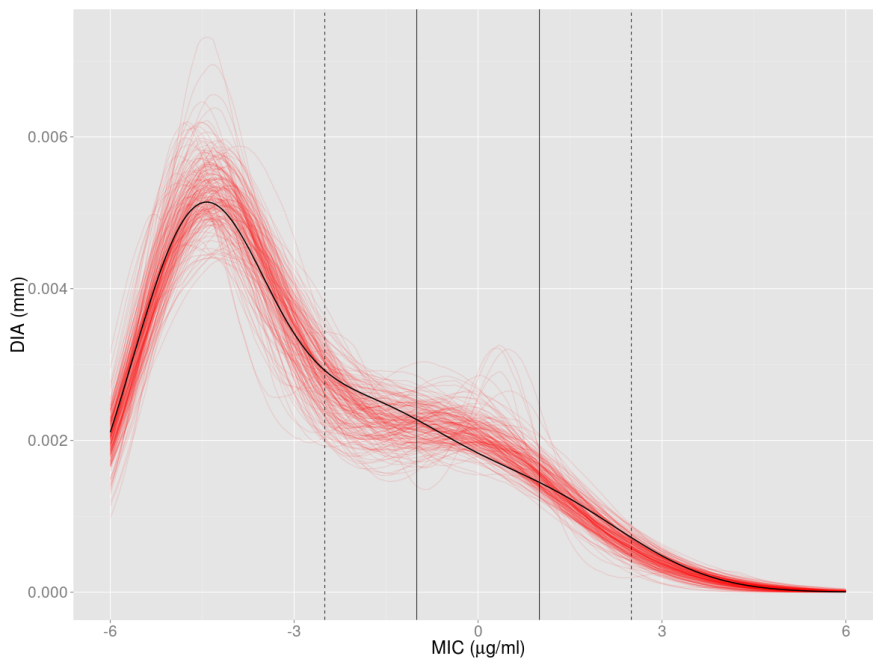
Figure 3.130. Median MIC Density Fits (BNP) for Relationship 4
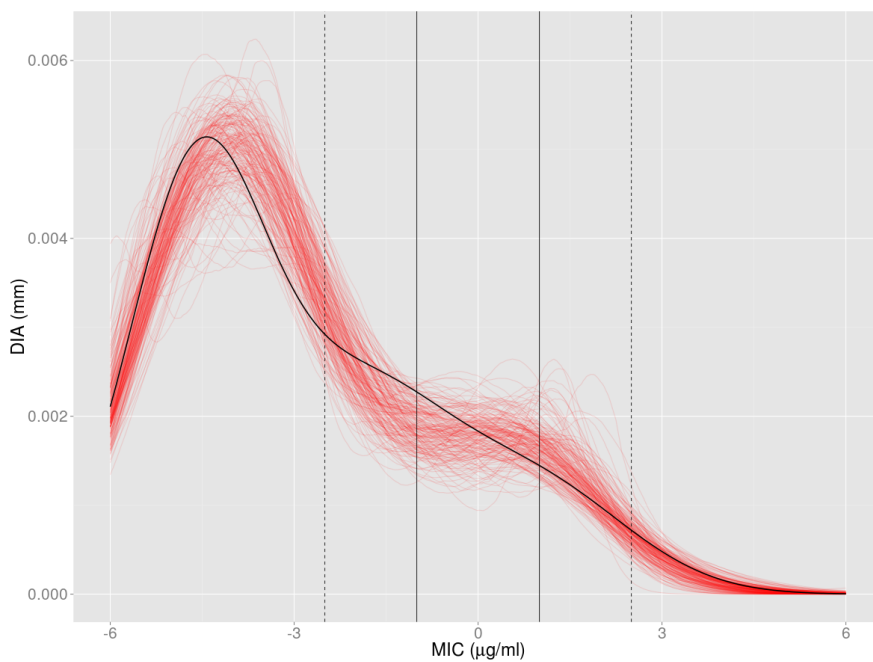
Figure 3.131. Median MIC Density Fits (BNP) for Relationship 5

Breakpoint performance was still very good for all approaches for MIC/DIA relationships 1, 2, and 3. For relationships 4 and 5 the BNP and FNP approach perform well while the L4P approach struggles.

3.3.7 Real Data Sets

We now look at several real data sets. These data sets are unique in that they feature uncommon characteristics of typical susceptibility experimental data. We analyze each data set using the BNP, L4P, and ERB approaches. For the ERB approach we use nonparametric bootstrap to get a distribution of DIA breakpoints. For the FNP approach we only include the estimated DIA breakpoints without any indication of uncertainty (the uncertainty could be quantified by nonparametric bootstrap). It should also be noted that our FNP approach does not take censoring into account, and should be interpreted with caution for these data sets.

Ciprofloxacin and Salmonella

We analyze the data from a susceptibility test involving the drug Ciprofloxacin and pathogen Salmonella. These experimental data were presented at CLSI in 2011. It is unique in that there is an extra large three dilution indeterminant range.

Approach	DIA Breakpoints	Probability/ Proportion
BNP	22, 31	0.65
	22, 32	0.35
L4P	22, 31	0.71
	22, 32	0.29
FNP	22, 30	
ERB	20, 29	0.74
	21, 29	0.23
	20, 28	0.03

Table 3.17 DIA breakpoint estimates for all approaches.

The BNP and L4P approaches give very similar DIA breakpoint distributions. The estimated fits for both the underlying MIC density and underlying MIC/DIA relationship are also very similar. The FNP approach breakpoints are close to the BNP and L4P approaches. The ERB breakpoints are also close to the model-based approaches but lower.

Figure 3.132. Resulting BNP fit. The black line represents the median estimate and the dotted red lines represent 95% credible intervals.

Figure 3.133. Median Density Estimate for BNP

Figure 3.134. Resulting L4P fit. The black line represents the median estimate and the dotted red lines represent 95% credible intervals.

Figure 3.135. Median Density Estimate for L4P

Ertapenem and Enterobacteriaceae

We analyze the data from a susceptibility test involving the drug Ertapenem and pathogen Enterobacteriaceae. These experimental data were presented at CLSI in 2011. It is unique in the underlying MIC density and features heavy censoring.

Approach	DIA Breakpoints	Probability/ Proportion
BNP	18, 22	0.73
	18, 23	0.19
	17, 22	0.08
L4P	17, 21	1
FNP	17, 21	
ERB	17, 21	0.44
	18, 22	0.21
	16, 21	0.11
	17, 22	0.07
	. . .	

Table 3.18 DIA breakpoint estimates for all approaches.

