Truncation methods on the SROC curve

TRUNCATION METHODS ON THE SROC CURVE

BY

JING CAI, B.Sc.

A thesis

submitted to the Department of Mathematics & Statistics

and the School of Graduate Studies

of McMaster University

in partial fulfilment of the requirements

for the degree of

Master of Science

Master of Science (2016) McMaster University
(Mathematics & Statistics) Hamilton, Ontario, Canada

TITLE: Truncation methods on the SROC curve

AUTHOR: Jing Cai
B.Sc., (Science & Mathematics)
University of Waterloo, Waterloo, Canada

SUPERVISOR: Dr. Stephen Walter

NUMBER OF PAGES: xviii, 99

To my family

Abstract

The Summary Receiver Operating Characteristic (SROC) curve is a method used to summarize the performance of a diagnostic test using data from a meta-analysis, and the Area Under the SROC Curve (AUC) is a measure of the performance of a diagnostic test. The Partial Area Under the SROC Curve (partial AUC) is sometimes used instead of the AUC, to include only the clinically relevant part. Several truncation methods are used on simulated data to determine the effect of the truncation methods on estimating the properties of partial AUC - mean, bias and standard deviation. Also, when part of the data is truncated before fitting the SROC curve, we examined how the properties of the SROC parameters are affected by the choice of truncation method. The results show that the truncation methods do affect the properties of partial AUC and the estimated SROC parameters. First of all, the estimated values of the partial AUC, with or without scaling, are increased as the value of truncation point increases. The standard deviation of the partial AUC has an increasing relationship with the value of truncation point, and the standard deviation of the scaled partial AUC is in contrast. Truncating a certain percentage of the data performs worse than truncating a part of the SROC curve with respect to the accuracy of estimation. As for estimating the SROC curve parameters a and b, the truncation method which keeps more studies gives more accurate estimation.

Acknowledgements

I will take this opportunity to thank my extraordinary supervisor Dr. Stephen Walter. It was a wonderful experience to work with such a knowledgeable and nice gentleman. Also, I would like to thank the supervisory committee, Dr. Shui Feng, and Dr. Roman Viveros-Aguilera for their instructions, my friend Atinder Bharaj for his kind help, Xuxuan Zhang, Yiliang Zhou, Angela Wang and Mu He for their suggestions and encouragement.

Notation and abbreviations

SROC curve...Summary receiver operating characteristic curve

AUC...Area under the SROC curve

TPR...true positive rate

FPR...false positive rate

Contents

List of Figures

Chapter 1

Introduction

The diagnosis for a disease can be represented in a two-by-two table, where one direction is the true disease stage, which could be a case or a non-case, and the other direction is the test result - positive or negative (Walter, 2002).

Table 1.1: Two-by-Two Contingency Table.

	Case	Non-case
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)
Column Total	n_1=TP+FN	n_2=FP+TN

The ratio of the number of true positive and the number of cases is defined as the true positive rate (TPR) or sensitivity. The ratio of the number of false positive and the number of non-cases is defined as the false positive rate (FPR), which is equivalent to 1-specificity.

$$
\begin{aligned}
TPR &= \frac{TP}{n_1} \\
FPR &= \frac{FP}{n_2}
\end{aligned}
\tag{1.1}
$$

The test result being positive (having the disease) or negative (not having the disease) is determined by the threshold. The threshold can be a numeric value, like the blood pressure. One doctor might diagnose the patients as having high blood pressure if the systolic pressure is greater than 140 mmHg, while another doctor may use 150 mmHg as the threshold. The threshold can also be a personal judgement, like a diagnosis of a dermatosis. Different doctors may make different judgements based on their experience.

Making wrong diagnosis could cause severe results. Giving a non-case a positive test result, which is false positive, will lead to a second test or a non-necessary surgery and treatment. However, the patient who has a disease being tested as negative (false negative) might miss the best treatment time. Our focus is to reduce the mistakes in diagnosis. In other words, we are trying to get as high TPR as possible while keeping the FPR to a low level. Then, the Summary Receiver Operating Characteristic (SROC) curve is introduced to summarize the relationship between the TPR and the FPR (Walter, 2005).

1.1 Area Under the SROC Curve

The Summary Receiver Operating Characteristic (SROC) curve is a method used to summarize the performance of a diagnostic test using data from a meta-analysis (Irwig *et al.*, 1993). On the SROC curve, each point comes from a separate study, and different thresholds may have been applied in each study. Therefore, higher true positive rate (TPR) values do not always associate with higher false positive rate (FPR) values. Then the SROC curve illustrates the relationship between the TPR and the FPR across studies. The SROC curve can be fitted by the model proposed by Moses et al. (Moses and Shapiro, 1993).

First, define

$$\begin{aligned} D &= \ln\left(\frac{TPR}{1-TPR}\right) - \ln\left(\frac{FPR}{1-FPR}\right), \\ S &= \ln\left(\frac{TPR}{1-TPR}\right) + \ln\left(\frac{FPR}{1-FPR}\right). \end{aligned} \tag{1.2}$$

The diagnostic odds ratio is the ratio between TPR/(1-TPR) and FPR/(1-FPR), then the diagnostic log-odds ratio is equivalent to D. The value of S increases as the product of TPR and FPR increases. For a single study, when the diagnostic threshold changes, the TPR and the FPR change monotonically. Therefore, S can be considered as a measure of the threshold. Estimates of D and S are calculated for each study, and a straight line can be fitted by

$$D = a + bS \tag{1.3}$$

using simple linear regression.

If $b \approx 0$, i.e. the value of D does not change with the change of S, or equivalently the log-odds ratio, ln(OR) does not change with the threshold, then the studies are called homogeneous. Otherwise, they are heterogeneous.

TPR can be written as a function of FPR,

$$TPR = \frac{\exp(\frac{a}{1-b})(\frac{FPR}{1-FPR})^{(1+b)/(1-b)}}{1+\exp(\frac{a}{1-b})(\frac{FPR}{1-FPR})^{(1+b)/(1-b)}}. \tag{1.4}$$

The Area Under the Curve (AUC) is achieved by,

$$AUC = \int_0^1 \frac{\exp(\frac{a}{1-b})(\frac{x}{1-x})^{(1+b)/(1-b)}}{1+\exp(\frac{a}{1-b})(\frac{x}{1-x})^{(1+b)/(1-b)}} dx. \tag{1.5}$$

The AUC is a measure of the overall performance of a test. AUC close to 1 indicates a test that discriminates most cases and non-cases correctly. The integration in (1.5) has no closed form.

1.2 Motivation and Partial Area Under the SROC Curve

In disease screening, high FPR corresponds to a large proportion of misclassifying non-cases as cases, and it will lead to a second test for them, which causes waste of medical resources. Therefore, the right-hand part of the SROC curve, where the FPR is particularly large, is considered clinically irrelevant.

As an alternative measure to the full AUC (1.5), the partial area under the SROC curve (partial AUC) is the area of a partial region under the SROC curve. This region could be where data have been observed, or restricted by the range of TPR or FPR which is clinically relevant. The partial AUC is denoted as AUC_{rs}.

$$AUC_{rs} = \int_r^s \frac{\exp(\frac{a}{1-b})(\frac{x}{1-x})^{(1+b)/(1-b)}}{1+\exp(\frac{a}{1-b})(\frac{x}{1-x})^{(1+b)/(1-b)}} dx. \tag{1.6}$$

$$\begin{aligned} \text{var}(\widehat{AUC}_{rs}) \approx & \left(\frac{\partial AUC}{\partial a}\right)^2 \text{var}(\hat{a}) + \left(\frac{\partial AUC}{\partial b}\right)^2 \text{var}(\hat{b}) + \\ & 2\left(\frac{\partial AUC}{\partial a}\right)\left(\frac{\partial AUC}{\partial b}\right) \text{cov}(\hat{a}, \hat{b}). \end{aligned} \tag{1.7}$$

Considering the clinically relevant area of the SROC curve, the partial AUC is calculated by taking the integration from 0 ($r = 0$) to a truncation point (s) defined using various truncation methods. Figure 1.1 shows the full AUC and $AUC_{0,0.2}$. The maximum of the AUC is 1, while the maximum of the AUC_{0s} is s. In order to compare the truncation methods without the effect of s, the scaled partial AUC is defined as,

$$AUC^*_{0s} = \frac{AUC_{0s}}{s}. \tag{1.8}$$

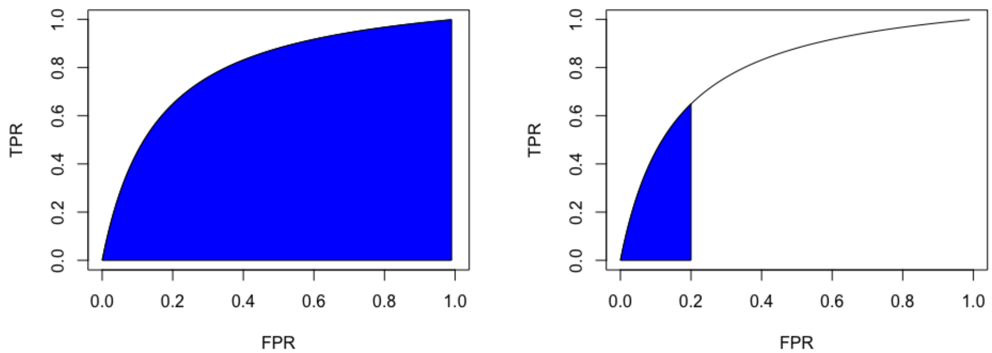

Figure 1.1: Comparison of the full AUC and the partial AUC.

In his paper (Walter, 2005), Walter discussed the numerical evaluations of the partial AUC and the scaled partial AUC, and their standard deviations. It was shown that when taking odds ratio to be 10, as the range restriction (s) becomes smaller, the values of the partial AUC are reduced; additionally, the effect of heterogeneity becomes more evident. The values of partial AUC are decreasing as b changes from -0.4 to 0.4. Under scaling, the decreasing trend of the partial AUC as b increasing is more dramatic.

Concerning the standard error of partial AUC, assuming that $\text{SE}[\ln(OR)] = 1$ and ignoring the terms of $\text{var}(b)$ and $\text{cov}(a, b)$ for simplicity, the standard error of partial AUC is increasing and the standard error of the scaled partial AUC is decreasing as s increases.

This thesis explores the properties of the partial AUC and the scaled partial AUC with the method of simulation. In this thesis, the properties of partial AUC are compared between alternative truncation methods. The mean and standard deviation of partial AUC are estimated by the sample mean and sample standard deviation, and the bias is the difference between the sample mean and the true value which is known. Additionally, the bias and the standard deviation for estimating parameters of the SROC curve, which are $\ln(\mathrm{OR}) := a$ and heterogeneity $:= b$, are also compared.

1.3 Objective

Each dataset consists of 1,000 groups of simulated clinical data from the same procedure which will be discussed in the next chapter. Each group, or say, replication, containing ten studies, will be used to generate an SROC curve. The truncation methods applied in this thesis are classified into two types.

One type of the method, called method A, is executed by fitting the SROC curve then truncating part of the curve. The SROC curve is estimated from the full data using the Moses's model, then it is truncated by eight methods, denoted as A-1, A-2 till A-8. The partial AUC is calculated by equation (1.6). The bias and the standard deviation of the partial AUC can be obtained by replicating this process 1000 times, then the properties of the estimated partial AUC can be compared between methods.

The other type of method, called method B, is truncating part of the data then fitting the SROC curve. Part of data is excluded via eight truncation methods, denoted as B-1, B-2 till B-8, then the SROC curve is fitted based on the remaining part of the data. Therefore, the SROC curve could be different under each truncation method. As well as the properties of partial AUC, the accuracy of estimating a and b are also discussed in order to reveal the effect of excluding data on fitting the SROC curve.

Chapter 2

Methodology

2.1 Data Simulation

The general process of simulating diagnostic data is by creating 10 pairs of TPR and FPR for 10 studies.

2.1.1 Simulation Settings

In the homogeneous case, the odds ratio (OR) is fixed between studies. OR is assigned to be 2, 5, 10 or 20, and the corresponding $a_0 = \ln(OR)$ are 0.693, 1.609, 2.303, 2.996.

The number of studies in each replication is set to be 10. 10 random variables are generated from a certain distribution as the values of specificity (s_2) for each study. Since the range of specificity is (0,1) and a low value of specificity (corresponding to a high FPR) is clinically irrelevant, possible distributions are required to be in the range of (0,1) and left-skewed. Two distributions are chosen to simulate the data.

One of them is the mixed uniform, which is a mixture of two uniform distributions - 8 random variables from Unif (0.5, 1), and 2 random variables from Unif (0, 0.5). The other one is a beta distribution with the shape parameters $\alpha = 5$ and $\beta = 2$. Additionally, a uniform distribution, Unif (0,1), is used in one experiment to check whether the performance of truncation methods is affected when the values of specificity are evenly distributed.

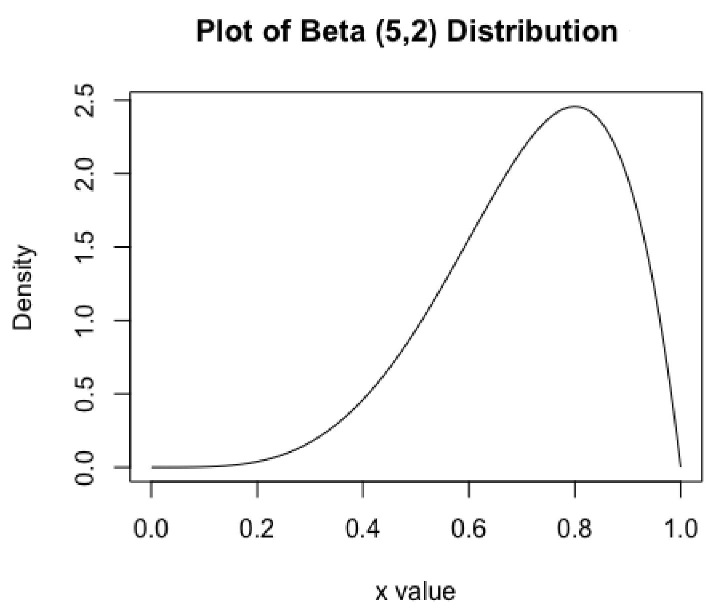

Figure 2.1: Plot of Beta (5,2) distribution.

The numbers of cases and non-cases are set to be n_1 and n_2 respectively, where n_1 and n_2 could be fixed or from a certain distribution.

The choices of the settings are unlimited, and only a limited number of them can be investigated in this thesis. Table 2.1 lists the values that have been explored in this thesis. The number of cases, number of non-cases, distribution of specificity, a (log-odds ratio) and b (heterogeneity) can be considered as the control variables.

Table 2.1: List of the simulation settings.

Experiment	n_1 (Number of the cases)	n_2 (Number of the non-cases)	Distribution of specificity	a_0 (log-odds ratio)	b_0 (heterogeneity)
1	50	200	Mixed-uniform	ln2	0
2	125	125	Mixed-uniform	ln2	0
3	25	225	Mixed-uniform	ln2	0
4	50	200	Mixed-uniform	ln5	0
5	50	200	Mixed-uniform	ln10	0
6	50	200	Mixed-uniform	ln20	0
7	50	200	Beta (5,2)	ln2	0
8	50	200	Uniform	ln2	0
9	Unif(40,60)	Unif(160,240)	Mixed-uniform	ln2	0
10	50	200	Mixed-uniform	ln2	various values $\neq 0$

The first experiment is defined as the base setting. In each experiments from 2 to 10, one of the control variables is changed to discover how these variables affect the performance of the truncation methods. Experiments 1, 2 and 3 vary in the ratio of cases versus non-cases. Experiments 1, 4, 5 and 6 have increasing log-odds ratio. Three different distributions of specificity are applied in Experiments 1, 7 and 8. Experiment 9 involves variability of numbers of cases and non-cases. Experiment 10 tests the heterogeneous studies.

The number of replications is fixed to be 1000, which is confirmed to be sufficiently large to give a satisfactory precision in estimating the partial AUC and the parameters a and b. Also, all experiments have 10 studies per replication.

2.1.2 Simulation Process

For each value of the specificity (s_2) generated from the selected distribution, the corresponding value of the sensitivity (s_1) for each study is calculated by the following formula. The formula is from the relationship between the TPR and the FPR (1.4), and s_1 = TPR, s_2 = 1-FPR.

$$s_1 = \frac{\exp(\frac{a_0}{1-b_0})(\frac{1-s_2}{s_2})^{(1+b_0)/(1-b_0)}}{1+\exp(\frac{a_0}{1-b_0})(\frac{1-s_2}{s_2})^{(1+b_0)/(1-b_0)}} \tag{2.1}$$

For homogeneous case (Experiment 1 to Experiment 9), (2.1) can be simplified as

$$s_1 = \frac{\exp(a_0)(\frac{1-s_2}{s_2})}{1+\exp(a_0)(\frac{1-s_2}{s_2})}. \tag{2.2}$$

The next step is to generate 10 pairs of numbers of true positive (TP) and numbers of false positive (FP) from a binomial distribution.

$$\begin{aligned} \text{TP} &\sim \text{Binomial}(n_1, s_1) \\ n_2 - \text{FP} &\sim \text{Binomial}(n_2, s_2) \end{aligned} \tag{2.3}$$

In case of 0 appearing in any cell in the 2-by-2 contingency table, then 0.5 correction (Fleiss *et al.*, 2003) is applied. Therefore, one two-by-two table is created for each study. Then, 10 pairs of the empirical true positive rate (TPR) and the empirical false positive rate (FPR) can be calculated from this table, displayed in Table 2.2.

Then, n pairs of empirical true positive rate (TPR) and empirical false positive rate (FPR) can be calculated from this table.

Table 2.2: 2 by 2 table for a single study.

	cases	non-cases
positive	TP+0.5	FP+0.5
negative	n_1-TP+0.5	n_2-FP+0.5

$$\begin{aligned} \text{TPR} &= (\text{TP} + 0.5)/(n_1 + 1) \\ \text{FPR} &= (\text{FP} + 0.5)/(n_2 + 1) \end{aligned} \tag{2.4}$$

The above process is replicated 1000 times, and each replication contains 10 studies, which is corresponding to 10 pairs of TPR and FPR. This forms a 2000×10 matrix. Each set of the data varies in n_1, n_2, the distribution of specificity and the values of a_0 and b_0.

2.2 Truncation

For one replication in the simulated data, the SROC curve is generated by the model introduced in Chapter 1. Then the AUC is calculated by (1.5). The partial AUC can be calculated by one of the following methods.

One type of truncation method is fitting the SROC curve by ten pairs of TPR and FPR, then truncating part of the curve. This type is named method A. The SROC curve can be truncated at a fixed point s, and s is between 0 and 1. The value of s is selected to be 0.1 (A-1), 0.2 (A-2), 0.3 (A-3), 0.4 (A-4), 0.6 (A-5) and 0.8 (A-6). (Inside the brackets are the names of the specific methods within method A.) Then the corresponding partial AUC is the partial area under the SROC curve in the range

0 to s. The SROC curve can also be truncated so that it covers a certain percentage of the data. Sorting the studies with increasing FPR, then p% of the data with the highest FPR is excluded. p is selected to be 10 (A-7), 20 (A-8). Since each study within a replication has the same number of cases and non-cases, and there are 10 studies in each replication, then the partial AUC for A-7 is the partial area under the SROC curve in the range 0 to the second highest FPR, and the partial AUC for A-8 is the partial area under the SROC curve in the range 0 to the third highest FPR.

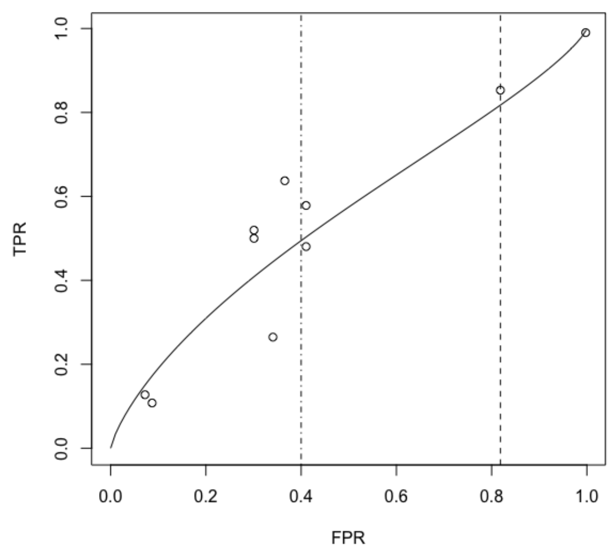

Figure 2.2: Sketch of truncation method A.

Figure 2.2 shows truncation method A, and the curve is fitted using all 10 points. The left vertical dotdash line shows truncating at $s = 0.4$ (A-4), and the right vertical dashed line shows truncating at $p = 10$ (A-7).

Since each method of type A is truncating on the same SROC curve, the truncation method would not affect the estimation of parameters a and b. For truncation methods A-1 through A-6, the partial AUC is recorded for each replication. For truncation methods A-7 and A-8, the partial AUC and the truncation point (the largest FPR in remaining studies) are recorded.

The other type (type B) of truncation method is excluding part of the data, then fitting the SROC curve on the remaining data. This method can also be truncated at a fixed value or a fixed percentage. For each fixed value s. 0.1 (B-1), 0.2 (B-2), 0.3 (B-3), 0.4 (B-4), 0.6 (B-5) and 0.8 (B-6), if there are more than half of the studies, which are five studies, have FPRs in the range of (0, s), then an SROC curve is fitted by the studies in this range, and the partial AUC is the integral of this curve from 0 to s. Otherwise, this replication is considered to be invalid under this method. For each percentage p, 10 (B-7) or 20 (B-8), p% of the data is excluded, then an SROC curve is fitted by the remaining studies, and the partial AUC is the integral of this curve from 0 to the highest FPR after excluding p% of the data.

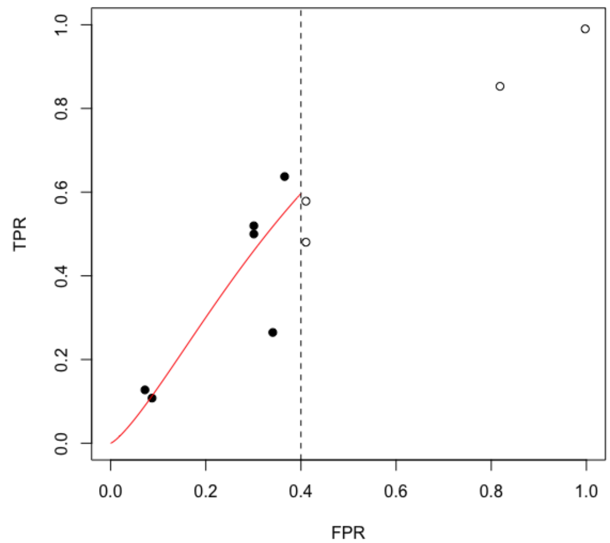

Figure 2.3: Sketch of truncation method B.

Figure 2.3 shows method B-4, as an example. Six out of ten studies have an FPR in the range (0,0.4), then it is valid to be truncated. Studies with an FPR greater than 0.4 are excluded (circles on Figure 2.3), and the SROC curve is fitted by those six studies with FPR less or equal to 0.4 (solid dots). Even though the data is restricted in the range (0,0.4), the fitted SROC curve is defined on the domain (0,1). Here, only part of the SROC curve is plotted, emphasizing that it is fitted by only the red dots. Also, the part of SROC curve not shown is irrelevant to calculate the partial AUC. For the same data shown in Figure 2.3, it is invalid for method B-2; since there are only two studies of which the FPR is in the range (0, 0.2).

For each method of type B, the remaining data could be different. Therefore, a new SROC curve is fitted for each method. Then the estimation of parameters a and

b is affected by the truncation method.

In summary, the truncation method can be categorized as in Table 2.3.

Table 2.3: Summary of the truncation methods.

	fixed point truncation	fixed percentage truncation
Method A (fitting the SROC curve with full data)	A-1, A-2, A-3, A-4, A-5, A-6	A-7, A-8
Method B (fitting the SROC curve with part of the data)	B-1, B-2, B-3, B-4, B-5, B-6	B-7, B-8

2.3 Truncation Methods Comparison

The truncation methods are compared in the following aspects. The first one is the theoretical value of the partial AUC, which is calculated by (1.5), and the values of a and b are from Table 2.1. For the fixed point truncation methods (A-1 to A-6 and B-1 to B-6), the theoretical values are the same for every replication and the average in each method, and the upper bound of the integration is the corresponding s. For the fixed percentage truncation methods (A-7, A-8, B-7 and B-8), s changes in each replication, hence the theoretical values are different for each replication. The theoretical value for each replication is defined as the integration of (1.6) over 0 to its own truncation point s. The theoretical values of the partial AUC for these four methods (generally, not for single replication) were considered to use the expectation of s as the upper bound of integration in (1.5). This works for certain situations. For examples, 10 studies with specificities generated from beta distribution, each study has the same numbers of cases and non-cases. When excluding 10% of the data, the expectation of s is the expectation of the second largest values in these 10 Beta random variables, which can be obtained by the order statistics. However, the expectation of s cannot always be expressed numerically. Therefore, it requires a definition of the theoretical value of partial AUC which works in any situation. The theoretical value of partial AUC for fixed percentage truncation methods is defined to be the average of the theoretical values for each replication.