The L4P and FNP approach result in the same DIA breakpoints, however they are different from the BNP and ERB approaches. The L4P approach estimates a linear fit for the underlying MIC/DIA relationship while the BNP approach estimates a fit that dips and flattens in various regions.
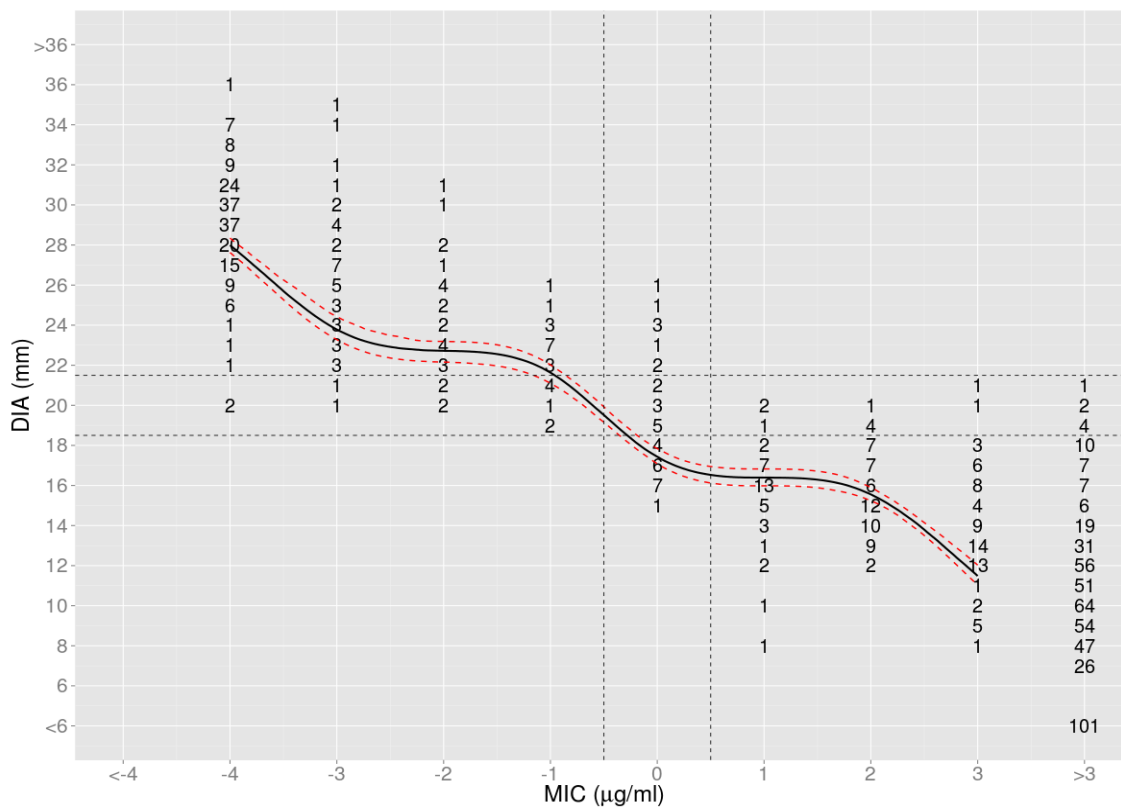
Figure 3.136. Resulting BNP fit. The black line represents the median estimate and the dotted red lines represent 95% credible intervals.
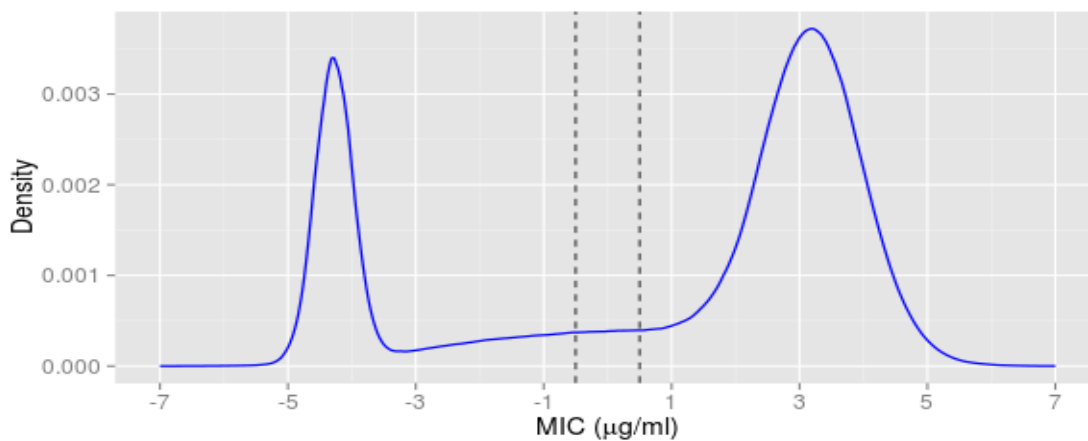
Figure 3.137. Median Density Estimate for BNP
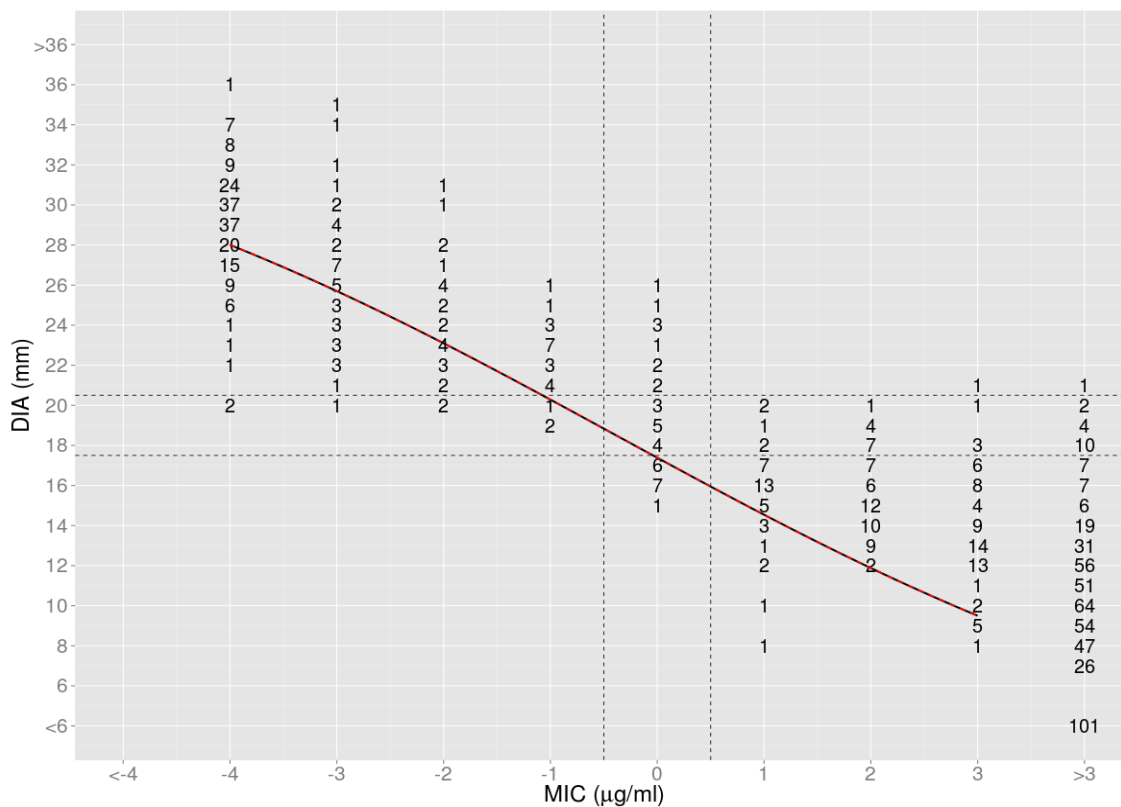
Figure 3.138. Resulting L4P fit. The black line represents the median estimate and the dotted red lines represent 95% credible intervals.
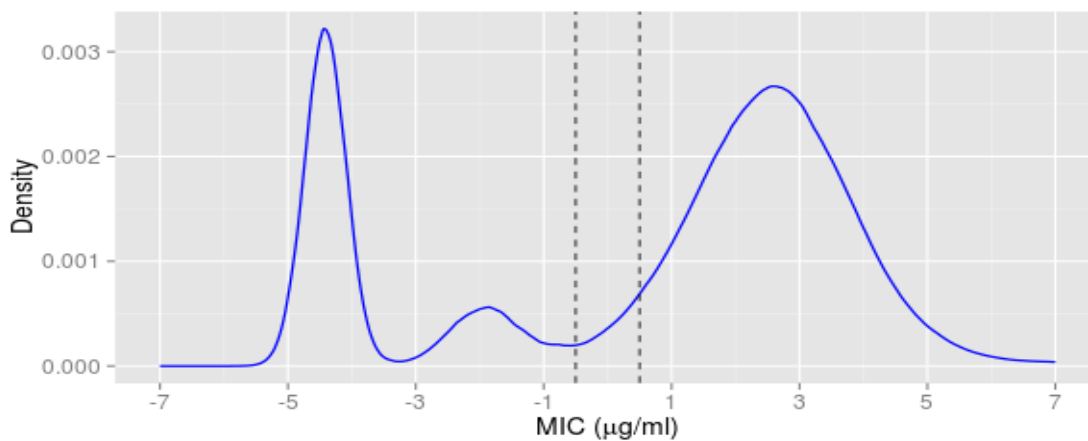
Figure 3.139. Median Density Estimate for BNP

Doripenem and Staphylococcus

This data set has no MIC indeterminant range and also features heavy censoring. We present the results for the BNP, L4P, and ERB approaches.

Approach	DIA Breakpoints	Probability/ Proportion
BNP	18, 22	0.81
	17, 21	0.16
	19, 23	0.03
L4P	25, 29	0.86
	24, 28	0.14
FNP	17, 23	
ERB	16, 25	0.30
	17, 26	0.16
	15, 24	0.13
	22, 31	0.10
	. . .	

Table 3.19 DIA breakpoint estimates for all approaches.

The DIA breakpoint distributions are substantially different for every approach. The BNP approach estimates much smaller breakpoints compared to the L4P approach. Looking at the estimated fit the for the underlying MIC/DIA relationship, it is questionable if a logistic model is appropriate here. The ERB DIA breakpoints are very wide while the FNP breakpoints are closer to the BNP breakpoints.
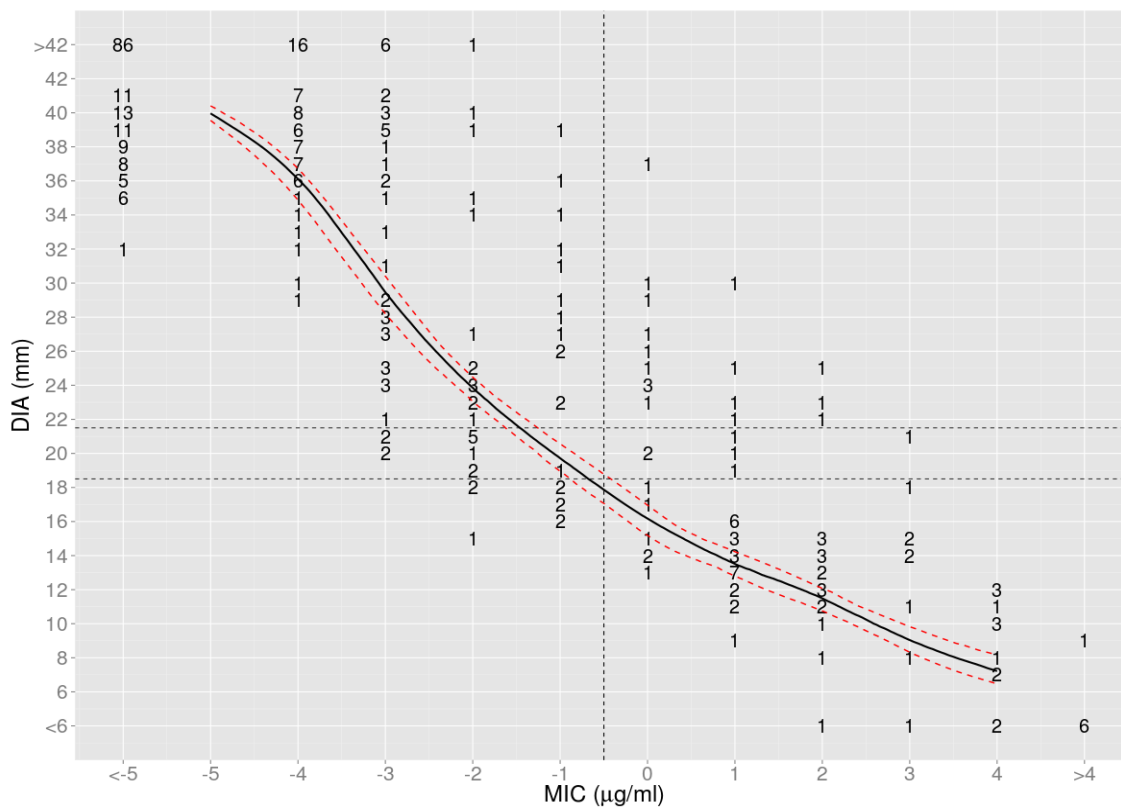
Figure 3.140. Resulting BNP fit. The black line represents the median estimate and the dotted red lines represent 95% credible intervals.