The second aspect is the distribution and the bias of the partial AUC for each method. Two important statistics are the sample mean and the standard deviation to explore the distribution of the partial AUC. The bias of the partial AUC is obtained by subtracting the theoretical value from the sample mean. Also, the estimated values of the partial AUC from 1000 replications are plotted with a boxplot. 1000 biases for each replication, which are the differences between the estimated partial AUC and its own theoretical value, are also shown in boxplots, so that the effect of the truncation method on the bias of estimating partial AUC can be compared without the effect of the true value.

Another aspect can be observed uniquely for method B. In each method in the type B method, the SROC curve for each replication is generated based on the truncated data. The essential parameters for the SROC curve are a (log-odds ratio) and b (heterogeneity), which are the coefficients of the linear regression in Moses model. In order to have enough data to generate an SROC curve, a valid replication is defined as the following. In one replication, the number of studies with an FPR in the truncation range (0, s) has to be no less than a half of the number of total studies. The valid number is expected to increase as s increases and has maximum value 1,000. The valid number for A-7 and A-8 is 1,000 for sure. Since the data being excluded is fixed to be 20% or 10%, and there are 10 studies with the same number of cases and non-cases for every replication, then the number of studies in the truncation range are correspondingly to be eight or nine. Only valid replications were processed to the following steps. An SROC curve is generated for every method in every replication. Therefore, the estimated parameter a and b are different between methods within one replication, and the effect of truncation method to estimating a and b also need to be compared. Here, similar statistics are applied as for the partial AUC, sample mean, bias, standard deviation and boxplot for visualization.

Let AUC^*_{0s} be the partial AUC scaled by its integration range (0,s),

$$\widehat{AUC^*_{0s}} = \frac{A\hat{U}C_{0s}}{s-0} = \frac{A\hat{U}C_{0s}}{s}. \tag{2.5}$$

For methods A-1 through A-6, mean of the scaled partial AUC equals the scaled mean of the partial AUC since the truncation range s is fixed in each replication. The theoretical value for AUC^*_{0s} is the theoretical value for AUC_{0s} divided by its corresponding truncation point,

$$AUC^*_{0s} = \frac{AUC_{0,s}}{s-0} = \frac{\int_0^s \frac{\exp(a)(\frac{x}{1-x})}{1+\exp(a)(\frac{x}{1-x})}dx}{s}. \tag{2.6}$$

The theoretical values of scaled partial AUC for A-7, A-8, B-7 and B-8 are the averaged ratio of the theoretical partial AUC for each replication and its corresponding truncation point.

Chapter 3

Results

3.1 Truncating on the Base Setting

Define the base setting (Experiment 1 in Table 2.1) to be

- studies per replication - 10
- number of cases in each study - 50
- number of non-cases in each study - 200
- values of specificity follows mixed uniform distribution
- $a = \ln 2$
- $b = 0$

3.1.1 $\widehat{AUC}_{0s}$ Under Method A

In table 3.1, column AUC refers to the full AUC, which can be considered as a partial AUC with range 0 to 1; "p1" refers to the partial AUC under truncation method A-1, and "p2" refers to the partial AUC under truncation method A-2 and so on. The theoretical values in the first row are calculated by taking integration of the theoretical function (3.1) from 0 to s for p1 through p6.

$$\widehat{AUC}^*_{0s} = \frac{\widehat{AUC}_{0s}}{s-0} = \frac{\widehat{AUC}_{0s}}{s}. \tag{3.1}$$

The row mean are the average partial AUC over 1000 replications. Bias is the difference between the mean and theoretical value, and SD is the standard deviation of the partial AUC.

Table 3.1: Properties of $\widehat{AUC}_{0s}$ under method A.

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.316	0.162
Mean	0.010	0.035	0.075	0.127	0.259	0.423	0.612	0.315	0.161
Bias ($\times 10^{-3}$)	0.169	0.038	-0.199	-0.419	-0.803	-1.363	-2.009	-1.167	-0.627
SD ($\times 10^{-2}$)	0.274	0.649	0.968	1.219	1.646	2.071	2.316	10.069	3.492

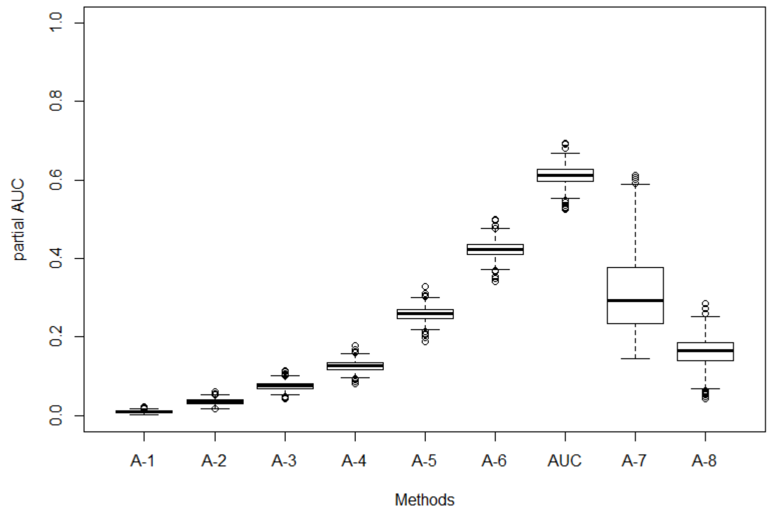

Figure 3.1: Boxplot showing the distribution of $\widehat{AUC}_{0s}$ under method A.

Table 3.1 shows that the mean of partial AUC increases and the standard deviation increases as s increases (Methods A-1 to A-6 and full AUC). Figure 3.1 is the boxplot of $\widehat{AUC}_{0s}$ under method A, where the middle line in each box is the median and the range of box is the first quartile to the third quartile. Table 3.1 and Figure 3.1 show similar pattern of how the partial AUC changes with a fixed truncation point. For methods A-7 and A-8, the truncation point in each replication varies, which results a greater standard deviation.

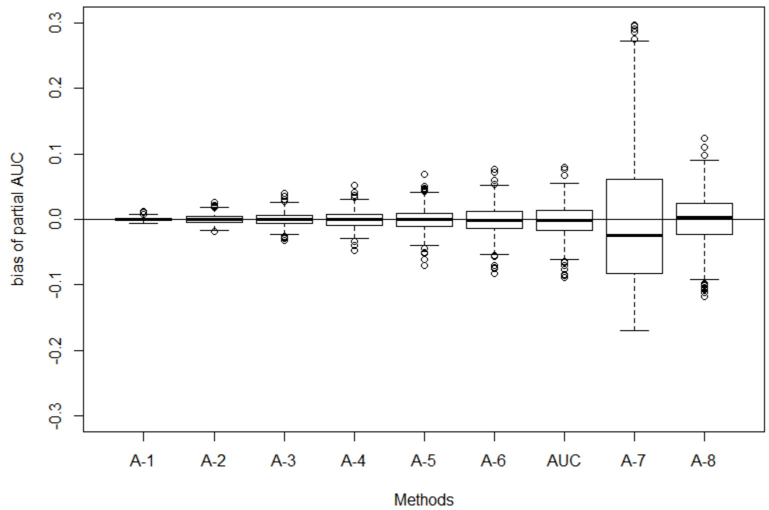

Figure 3.2: Boxplot showing the distribution of bias ($\widehat{AUC}_{0s}$) under method A.

From Table 3.1 and Figure 3.2, bias ($\widehat{AUC}_{0s}$) under methods A-1 through A-6 is approximately centered at 0, which implies that the simulation process estimates partial AUC is estimated unbiasedly in the fixed point truncation methods. Considering the absolute values of bias for method A-7 and A-8, they are not significantly larger than the absolute values of bias for method A-1 to A-6. However, from Figure 3.2, the differences between $\widehat{AUC}_{0s}$ and AUC_{0s} are not symmetrically distributed for method A-7. The range of the bias is wider than the range for A-8, and the range for A-8 is greater than the range for other methods. Therefore, truncating at a fixed point of the SROC curve performs better than truncating a fixed percentage of the data with respect to an unbiased and precise estimate.

3.1.2 $\widehat{AUC}^*_{0s}$ Under Method A

AUC^*_{0s} is the partial AUC scaled by its integration range (0, s), as shown in (2.4). In this section, it is examined that the effect of the truncation methods on the partical AUC when scaling is involved. In Table 3.2, all the row names and column names are the same as in Table 3.1, and the values in row theoretical are the theoretical partial AUC calualated via the method introduced in Chapter 2.

Table 3.2: Properties of $\widehat{AUC}^*_{0s}$ under method A.

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.464	0.351
Mean	0.095	0.177	0.250	0.317	0.432	0.529	0.612	0.462	0.350
Bias ($\times 10^{-3}$)	1.692	0.193	-0.664	-1.047	-1.338	-1.703	-2.009	-1.756	-1.096
SD ($\times 10^{-2}$)	2.742	3.247	3.225	3.047	2.744	2.589	2.316	6.363	4.308

Similar to the AUC^*_{0s}, the estimated AUC^*_{0s} in 1000 replications and the differences between the estimated AUC^*_{0s} and the theoretical values for AUC^*_{0s} are visualized as boxplots.

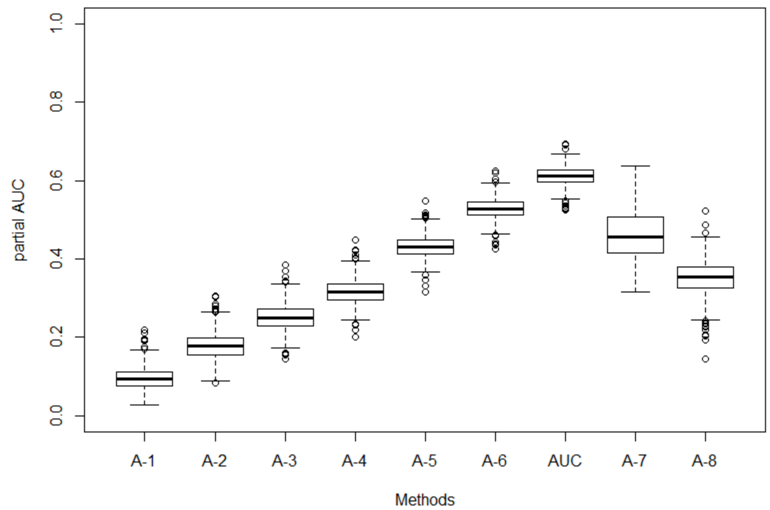

Figure 3.3: Boxplot showing the distribution of $(\widehat{AUC}^*_{0s})$ under method A.

From table 3.2 and figure 3.3, the estimated partial AUC $\widehat{AUC}^*_{0s}$ increases as s increases with scaling, while the standard deviation and the range become more constant as the fixed truncation point s increases. Specifically, except for the first one, the standard deviation shows a slightly decreasing relationship with s. The scaling increases the standard deviations for method A-1 to A-6 and A-8, and it decreases the standard deviation for method A-7.

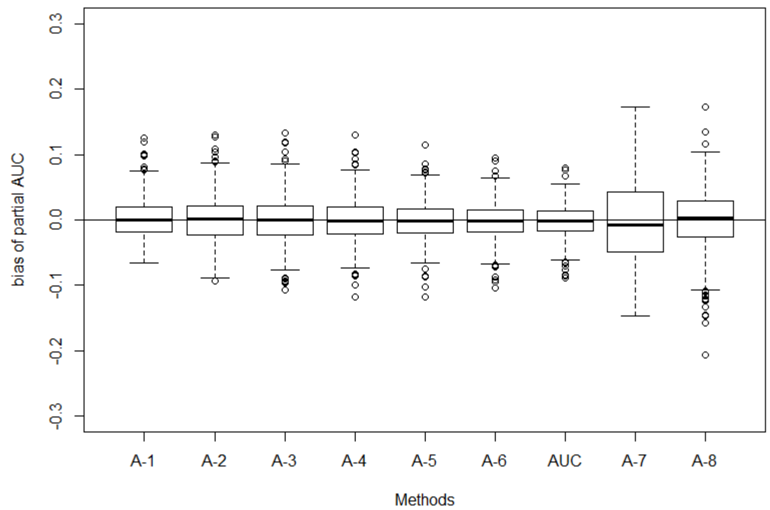

Figure 3.4: Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0s}$) under method A.