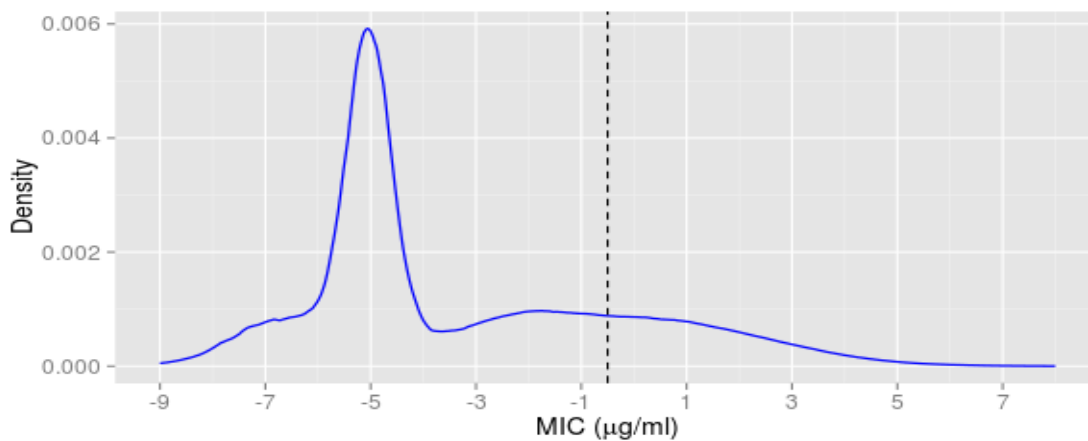
Figure 3.141. Median Density Estimate for BNP
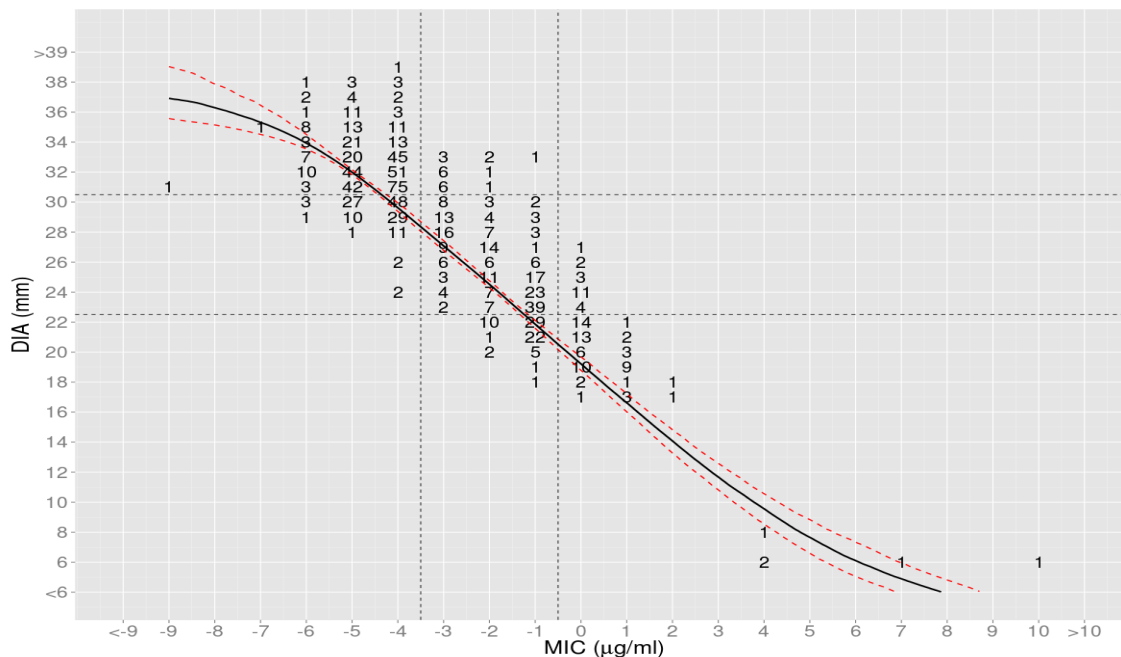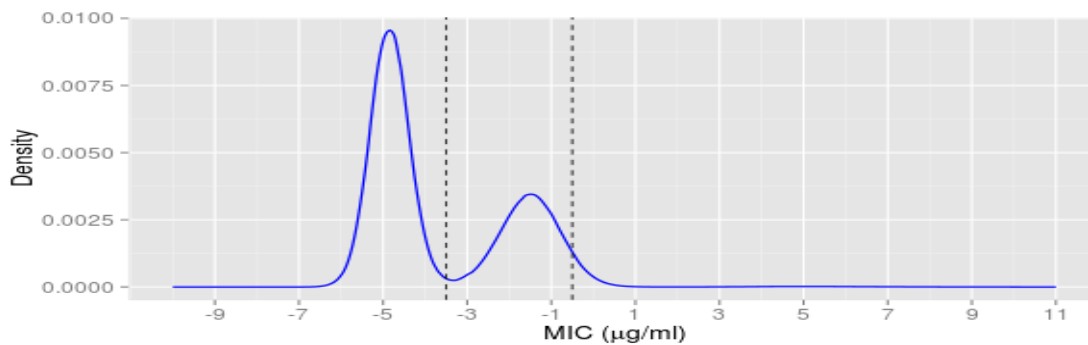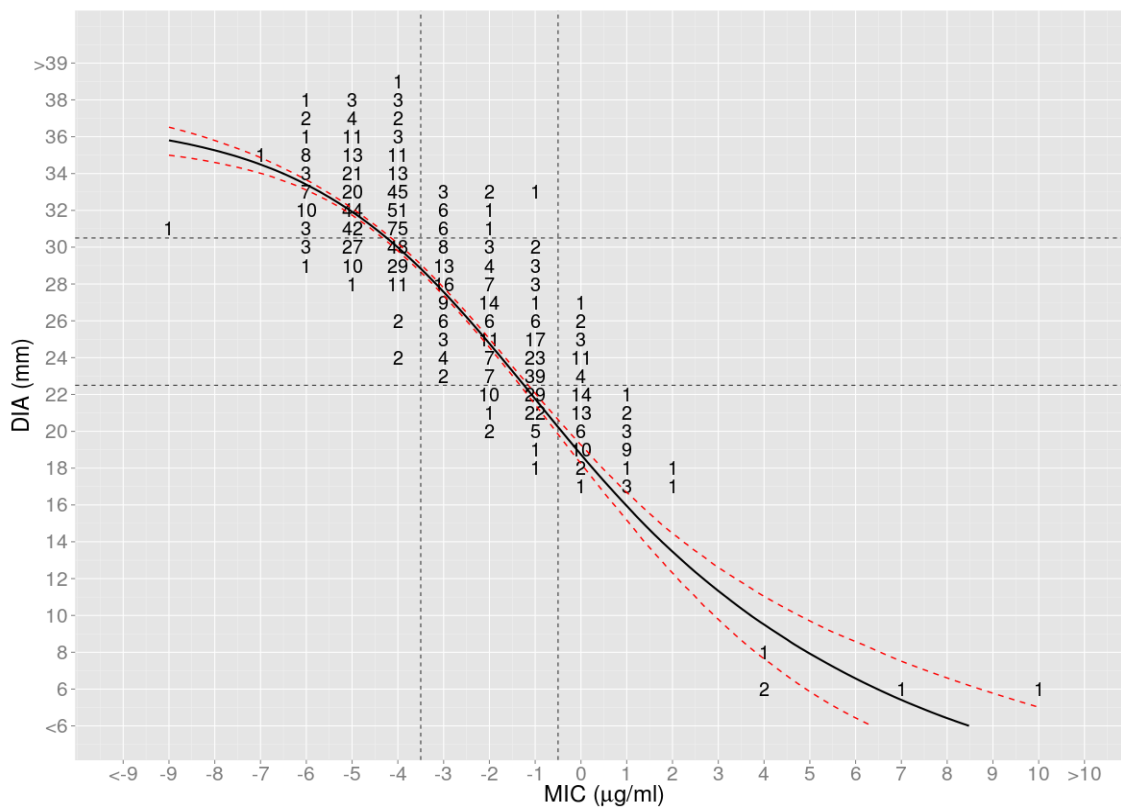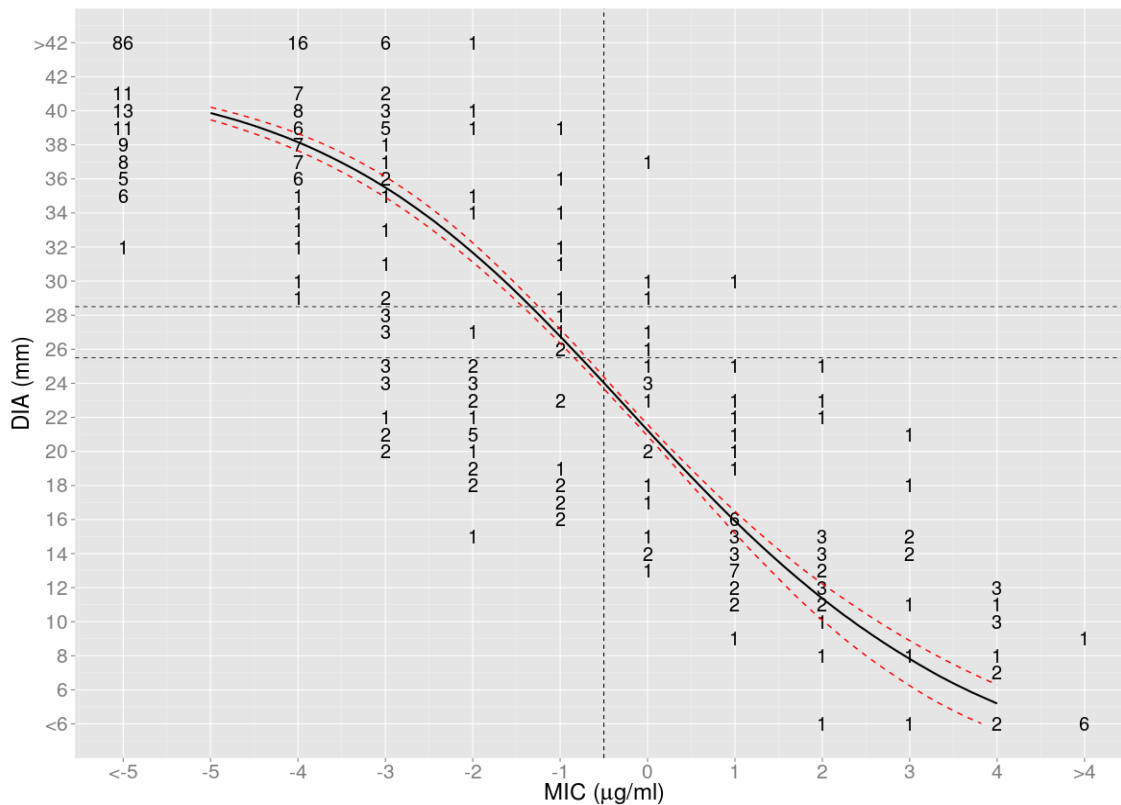
Figure 3.142. Resulting L4P fit. The black line represents the median estimate and the dotted red lines represent 95% credible intervals.
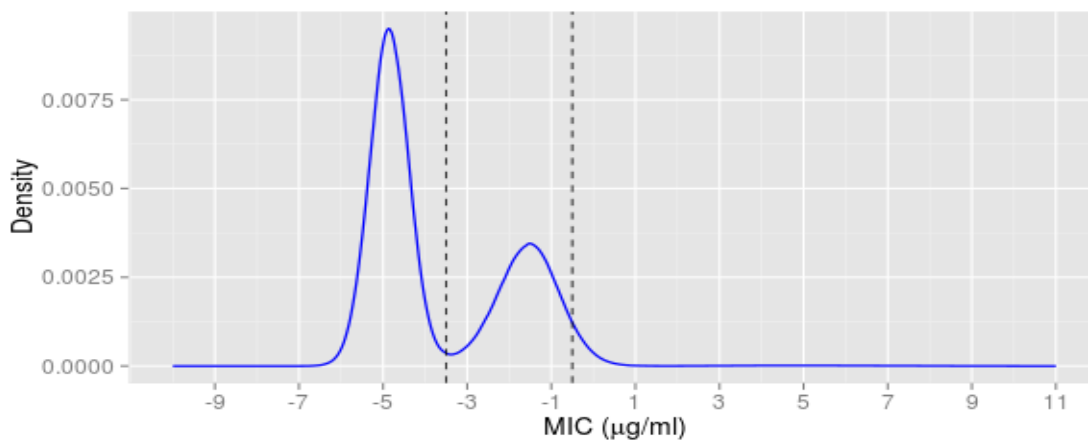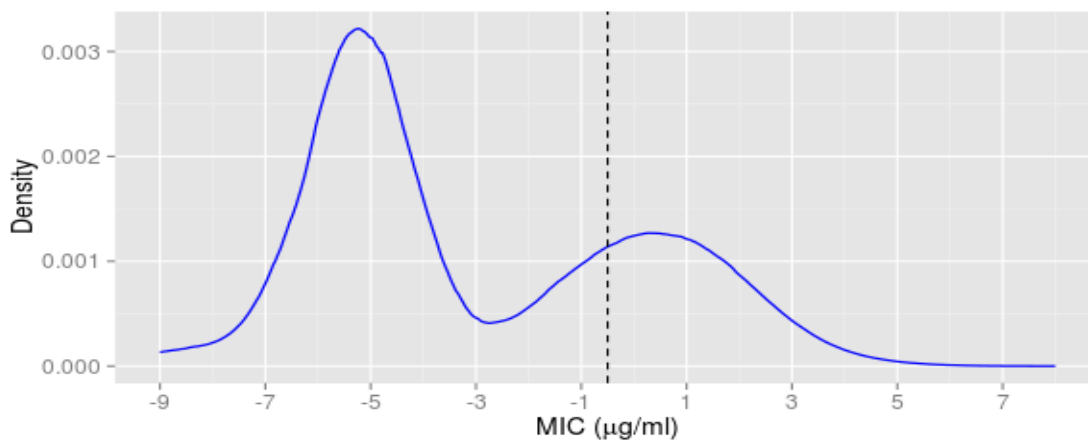