The boxplot of bias of the scaled partial AUC displayed in Figure 3.4 also shows that the scaling makes a difference to the standard deviation of $\widehat{AUC}^*_{0s}$. The standard deviation is no longer increasing as s increases. In the fixed percentage truncation method, the variation in the truncation point increases the variation in the partial AUC. The range of the bias for A-7 and A-8 are larger than the range of the bias for the fixed point truncation method even with scaling.

3.1.3 $\widehat{AUC}_{0s}$ Under Method B

Instead of applying all the truncations on the same SROC curve (method A), the partial AUC can also be obtained by calculating the partial area under the SROC curve which is fitted by only part of the studies (method B). Selection of the studies and determination of the truncation point follow the procedure described in Chapter 2.

Table 3.3: Properties of $\widehat{AUC}_{0s}$ under method B.

row.names	p1	p2	p3	p4	p5	p6	AUC	p7	p8
Valid	10	158	602	938	1000	1000	1000	1000	1000
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.316	0.162
Mean	0.011	0.037	0.078	0.130	0.262	0.425	0.612	0.317	0.162
Bias ($\times 10^{-3}$)	1.329	1.866	2.584	2.463	1.693	0.914	-2.009	1.119	0.208
SD ($\times 10^{-2}$)	0.110	0.633	1.083	1.278	1.631	2.034	2.316	10.173	3.444

Table 3.3 gives the valid number of replications. For the fixed point truncation method, a replication is defined to be valid if it contains a half or more studies which have FPR in the range (0, s). For example, the valid number for the partial AUC in method B-1 (p1) is 10, which means that only 10 out of 1000 replications that there are 5 or more studies have FPR in the range (0, 0.1) in this replication. As for the full AUC, the partial AUC in method B-7 (p7) and the partial AUC in method B-8 (p8), the SROC curve is designed to be fitted by the full data, 90% of the data or 80% of the data; i.e. every replication is valid. The values of partial AUC and its standard deviation has an increasing relationship of s as in method A.

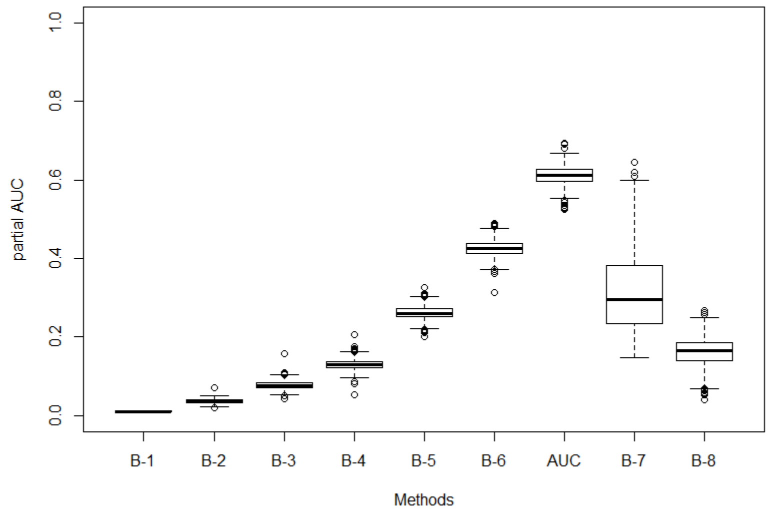

Figure 3.5: Boxplot showing the distribution of $\widehat{AUC}_{0s}$ under method B.

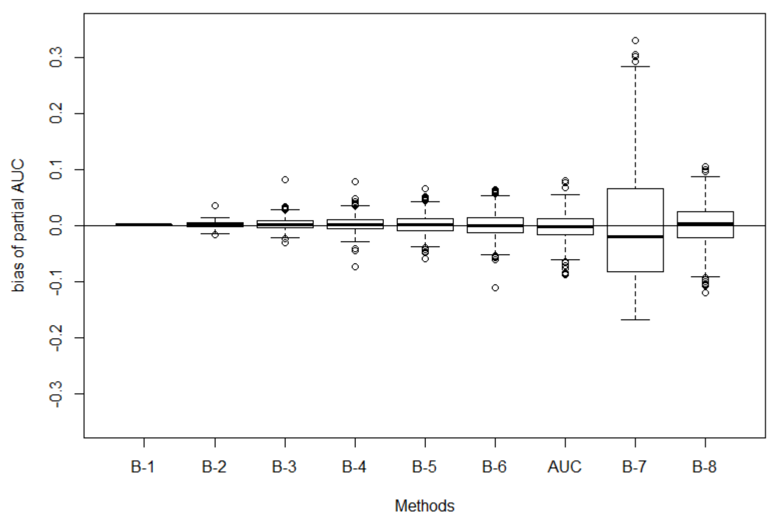

Figure 3.6: Boxplot showing the distribution of bias ($\widehat{AUC}_{0s}$) under method B.

Compared with method A, the absolute values of bias for p1 to p5 in method B are greater. Since only part of the data is used to fit the SROC curve, the fitted curve could be far away from the true curve. For p6, p7, p8 and AUC, most (or all) data is used to fit the SROC curve, then the absolute values of bias are similar for method A and B. Method B-1 and B-2 give small standard deviations of the partial AUC, however, the valid numbers are so small that methods B-1 and B-2 and meaningless.

For each method in method B, the SROC curve is fitted with different set of the data points, i.e. the simulated studies are selected to fit the SROC curve when a different truncation method applied. Therefore, the estimates of the SROC parameter log-odds ratio (a) and heterogeneity (b) could be different for each truncation method and each replication. a is set to be ln2 and b is set to be 0 in the simulation step. The sample mean, the bias and the standard deviation of estimated a and b in 1000 replication are listed below. Also, boxplots are used for visualization.

Table 3.4: Properties of $\hat{a}$ (log-odds ratio) under method B.

	a1	a2	a3	a4	a5	a6	a	a7	a8
Mean	0.883	0.837	0.805	0.754	0.713	0.700	0.683	0.705	0.713
Bias	0.190	0.144	0.112	0.061	0.020	0.007	-0.010	0.011	0.020
SD	1.382	0.672	0.449	0.296	0.176	0.149	0.146	0.147	0.190

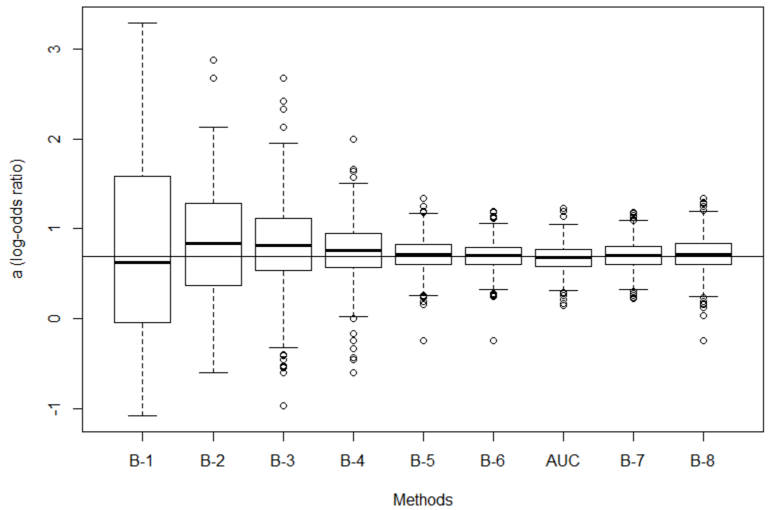

Figure 3.7: Boxplot showing the distribution of $\hat{a}$ (log-odds ratio) under method B.

Table 3.5: Properties of $\hat{b}$ (heterogeneity) under method B.

	b1	b2	b3	b4	b5	b6	b	b7	b8
Mean	0.019	0.034	0.034	0.022	0.012	0.006	-0.001	0.007	0.012
Bias	0.019	0.034	0.034	0.022	0.012	0.006	-0.001	0.007	0.012
SD	0.303	0.194	0.169	0.138	0.101	0.086	0.073	0.086	0.105

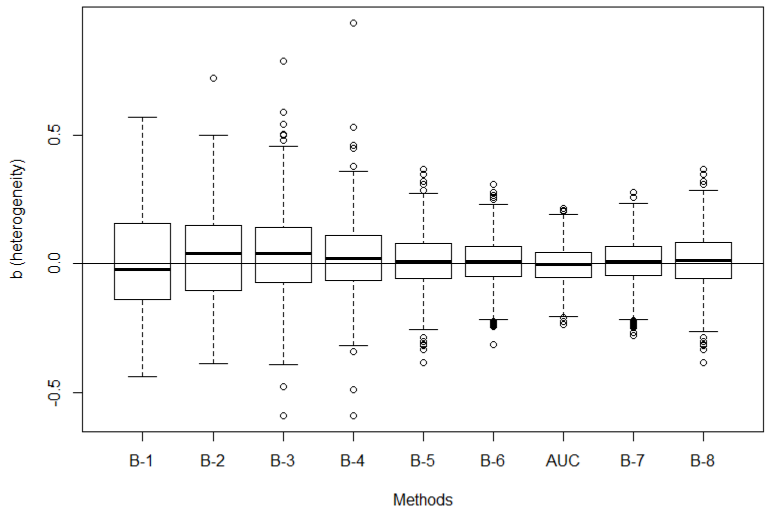

Figure 3.8: Boxplot showing the distribution of $\hat{b}$ (heterogeneity) under method B.

The estimates of the SROC curve parameters a (log-odds ratio) and b (heterogeneity) based on part of the data are cloSD to the true value shown as the solid line in Figure 3.7 and Figure 3.8. From table 3.4 and table 3.5, the estimating of a and b becomes more accurate and precise as s increases. More data is included in fitting SROC curve; the fitted curve is closer to the true curve. For B-5 to B-8, the bias and standard deviation for both a and b are similar to what in column AUC. Therefore, estimating the values of a and b with part of the data is as good as using the full data when s is large enough.

3.1.4 $\widehat{AUC}^*_{0s}$ Under Method B

The partial AUC obtained in method B is also scaled. Scaling is applied after the SRCO curve being fitted, so it does not affect the estimates of the SROC parameters a and b.

Table 3.6: Properties of $\widehat{AUC}^*_{0s}$ under method B.

row	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.464	0.351
Mean	0.107	0.186	0.260	0.324	0.436	0.532	0.612	0.465	0.352
Bias ($\times 10^{-3}$)	13.293	9.329	8.613	6.159	2.821	1.143	-2.009	1.459	0.819
SD ($\times 10^{-3}$)	1.101	3.163	3.610	3.194	2.718	2.543	2.316	6.432	4.140

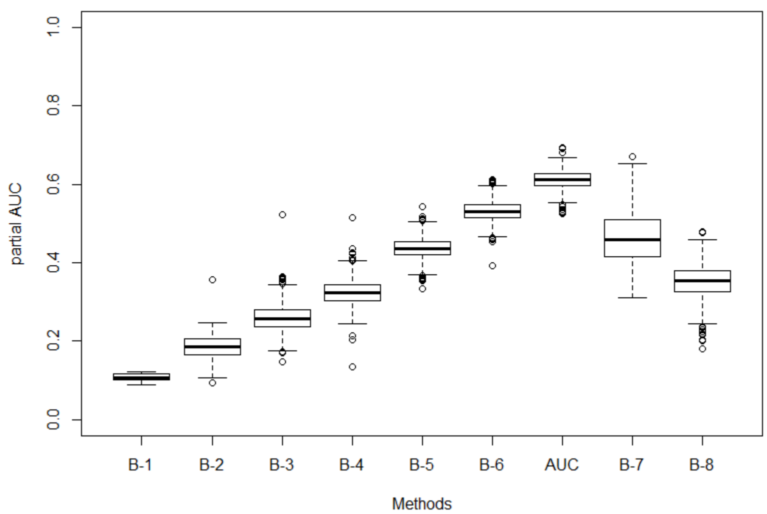

Figure 3.9: Boxplot showing the distribution of $\widehat{AUC}^*_{0s}$ under method B.

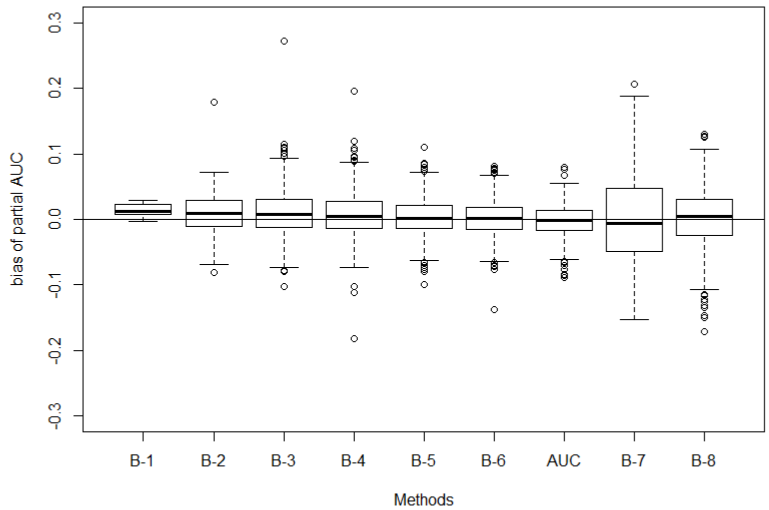

Figure 3.10: Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0s}$) under method B.

Similar to what has been observed in method A, the scaling make the standard deviation more constant between methods. The estimated partial AUC is more biased with scaling because of dividing by a number less than one, especially for methods with small s.

Similar to the base setting, all the experiments listed in Table 2.1 requires 6 tables and 10 figures to illustrate. To reduce the length of contents, most of the tables and figures will be attached in the appendix.

3.2 Effect of Truncation Methods as n_1 and n_2 Change

3.2.1 The Ratio of n_1 and n_2 Changes

Experiments 1, 2 and 3 listed in Chapter 2 have the same settings except for the ratio of numbers of cases and non-cases. Experiment 1 - the base setting has 50 cases and 200 cases for each study, which is a commonly seen ratio in empirical data from the literature. However, in practice, the ratio could be very different. If the effect of truncation method is similar for all these three experiment, then the choice of truncation method would be easy for there is no need to consider the ratio of numbers of cases and non-cases within studies. Experiment 2 has 125 cases and 125 non-cases for each study (more balanced than experiment 1) and experiment 3 has 25 cases and 225 non-cases for each study (more unbalanced than experiment 1). All three experiments have the same total number of cases and non-cases.

Starting from the estimates of the partial AUC, the changes of results with different ratio of numbers of cases and non-cases are similar under method A and method B. In the view of the bias, there is no significant difference among three settings, all the biases are small relative to the sample means of the partial AUC. As for the standard deviation, the ratio of numbers of cases and non-cases does not affect the standard deviation for the fixed percentage truncation method (A-7 and A-8). However, for the fixed point truncation methods, when the numbers of cases and non-cases are more balanced (Experiment 2), the standard deviation decreases, and the standard deviation increases when the numbers of cases and non-cases are more unbalanced (Experiment 3). This implies that the variation inside a study does affect the variation to the curve.

The figures of boxplots of the partial AUC and the bias for these three experiments show almost the same pattern (See Appendix, Figure A.1, A.2, A.11, A.12, A.21 and A.22). Then the performance of the truncation methods is not affected by the ratio of the numbers of cases and non-cases.

The observations for the partial AUC are also true for the scaled partial AUC. Since the ratio of numbers of cases and non-cases does not affect the values of the estimated partial AUC and the truncating range, the scaling changes the results with the same scale, and the increasing and decreasing relationships remain the same.

As for the estimated SROC parameter a and b in method B, the standard deviation in Experiment 3 is greater than that in experiment 1, and Experiment 2 has the lowest standard deviation, which is the same as the estimated partial AUC. Same as in the base setting, when the valid number is large enough, equivalently, when the truncation point s is large enough, fitting the SROC curve with only part of the data gives sufficiently accurate and precise estimates of both a and b.

In summary, the ratio of number of cases and non-cases does effect some properties of the partial AUC and the SROC parameters a and b, but it does not change the effect of truncation methods on them. Therefore, the ratio of number of cases and non-cases can be ignored when choosing a truncation method to an empirical clinical dataset.

3.2.2 n_1 and n_2 are not Fixed

Experiment 9 in the appendix tested the situation that the numbers of cases and non-cases are from a uniform distribution, and the numbers are kept to be the same between studies. In the other word, the variation of the numbers of cases and non-cases between replications is added to experiment 9. The results of experiment 9 are almost the same as Experiment 1. This implies that when keeping n_1 and n_2 to be the same between studies, the variation of n_1 and n_2 between replications does not affect the performance of the truncation methods.

3.3 Effect of Truncation Methods as a_0 and b_0 Change

3.3.1 a_0 Increases

When increasing the log-odds ratio, the SROC curve is moving towards the upper left corner shown in the following figure. Therefore, the region under the SROC curve gains more area, causing the partial AUC to also increase for some fixed s.

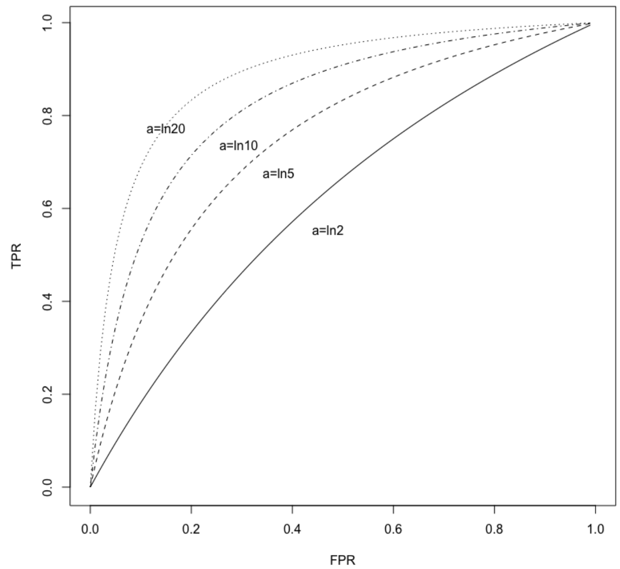

Figure 3.11: SROC curves with different values of log-odds ratio (a).

Experiments 1, 4, 5 and 6 apply increasing values of log-odd ratio (a), and all other parameters and distribution are kept to be the same.

First, comparing the results under a certain method of type A, the sample mean and the absolute value of bias of the partial AUC increase as the log-odds ratio increases. As for the standard deviation, it is increasing as the log-odds ratio when truncating at low values of s, and it is decreased for high values of s. The standard deviations are always increasing as a increases for A-7 and A-8. However, the pattern of how the bias and standard deviation are effected by the truncation methods are the same for any value of a. The sample mean, the absolute value of bias and the standard deviation of partial AUC are increasing as s increases for the fixed point truncation methods, and A-7 always give the largest bias and standard deviation.

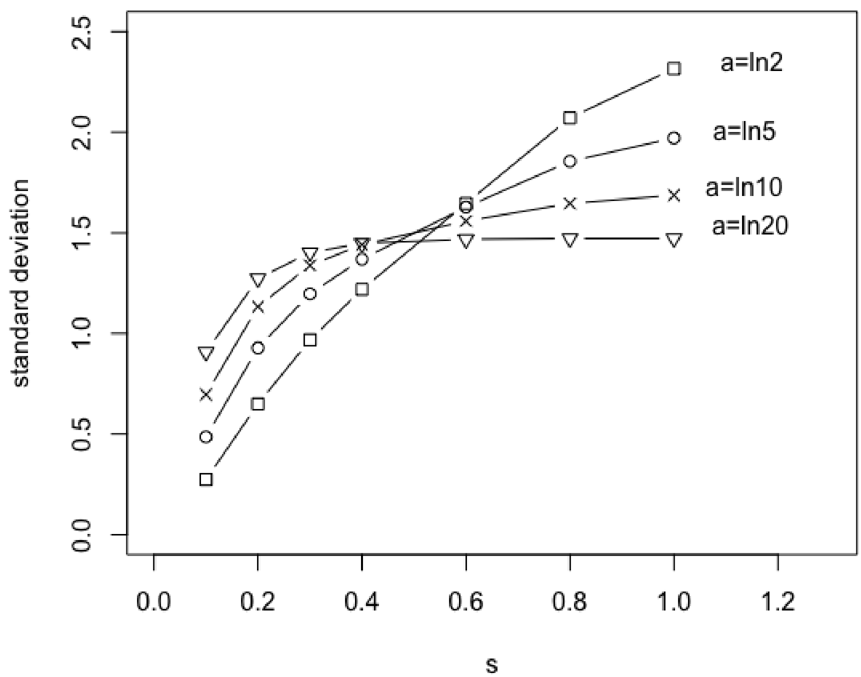

Figure 3.12: Plot of standard deviations of AUC_{0s} with different values of log-odds ratio (a).

After scaling, the decreasing relationship between the standard deviation and s becomes more dramatic as the log-odds ratio increases. Visually, from Figure 3.14, Figure 3.15 and Figure 3.16, the range of both estimated AUC_{0s} and the bias shrinks along the x-axis from “p1” to “AUC” more obviously as a increases. This suggests, in practice, truncating at a high value of the FPR is suggested for a large log-odds ratio study if the scaled partial AUC is applied.

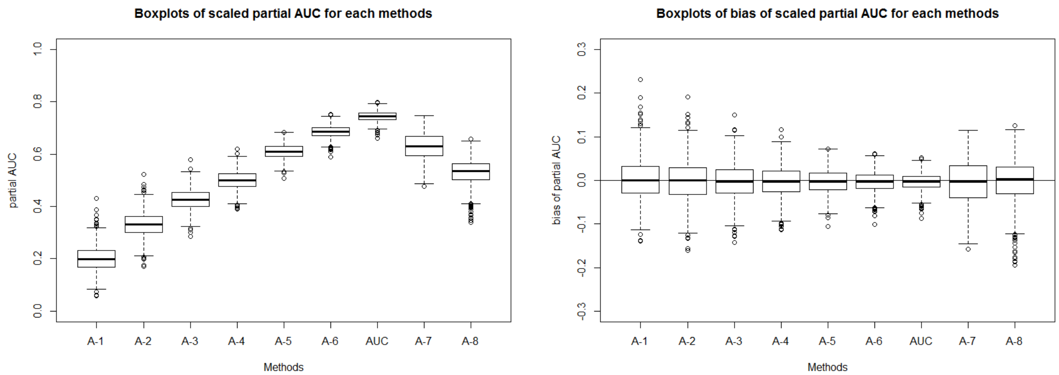

Figure 3.13: Boxplots showing the distribution of $\widehat{AUC}_{0s}$ and bias $(\widehat{AUC}_{0s})$ under method A with $a = \ln 5$.

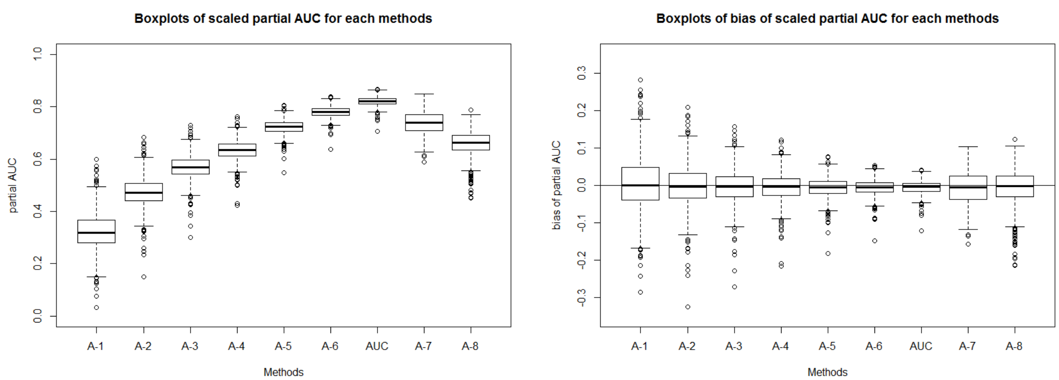

Figure 3.14: Boxplots showing the distribution of $\widehat{AUC}_{0s}$ and bias $(\widehat{AUC}_{0s})$ under method A with $a = \ln 10$.