Figure 3.143. Median Density Estimate for L4P

3.4   Summary

The goal of the ERB approach is to minimize observed discrepancies. Therefore the approach does not adequately take into measurement error nor try to get to the underlying truth. This results in very poor and inconsistent performance as demonstrated in the first scenario. When the measurement error was increased the ERB approach performed even worse, while the impact on the BNP approach was minimal. This provides compelling evidence for practitioners to abandon the use of the ERB approach.

The second scenario compared the BNP and FNP approaches by looking at the estimate of the underlying MIC density, MIC/DIA relationship, and DIA breakpoints. The BNP approach did a slightly better job of estimating the underlying MIC density compared to the FNP approach. This performance difference increased as the MIC measurement error increased. The resulting MIC/DIA fits for the BNP approach were also better, but this increased performance was similar between the two approaches in terms of breakpoint accuracy. However, when the MIC breakpoints were shifted towards the right of the data, the BNP approach performed better.

The third scenario compared the BNP and L4P approach. Performance was similar when the underlying MIC/DIA relationship followed a logistic curve, however the BNP approach exceeded the L4P approach when this relationship deviated from a logistic fit.

The fourth scenario showed the two solutions we proposed for knot selection, many equally spaced knots (BNP) or treating the number and location of knots as unknown parameters (BNPRJ), resulted in very similar DIA breakpoint estimation. This is expected since the two approaches produce relatively similar fits for the underlying relationships, although there could be differences.

The fifth scenario showed the BNP approach worked surprisingly well when there were few isolates near the indeterminant range. However, when the underlying MIC/DIA relationship shifts in this range, the BNP approach performs poorly as there is little information in the data to estimate this change. The best option in this case is to choose DIA breakpoints as conservatively as possible.

In the sixth scenario we assessed the performance of the BNP, L4P, and FNP approaches for a smaller sample of isolates. In all previous scenarios the number of isolates was 1000, for this scenario we used only 500. Performance was similar as when there were 1000 isolates. This gives promising results that clinicians can use a smaller number of isolates in practice.

Finally, we applied the BNP, L4P, and ERB approaches to three real data sets. These data sets were unique in that they were uncommon in what would typically be seen in susceptibility experiment data. They featured heavy censoring, indeterminant ranges different than the common one $\log_2$ dilution range, and uncommon MIC densities. The BNP and L4P approaches produced similar results for the first data set (Ciprofloxacin and Salmonella), however the results were quite different for the other two data sets. Looking at the estimated underlying relationship, it appeared that the BNP approach results were more reasonable. For example, with the second data set the L4P approach estimated an exact linear relationship. While this is possible, it is much more likely the underlying relationship differs significantly from a logistic fit.

4. CONCLUSION

Resistance to antibiotics is a world-wide concern. More and more infections that were previously easily treatable have become life-threatening. Humans must address the resistance issue before it is too late. One way to slow down the resistance is through a better choice of antibiotics. This thesis addresses a step in that direction.

Antimicrobial susceptibility testing is used to help clinicians choose the appropriate drug to treat an infection. Clinicians count on this method to give accurate results. The current method to determine breakpoints has been shown to perform poorly. Our goal was to not only propose a better method, but also provide the tools necessary for clinicians to use it in practice.

While we discussed our approach in terms of calibrating antimicrobial susceptibility tests, it can be used in other contexts as well. Data featuring measurement error and/or rounding are common and found in all disciplines. In addition, there is often prior knowledge of a monotonicity constraint. Taking this information into account increases precision by filtering out some of the noise. Our method is the first, that we know of, to take these factors into account in a nonparametric Bayesian framework.

4.1 Summary

We proposed a novel errors-in-variables Bayesian nonparametric model that can be used to estimate DIA breakpoints. This model assumes no parametric relationship between the true MIC and DIA, accounts for the measurement error and rounding in both tests, and utilizes both the observed MIC and DIA results when estimating the underlying true MIC distribution.

Our simulation studies show the model performs well across a variety of scenarios. It performs comparably to the extended four-parameter logistic model when the underlying MIC/DIA relationship follows a logistic relationship, and exceeds it when it does not. The approach also outperforms, in most cases, our two-stage frequentist approach, where the underlying MIC density estimate and underlying MIC/DIA relationship are estimated in two separate steps. In addition, we showed the current method used in practice, ERB, is biased and is not precise. Our approach greatly outperforms the ERB and hopefully these results will encourage clinicians to abandon its use.

4.2 Software

Model-based methods have been proposed in the past but not implemented by clinicians. Perhaps the biggest hurdle in encouraging their implementation was the lack of easy-to-use software. We have created software, available freely on the web, for clinicians to use in practice. We hope this starts a movement to discontinue the use of the ERB method and push towards model-based approaches when determining DIA breakpoints.

dBETS (**d**iffusion **B**reakpoint **E**stimation **T**esting **S**oftware) was developed with the R package shiny [72]. It is currently hosted on a public server and can be accessed freely via the Internet, `http://glimmer.rstudio.com/dbets/dBETS/`. Organizations can also make it available on private networks to have more control of the underlying server.

dBETS software enables clinicians to easily upload susceptibility data and perform analysis via a common tool. Previously, clinicians analyzed susceptibility data with their own software. While everyone was following the same guidelines, it's very likely there were small differences between programs. These differences could result in

different conclusions for the same data across labs. dBETS provides a standard format for all clinicians to use.

dBETS provides a variety tools for clinicians. Data can be uploaded from an online source or via a local machine. The ERB method, and both spline and logistic models are implemented. In addition, various data visualization techniques are provided.

4.3   Future Work

Our approach assumes the DIA be a function of the MIC. However we can flip this around and assume the MIC is a function of the DIA. We can then use our approach to estimate the underlying MIC/DIA relationship, the underlying DIA density, and DIA breakpoints. No matter which way the data are modeled, the resulting MIC/DIA relationship fit and DIA breakpoints should be similar. In fact for our approach the results should be exactly the same, however the results may be different for other approaches. This is something that needs to be investigated.

We also did not incorporate Qi's approach directly into our simulation study; we instead used an approximation. This was partly due to the lack of available software to incorporate his approach easily. It would a more valid comparison to use Qi's approach exactly as described as our FNP approach in the simulation studies. It may be that Qi's approach does better than our approximation. This requires further investigation. This is especially important in data sets that are heavily censored, as seen in the real data sets analysis.

The algorithm for the reversible jump approach could be improved. Occasionally, the chain exhibits poor mixing for several of the parameters. This is likely due to poor proposal distributions. More research is needed to improve the algorithm and improve efficiency.

More investigation is needed in how DIA breakpoint performance varies with sample size. Most experimental data uses around 500 isolates. Except for simulation scenario six, we used 1000 isolates. We believe our approach will show similar performance with a smaller sample size however this is planned future work. In addition we would like to further investigate how our approach performs with different combinations of measurement error and sample size. It would be very beneficial to practitioners if they knew they could achieve the same performance with 300 pathogens compared to 700 pathogens as this will save them money. A simulation based study varying measurement error and sample size would be very informative.

One of the ongoing works will be to continually update dBETS software. dBETS is still in the beta stage of development and ongoing work is being done to improve the application in order for it to be as easy to use as possible for clinicians. In addition, the underlying code can be improved upon to increase performance, both from an application and algorithmic perspective.

The software program dBETS is built on, Shiny, is still in the beta stage of development. We therefore can expect changes to the underlying structure. Because of this, the current code may become outdated and no longer functional. Consistent maintenance of the code is needed to not only ensure the application works, but also utilize future enhancements to make dBETS more productive for clinicians.

LIST OF REFERENCES

LIST OF REFERENCES

[1] SA Waksman. What is an antibiotic or an antibiotic substance? *Mycologia*, 29(5):565–569, 1947.

[2] John Parascandola. The history of antibiotics: a symposium. *American Insitute of the History of Pharmacy*, 5, 1980.

[3] Julian Davies and Dorothy Davies. Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews*, 3(74):417–433, 2010.

[4] R Sykes. Penicillin: from discovery to product. *Pharmacology*, 79(8):778–779, 2001.

[5] F Bosch and L Rosich. The contributions of paul ehrlich to pharmacology: a tribute on the occasion of the centenary of his nobel prize. *Pharmacology*, 82(3):171–179, 2008.

[6] M. Wainwright and H.T. Swan. C.g. paine and the earliest surviving clinical records of penicillin therapy. *Medical History*, 30(1):42–56, 1986.

[7] PM Hawkey. The growing burden of antimicrobial resistance. *J. Antimicrob. Chemother*, 62:Suppl 1:i1–9, 2008.

[8] Stephen Hawser. Surveillance programmes and antibiotic resistance: Worldwide and regional monitoring of antibiotic resistance trends. *Antibiotic Resistance.*

[9] PL Marino. *The ICU book*. Lippincott Williams and Wilkins; Hagerstown, MD, 2007.

[10] TG Slama, A Amin, and SA et al. Brunton. A clinician's guide to the appropriate and accurate use of antibiotics: the council for appropriate and rational antibiotic therapy (carat) criteria. *Am. J. Med*, 118:Suppl 7A (7): 1S6S, 2005.

[11] E Larson. Community factors in the development of antibiotic resistance. *Annu Rev Public Health*, 28:435–447, 2007.

[12] James Gallagher. Antibitoic "apocalypse" warning, January 2013.

[13] Brad Spellberg, John G Bartlett, and David N Gilber. The future of antibiotics and resistance. *N Engl J Med*, 368(4):299–302, 2013.

[14] Dan I Andersson and Diarmaid Hughes. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature Reviews Microbiology*, 8, 2010.

[15] CX Chan, RG Beliko, and MA Ragan. Lateral transfer of genes and gene fragments in staphylococcus extends beyond mobile elements. *Bacteriol*, 193 (15):3964–3977, 2011.

[16] James H Jorgensen and Mary Jane Ferraro. Antimicrobial susceptibility testing: A review of general principles and contemporary practices. *Medical Microbiology*, 49:1749–1755, 2009.

[17] D.H. Annis and B.A. Craig. The effect of inter-laboratory variability on antimicrobial susceptibility determination. *Diagnostic Medicine and Infectious Diseases*, 53(1):61–64, 2005.

[18] D.H. Annis and B.A. Craig. Statistical properties and inference of the antimicrobial mic test. *Statistics in Medicine*, 24:3631–3644, 2005.

[19] Developement of in vitro susceptibility testing cirteria and quality control parameteres; approved guidelines-second edition. 21(7), 2001.

[20] D.M. Metzler and R.M. DeHaan. Susceptibility tests of anaerobic bacteria: Statistical and clinical considerations. *Journal of Infectious Diseases*, 130:588–594, 1974.

[21] Marshall N. Brunden, Gary E. Zurenko, and Barry Kapik. Modification of the error-rate bounded classification scheme for use with two mic break points. *Diagnostic Microbiology and Infectious Disease*, 15 (2):135–140, 1992.

[22] M.N. Brunden, G.E. Zurenko, and B. Kapik. Modification of the error-rate bounded classification scheme for use with two mic breakpoints. *Diagnostic microbiology and infectious disease*, 15(2):135–140, 1992.

[23] Bruce A. Craig. Modeling approach to diameter breakpoint determination. *Diagnostic Microbiology & Infectious Disease*, 36.3:193–202, 2000.

[24] James T. Rudrik, Stephen J. Cavalieri, and Eugene M. Britt. Proposed quality control and interpretive criteria for disk diffusion susceptibility testing with enoxacin. *Journal of Clinical Microbiology*, 21 (3):332–334, 1985.