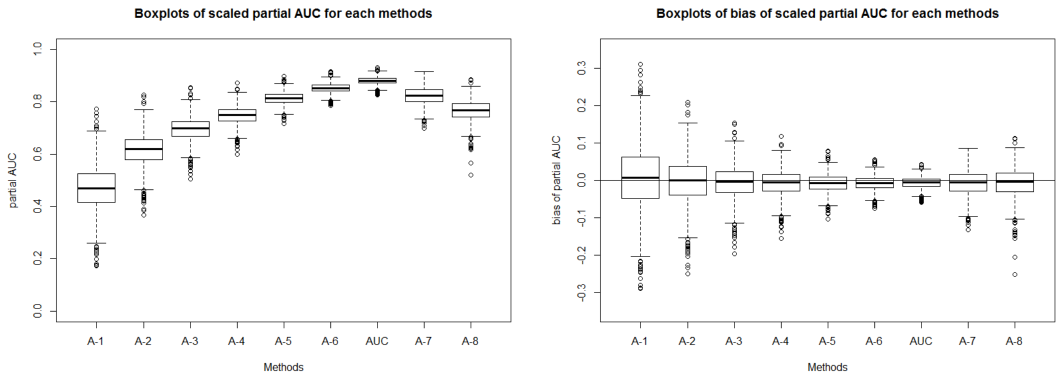

Figure 3.15: Boxplot showing the distribution of $\widehat{AUC}_{0s}$ and bias $(\widehat{AUC}_{0s})$ under method A with $a = \ln 20$.

Secondly, for all type B methods, the sample mean of the partial AUC increases as the log-odds ratio increase, but there is no clear trend of how the bias and standard deviation changes with the log-odds ratio. Additionally, increasing log-odds ratio does not make a great difference on how the truncation methods affect the estimates of a and b with respect to the bias and the standard deviation. The scaling has the same effect as for method A.

3.3.2 $b_0 \neq 0$

When the parameter b is not equal to 0, the study is called heterogeneous. The same process of simulation and truncation is also applied to heterogeneous cases, and the only difference is when generating the data, the values of sensitivity are calculated by equation 2.1 with non-zero value of b.

Not yet being truncated, there is a problem when generating the SROC curve by simulated data. The log-odds ratio is under-estimated when b_0 is positive and over-estimated when b_0 is negative. Also, b is biased toward 0. Here is a table of one estimation at $a_0 = \ln 2 = 0.693$.

Table 3.7: Estimations of a, b and AUC in heterogeneous case.

b_0	-0.8	-0.6	-0.3	-0.2	-0.1	0.1	0.2	0.3	0.6	0.8
$\hat{a}$	0.717	0.711	0.697	0.701	0.695	0.676	0.656	0.619	0.509	0.424
$\hat{b}$	-0.751	-0.564	-0.282	-0.184	-0.093	0.084	0.159	0.224	0.374	0.448
SD(b)	0.117	0.096	0.077	0.080	0.078	0.075	0.078	0.089	0.106	0.112
$\widehat{AUC}$	0.599	0.606	0.611	0.613	0.613	0.610	0.606	0.599	0.579	0.564
bias (AUC)	0.005	0.004	0.001	0.001	0.000	-0.003	-0.006	-0.011	-0.023	-0.030

When b_0 is negative, the estimated b is larger than the theoretical value, which is less heterogeneous. The estimated b gets more biased as the absolute value of b increases. When b_0 is positive, estimated b is smaller than the theoretical value, which is also less heterogeneous. The estimated b is getting more biased as the absolute value of b increases. The estimated b is more biased when b_0 is positive than when b_0 is negative. The parameter a is over-estimated when b_0 is negative, and it is under-estimated when b_0 is positive. Balanced by the estimation of a and b, estimated AUC is not that biased.

A lot work has been done to find out the reason behind the biasness of the estimates $\hat{a}$ and $\hat{b}$ in heterogeneous case. The first guess is that the 0.5 correction causes the bias, then the numbers of cases and non-cases are increased to reduce the effect of 0.5 correction. Then the estimates become less biased; however, even when the number of cases is increased to 50,000 and the number of non-cases is increased to 200,000, which are incredibly large for a clinical data, the bias is still present. This means that the effect of 0.5 correction cannot be eliminated by increasing the sample size. Then the linear regression in Moses's model is replaced with the weighted linear regression (Walter, 2002). The bias is less but still not satisfactory.

To find out at which step in the simulation the problem happens, some lines of the R code are skipped. The finding is that when the step of generating binomial random variables is skipped, the parameters a and b are estimated perfectly for both homogeneous cases an heterogeneous cases. Binomial sampling is adding noise to the true function, and the noise is strong when one of the TPR and FPR is close to 0 or 1. Figure 3.16 shows the SROC curve with a being 2 and b equals to 0, 0.5 and -0.5 respectively. It can be seen that on the SROC curves for heterogeneous case, there are more parts of the curve that corresponding to a TPR or FPR close to 0 or 1. However, this guess fails to explain the asymmetry of the bias and the bias occur so quickly as b slightly approaches away from 0.

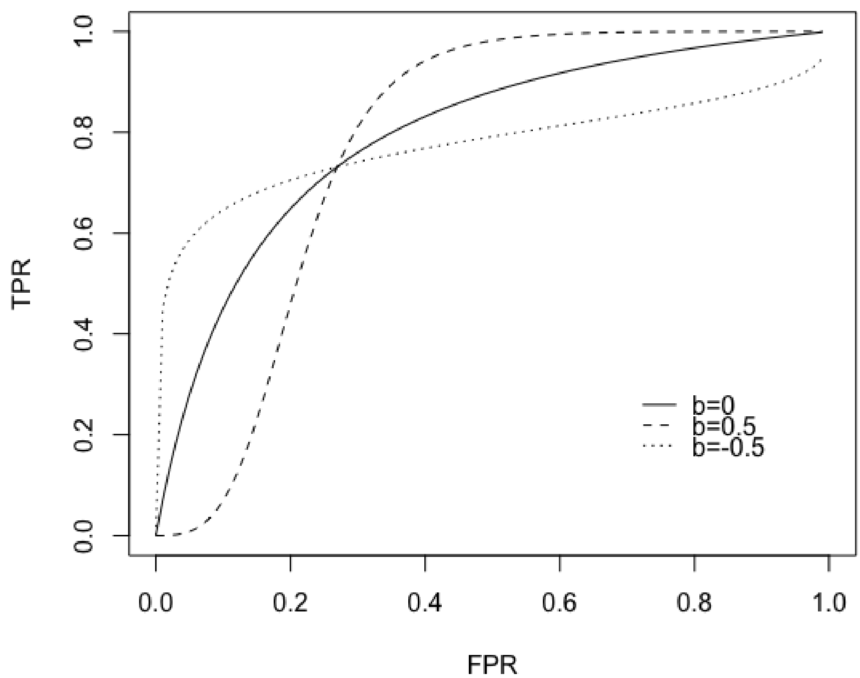

Figure 3.16: SROC curve with different values of heterogeneity (b).

3.4 Effect of Truncation Methods as the Distribution of Specificity Changes

Experiment 1, 7 and 8 have the same settings except for the distribution of specificity.

One major difference when comparing these three experiments is the performance of the fixed percentage truncation method. Applying beta (5,2) or uniform (0,1) distributions significantly decrease the standard deviation of the partial AUC under method A-7, A-8, B-7 and B-8. The other difference is that for method B in Experiment 7 and 8, the valid numbers for small s are much less than the valid number in Experiment 1.

Chapter 4

Summary and Discussion

4.1 Summary

This thesis examined the effect of truncation methods on the properties of partial AUC, such as the bias, standard deviation, accuracy and precision for estimating SROC parameters a and b.

Applying different truncation methods does change the value of partial AUC. Specifically, $\widehat{AUC}_{0,s}$ increases as s increases, whether the value of s is fixed or determined by other methods. Either the SROC curve is fitted by the full data then truncated (method A) or the SROC curve is fitted by the truncated data (method B). Under the fixed point truncation method, which is truncating the SROC curve at a fixed point s, the standard deviation of the partial AUC increases as s increases and the standard deviation of the scaled partial AUC decreases as s increases. This is corresponding to the analytical results from (Walter, 2005). This suggests that if the value of the partial AUC is used, then it is possible to truncate at a low value of the FPR, and if the value of the scaled partial AUC is used, then it is recommended to truncate at a high value of the FPR.

The fixed percentage truncation method is less recommended than the fixed point truncation method. In the fixed percentage truncation method, the variation in truncation point increases the variation in partial AUC. Even with scaling, the disadvantage of the fixed percentage truncation method being less precise than the fixed point truncation method cannot be eliminated. Comparing two fixed percentage truncation methods, excluding 20% of the data (A-8 and B-8) gives less bias and lower standard deviation than excluding 10% of the data (A-7 and B-7). Truncating the SROC which is fitted by either the full data or with part of the data, excluding 20% of the data gives less biased estimated partial AUC and smaller standard deviation.

Comparing methods A and B, method A is recommended. When applying method B, part of the data is excluded before fitting the SROC curve, which leads to two problems. First, there may not be enough data to fit the curve, especially when the truncating range is small. Second, fitting the SROC curve with only part of the data could be more variable due to the loss of information. What the simulation and truncation under method B provide is that when the truncation point s is sufficiently large or a large enough proportion of the data is used to fit the SROC, estimating the partial AUC and the SROC parameters a and b using part of the data is as good as using the full data.

Changing the simulation settings such as the ratio of the numbers of cases and non-cases, the log-odds ratio and the distribution of specificities does change the values of the bias and the standard deviation for partial AUC. However, how these properties are affected by the truncation methods is similar in all settings.

Combining all the results, truncating at FPR = 0.6 or 0.8 performs well no matter which method is used, with or without scaling. It gives low bias and a relatively low standard deviation compared with the estimated AUC. Truncating at a higher FPR would include more clinically irrelevant region, and truncating at a lower FPR would lose more information.

In practice, the choice of the truncating point is also determined by what region is the most interesting. When using the AUC or the partial AUC as a measure to compare diagnostic tests, multiple comparisons are applied. One can compare the full AUC, the partial AUC on a certain range and the partial AUC under the SROC curve fitted by a part of the data for the tests. If a diagnostic test has the highest value of the full AUC or the partial AUC, then it is convincing that this test performs the best in classifying cases and non-cases.

4.2 Future Work

For the homogeneous case, there are more choices of the settings when simulating the data. For example, the number of studies is not necessary to be 10. When there are more or less studies to fit an SROC curve, then the effect of truncation methods could be different from what performed as in this thesis. Also, the variation of the numbers of cases and non-cases between studies could be introduced. This could be achieved by simulating a pair of the numbers of cases and non-cases from a distribution for each study.

In this thesis, only one variable is changed from experiment 1, while in the clinical data, the situation is more complicated. For example, there might be interaction when changing the number of studies and the distribution of specificities at the same time. Also the distribution of specificities is not always able to be approximated by a named distribution. Other than the simulation method proposed in this thesis, an alternative way is to approximate the distributions of cases and non-cases. The distribution could be normal, logistic, or others. Then by moving the threshold, the empirical TPR and FPR are obtained.

Moreover, the reason for the unsatisfactory estimation in heterogeneous case needs more work to find out.

Appendix A

Results for Experiment 2 to Experiment 9

A.1 Experiment 2 - 125 cases and 125 non-cases per study, mixed uniform distribution of specificities, a=ln2, b=0

A.1.1 $\widehat{AUC}_{0,s}$ under method A

Table A.1: Experiment 2 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.315	0.164
Mean	0.009	0.035	0.075	0.127	0.259	0.423	0.612	0.313	0.163
Bias ($\times 10^{-3}$)	0.102	-0.055	-0.304	-0.535	-0.912	-1.356	-1.826	-1.496	-0.869
SD($\times 10^{-2}$)	0.240	0.569	0.843	1.050	1.379	1.702	1.891	9.796	3.863

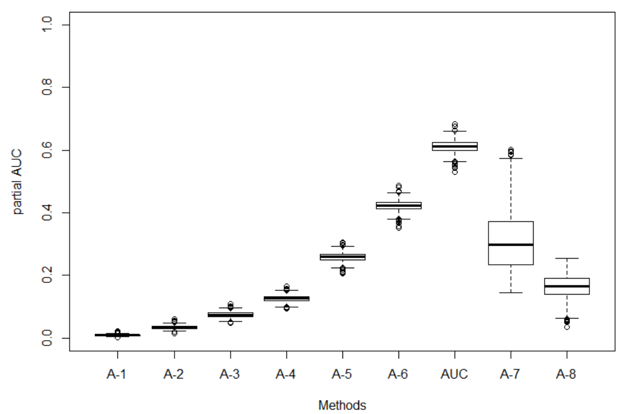

Figure A.1: Experiment 2 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method A

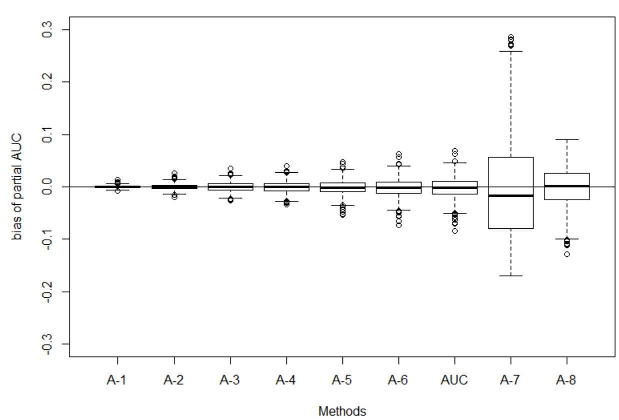

Figure A.2: Experiment 2 - Boxplot showing the distribution of bias $(\widehat{AUC}_{0,s})$ under method A

A.1.2 $\widehat{AUC}^*_{0,s}$ under method A

Table A.2: Experiment 2 - Properties of $\widehat{AUC}^*_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.463	0.353
Mean	0.095	0.177	0.250	0.316	0.432	0.529	0.612	0.462	0.351
bias ($\times 10^{-3}$)	1.021	-0.277	-1.015	-1.338	-1.521	-1.695	-1.826	-1.805	-1.707
SD($\times 10^{-2}$)	2.403	2.844	2.809	2.626	2.298	2.128	1.891	6.085	4.460

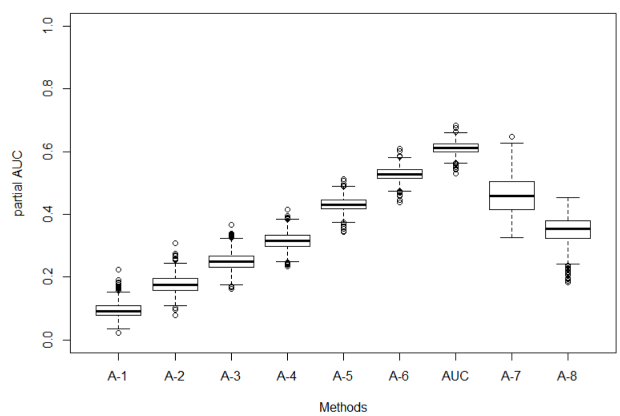

Figure A.3: Experiment 2 - Boxplot showing the distribution of ($\widehat{AUC}^*_{0,s}$) under method A

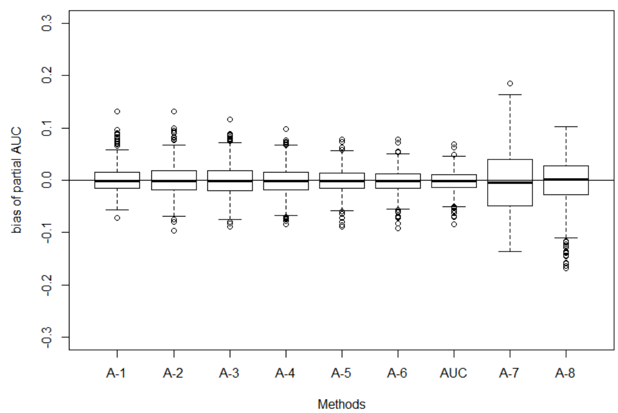

Figure A.4: Experiment 2 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method A

A.1.3 $\widehat{AUC}_{0,s}$ under method B

Table A.3: Experiment 2 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
Valid	13	171	579	938	1000	1000	1000	1000	1000
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.316	0.162
Mean	0.010	0.038	0.078	0.129	0.260	0.425	0.612	0.315	0.163
Bias ($\times 10^{-3}$)	1.038	2.235	2.287	1.915	0.366	0.227	-1.826	-1.368	1.534
SD($\times 10^{-2}$)	0.162	0.606	0.903	1.167	1.412	1.708	1.891	9.850	3.767

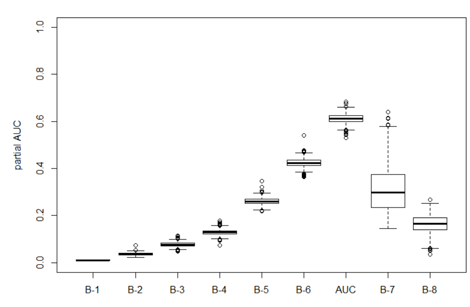

Figure A.5: Experiment 2 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method B

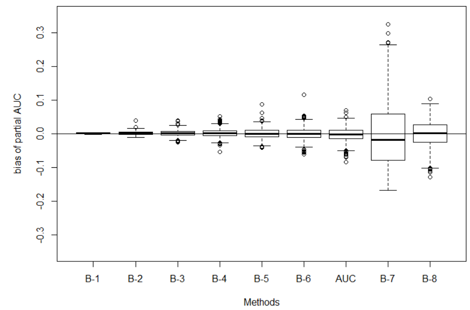

Figure A.6: Experiment 2 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method B

Table A.4: Experiment 2 - Properties of $\hat{a}$ (log-odds ratio) under method B

	a1	a2	a3	a4	a5	a6	a	a7	a8
Mean	1.692	0.910	0.772	0.742	0.697	0.695	0.683	0.696	0.696
Bias	0.567	0.369	0.165	0.073	0.023	0.015	0.014	0.016	0.028
SD	0.753	0.608	0.406	0.270	0.153	0.124	0.119	0.126	0.168

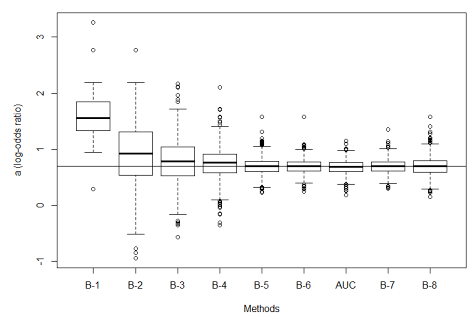

Figure A.7: Experiment 2 - Boxplot showing the distribution of $\hat{a}$ (log-odds ratio) under method B

Table A.5: Experiment 2 - Properties of $\hat{b}$ (heterogeneity) under method B

	b1	b2	b3	b4	b5	b6	b	b7	b8
Mean	0.196	0.050	0.021	0.018	0.005	0.004	-0.001	0.004	0.004
Bias	0.026	0.029	0.023	0.015	0.008	0.005	0.004	0.005	0.009
SD	0.161	0.169	0.152	0.123	0.089	0.074	0.063	0.074	0.093

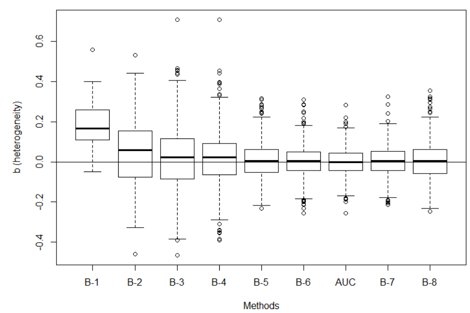

Figure A.8: Experiment 2 - Boxplot showing the distribution of $\hat{b}$ (heterogeneity) under method B

A.1.4 $\widehat{AUC}^*_{0,s}$ under method B

Table A.6: Experiment 2 - Properties of $\widehat{AUC}^*_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.463	0.353
Mean	0.104	0.188	0.259	0.322	0.434	0.531	0.612	0.463	0.351
Bias ($\times 10^{-3}$)	10.380	11.177	7.624	4.787	0.609	0.284	-1.826	-0.237	-1.106
SD ($\times 10^{-2}$)	1.623	3.028	3.011	2.919	2.353	2.135	1.891	6.105	4.205

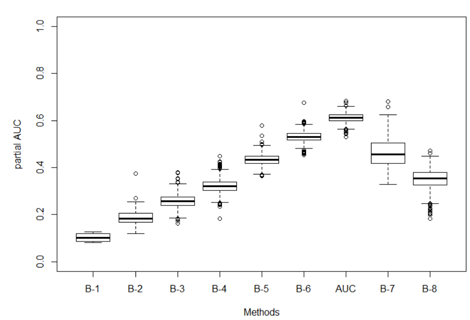

Figure A.9: Experiment 2 - Boxplot showing the distribution of $\widehat{AUC}^*_{0,s}$ under method B

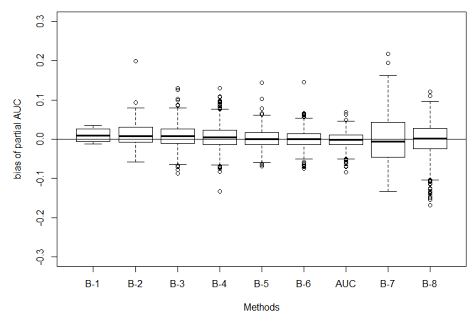

Figure A.10: Experiment 2 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method B

A.2 Experiment 3 - 25 cases and 225 non-cases per study, mixed uniform distribution of specificities, a=ln2, b=0

A.2.1 $\widehat{AUC}_{0,s}$ under method A

Table A.7: Experiment 3 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.314	0.159
Mean	0.010	0.036	0.076	0.128	0.260	0.424	0.612	0.313	0.159
Bias ($\times 10^{-3}$)	0.503	0.734	0.773	0.726	0.322	-0.707	-1.879	-0.737	0.552
SD ($\times 10^{-2}$)	0.330	0.777	1.168	1.495	2.100	2.695	3.031	9.740	3.707

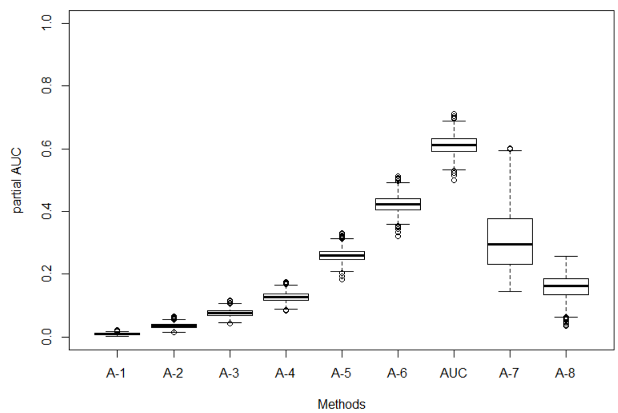

Figure A.11: Experiment 3 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method A

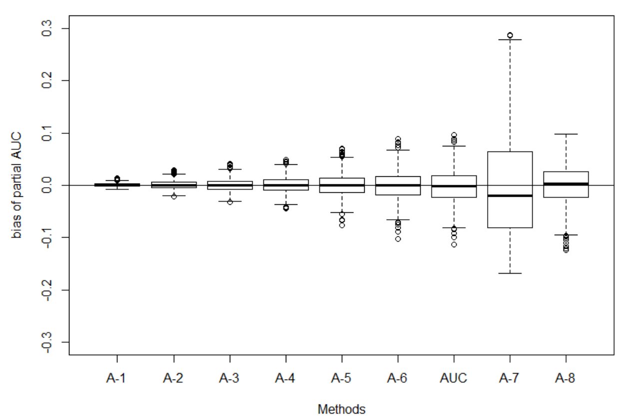

Figure A.12: Experiment 3 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method A

A.2.2 $\widehat{AUC}^*_{0,s}$ under method A

Table A.8: Experiment 3 - Properties of $\widehat{AUC}^*_{0,s}$ under method A

	p1	p2	p3	p4	p5	s	0.1	0.2	0.3
0.4	0.6	0.8	1						
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.463	0.348
Mean	0.099	0.180	0.253	0.319	0.434	0.530	0.612	0.462	0.349
Bias ($\times 10^{-3}$)	5.028	3.670	2.577	1.816	0.536	-0.884	-1.879	-0.691	1.502
SD ($\times 10^{-2}$)	3.299	3.887	3.892	3.737	3.500	3.369	3.031	6.557	4.948