[25] Bruce A. Craig. Drug dilution vs drug diffusion: Calibrating the two suusceptibilty tests. Technical Report 99-15, Purdue University, Department of Statistics, 01 2010.

[26] Xiaoli Qi. *Nonparametric Calibration of Two Common Susceptibility Tests using Interval-Censored Data with Measurement Error*. PhD thesis, Purdue University, 2008.

[27] John P. Buonaccorsi. *Measurement Error: Models, Methods and Applications*. Chapman and Hall/CRC, 2010.

[28] I.M. Heid, H. Kuchenhoff, J. Miles, L. Kreienbrock, and Wichmann H.E. Two dimensions of measurement error: Classical and berkson error in residential radon exposure assessment. *Journal of Exposure Analysis and Environmental Epidemiology*, 14:365–377, 2004.

[29] R.J. Carroll, D. Ruppert, L. Stefanski, and C. Crainiceanu. *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. Chapman and Hall/CRC, 2006.

[30] P.J. Brown. *Measurement, Regression, and Calibration*. Oxford Science Publications, 1994.

[31] J.R. Cook and L.A. Stefanski. Simulationextrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89:1314–1328, 1994.

[32] R.J. Carroll and L.A. Stefanski. Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. *Statistics in Medicine*, 13(12):1265–1282, 1994.

[33] William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979.

[34] David Ruppert, M.P. Wand, and Raymond J Carroll. *Semiparametric Regression*. Cambridge University Press Cambridge; New York, 2003.

[35] Jeffrey S. Simonoff. *Smoothing Methods in Statistics*. Springer, 1996.

[36] E.A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 9(1):141–142, 1964.

[37] William S. Cleveland. Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistical Association*, 35(1):54, 1981.

[38] T.J. Hastie and R.J. Tibshirani. *Generalized Additive Models*. Chapman and Hall, 1990.

[39] J.H Friedman. Multivariate adaptive regression splines. *Annals of Statistics*, 19(1):1–67, 1991.

[40] Brian S Yandell. Smoothing splines - a tutorial. *The Statistician*, 42:317–319, 1993.

[41] Edward J Wegman and Ian W Wright. Splines in statistics. *Journal of the American Statistical Assocation*, 78:351–365, 1983.

[42] T. Robertson, F.T. Wright, and R.L. Dykstra. *Order Restricted Statistical Inference*. John Wiley and Sons, 1988.

[43] Hari Mukerjee. Monotone nonparametric regression. *The Annals of Statistics*, 16(2):741–750, 1988.

[44] B. Neelon and D.B. Dunson. Bayesian isotonic regression and trend analysis. *Biometrics*, 60(2):398–406, 2004.

[45] Holger Dette, Natalie Neumeyer, and Kay F. Pilz. A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12(3):469–490, 2006.

[46] J.O. Ramsay. Monotone regression splines in action. *Statistical Science*, 3(4):425–461, 1988.

[47] Mary C Meyer, Amber J Hackstadt, and Jennifer A Hoeting. Bayesian estimation and inference for generalised partial linear models using shape-restricted splines. *Journal of Nonparametric Statistics*, 23(4):867–884, 2011.

[48] Scott M. Berry, Raymond J. Carroll, and David Ruppert. Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, 97(457):160–169, 2002.

[49] S. Spiriti, R. Eubank, P.W. Smith, and D. Young. Knot selection for least squares and penalized splines. *Journal of Statistical Computing & Simulation*, 0(0):1–17, 2008.

[50] Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1):267–288, 1996.

[51] Maja Miloslavsky and Mark J. van der Laan. Fitting of mixtures with unspecified number of components using cross validation distance estimate. *Computational Statistics and Data Analysis*, 41(3-4):413–428, 2003.

[52] J.K. Ghosh and R.V. Ramamoorthi. *Bayesian Nonparametrics*. Springer, 2003.

[53] Martyn Plummer. *rjags: Bayesian graphical models using MCMC*, 2013. R package version 3-10.

[54] William Bolstad. *Understanding Computational Bayesian Statistics*. John Wiley and Sons, 2010.

[55] Ronald Christensen, Wesley Johnson, Adam Branscum, and Timothy E. Hanson. *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*. Chapman and Hall/CRC, 2010.

[56] T. Park and G. Casella. The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

[57] P. J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732, 1995.

[58] Faming Liang, Chuanhai Liu, and Raymond Carroll. *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*. Wiley, 2010.

[59] Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo (Chapman & Hall/CRC Handbooks of Modern Statistical Methods)*. Chapman and Hall/CRC, 2011.

[60] W.K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[61] Ilaria DiMatteo, Christopher R. Genovese, and E. Kass, Robert. Bayesian curve-fitting with free-knot splines. *Biometrika*, 88(4):1055–1071, 2001.

[62] X. Wang. Bayesian free-knot monotone cubnic spline regression. *Journal of Computational and Graphical Statistics*, 17(2):373–387, 2008.

[63] Paul G. Gottschalk and John R. Dunn. The five-parameter logistic: A characterization and comparison with the four-parameter logistic. *Analytical Biochemistry*, 343(1):54–65, 2005.

[64] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[65] Alan Genz, Frank Bretz, Tetsuhisa Miwa, Xuefei Mi, Friedrich Leisch, Fabian Scheipl, and Torsten Hothorn. *mvtnorm: Multivariate Normal and t Distributions*, 2012. R package version 0.9-9994.

[66] Simon Urbanek. *multicore: Parallel processing of R code on machines with multiple cores or CPUs*, 2011. R package version 0.1-7.

[67] Tatiana Benaglia, Didier Chauveau, David R. Hunter, and Derek Young. mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software*, 32(6):1–29, 2009.

[68] Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.

[69] Yu-Sung Su and Masanao Yajima. *R2jags: A Package for Running jags from R*, 2012. R package version 0.03-08.

[70] Chris Fraley and Adrian E. Raftery. Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association*, 97:611–631, 2002.

[71] This version Kay Pilz, Steffanie Titoff. Earlier developements by Holger Dette, and Kay Pilz. *monreg: Nonparametric monotone regression*, 2013. R package version 0.1.2.

[72] RStudio and Inc. *shiny: Web Application Framework for R*, 2013. R package version 0.6.0.

VITA

VITA

Glen DePalma was born in Cincinnati, Ohio. He obtained a bachelor's degree in Mathematics and Statistics in 2006 from Miami University. In May 2007, he enrolled into the Statistics Department at Purdue University.

Glen DePalma obtained a Master's degree in Applied Statistics from the Department of Statistics at Purdue University in 2009.