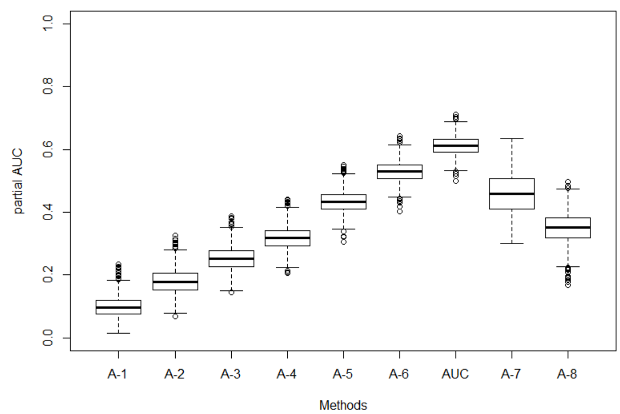

Figure A.13: Experiment 3 - Boxplot showing the distribution of ($\widehat{AUC}^*_{0,s}$) under method A

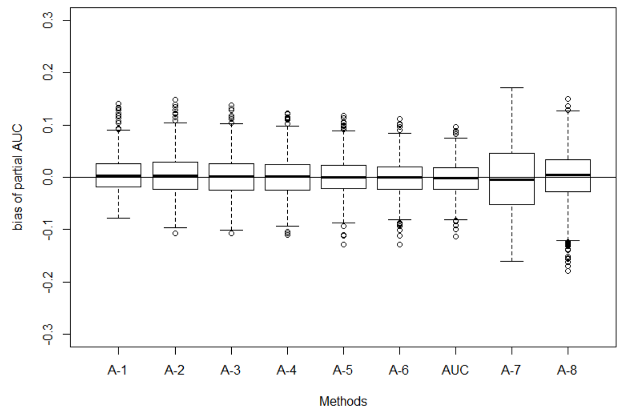

Figure A.14: Experiment 3 - Boxplot showing the distribution of bias ($\widehat{AUC}^{*}_{0,s}$) under method A

A.2.3 $\widehat{AUC}_{0,s}$ under method B

Table A.9: Experiment 3 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
Valid	12	161	611	947	1000	1000	1000	1000	1000
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.314	0.159
Mean	0.012	0.040	0.080	0.131	0.264	0.428	0.612	0.316	0.161
Bias ($\times 10^{-3}$)	2.292	4.186	4.579	4.231	4.155	3.778	-1.879	2.211	1.787
SD ($\times 10^{-2}$)	0.457	0.894	1.511	1.736	2.150	2.687	3.031	9.864	3.673

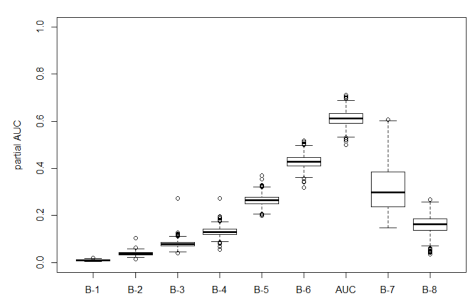

Figure A.15: Experiment 3 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method B

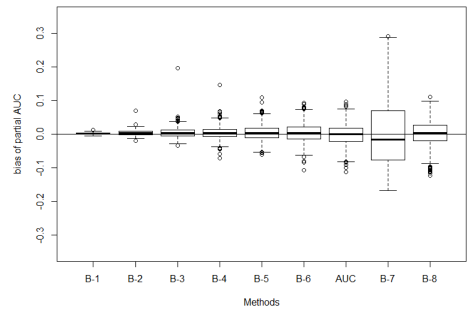

Figure A.16: Experiment 3 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method B

Table A.10: Experiment 3 - Properties of $\hat{a}$ (log-odds ratio) under method B

	a1	a2	a3	a4	a5	a6	a	a7	a8
Mean	0.892	0.879	0.818	0.774	0.733	0.720	0.686	0.718	0.737
Bias	2.463	0.973	0.337	0.162	0.056	0.040	0.037	0.039	0.065
SD	1.569	0.986	0.581	0.403	0.236	0.199	0.192	0.197	0.256

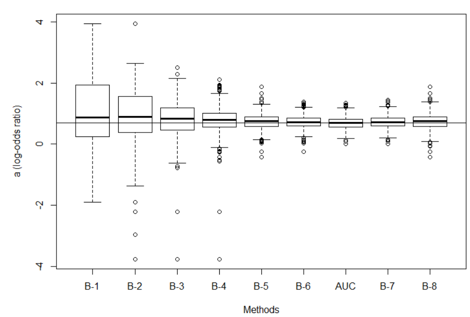

Figure A.17: Experiment 3 - Boxplot showing the distribution of $\hat{a}$ (log-odds ratio) under method B

Table A.11: Experiment 3 - Properties of $\hat{b}$ (heterogeneity) under method B

	b1	b2	b3	b4	b5	b6	b	b7	b8
Mean	0.013	0.033	0.033	0.027	0.016	0.011	-0.007	0.009	0.017
Bias	0.123	0.069	0.044	0.030	0.016	0.011	0.008	0.011	0.017
SD	0.351	0.262	0.209	0.172	0.125	0.107	0.089	0.105	0.132

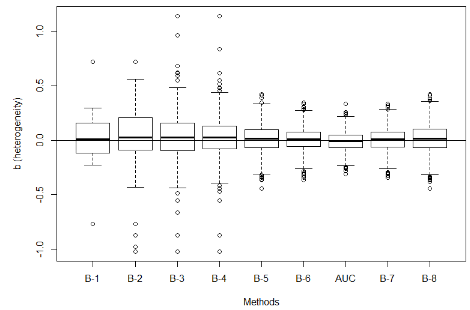

Figure A.18: Experiment 3 - Boxplot showing the distribution of $\hat{b}$ (heterogeneity) under method B

A.2.4 $\widehat{AUC}^*_{0,s}$ under method B

Table A.12: Experiment 3 - Properties of $\widehat{AUC}^*_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.463	0.348
Mean	0.117	0.198	0.266	0.328	0.440	0.535	0.612	0.467	0.352
Bias ($\times 10^{-3}$)	22.920	20.929	15.262	10.579	6.924	4.723	-1.879	3.536	4.350
SD ($\times 10^{-2}$)	4.566	4.471	5.036	4.341	3.583	3.358	3.031	6.608	4.792

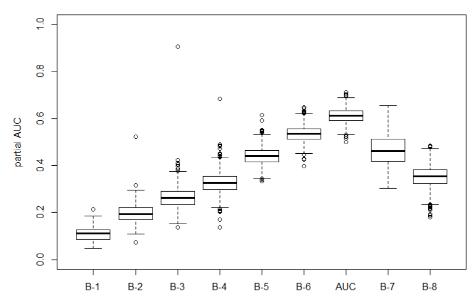

Figure A.19: Experiment 3 - Boxplot showing the distribution of $\widehat{AUC}^*_{0,s}$ under method B

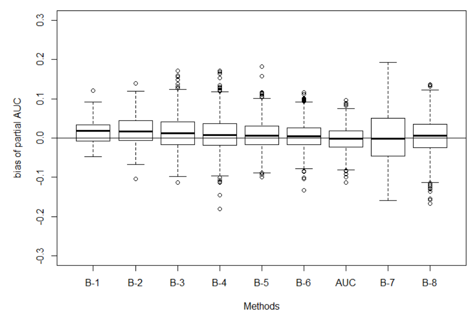

Figure A.20: Experiment 3 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method B

A.3 Experiment 4 - 50 cases and 200 non-cases per study, mixed uniform distribution of specificities, a=ln5, b=0

A.3.1 $\widehat{AUC}_{0,s}$ under method A

Table A.13: Experiment 4 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.020	0.066	0.129	0.201	0.368	0.552	0.747	0.424	0.243
Mean.	0.020	0.066	0.128	0.200	0.365	0.548	0.743	0.421	0.241
Bias ($\times 10^{-3}$)	0.288	-0.096	-0.592	-1.087	-2.088	-3.073	-3.725	-2.784	-1.446
SD($\times 10^{-2}$)	0.486	0.928	1.197	1.369	1.628	1.855	1.971	10.576	4.817

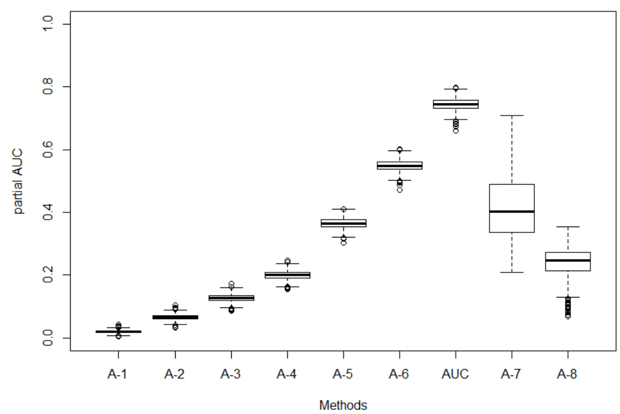

Figure A.21: Experiment 4 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method A

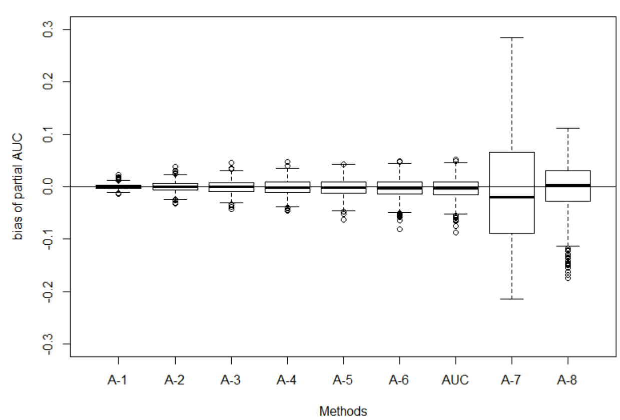

Figure A.22: Experiment 4 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method A

A.3.2 $\widehat{AUC}^*_{0,s}$ under method A

Table A.14: Experiment 4 - Properties of $\widehat{AUC}^*_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.1985243	0.332	0.429	0.504	0.613	0.689	0.747	0.634	0.533
Mean	0.201	0.331	0.427	0.501	0.609	0.686	0.743	0.630	0.530
Bias ($\times 10^{-3}$)	2.884	-0.479	-1.975	-2.718	-3.480	-3.842	-3.725	-3.653	-2.878
SD ($\times 10^{-2}$)	4.855	4.639	3.991	3.423	2.714	2.318	1.971	5.115	4.812

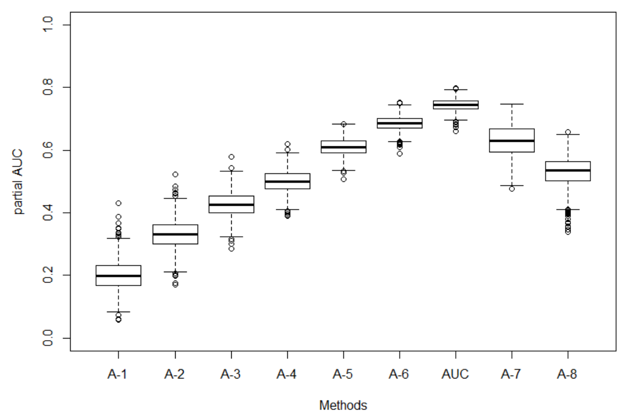

Figure A.23: Experiment 4 - Boxplot showing the distribution of ($\widehat{AUC}^*_{0,s}$) under method A

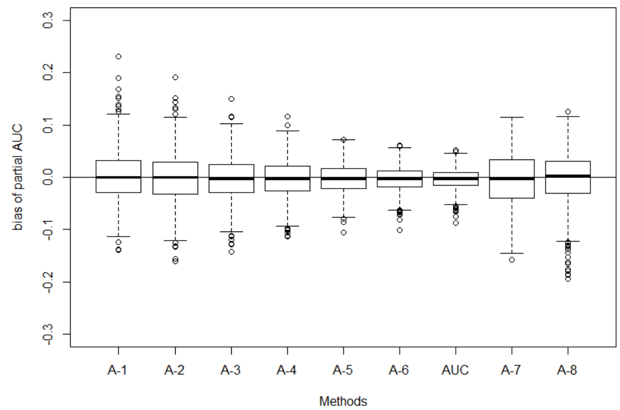

Figure A.24: Experiment 4 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method A

A.3.3 $\widehat{AUC}_{0,s}$ under method B

Table A.15: Experiment 4 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
Valid	10	172	571	943	1000	1000	1000	1000	1000
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.020	0.066	0.129	0.201	0.368	0.552	0.747	0.424	0.243
Mean	0.020	0.069	0.130	0.201	0.367	0.551	0.743	0.423	0.241
Bias ($\times 10^{-3}$)	0.647	2.533	1.019	-0.197	-0.897	-0.479	-3.725	-0.811	-1.514
SD ($\times 10^{-2}$)	0.457	0.894	1.511	1.736	2.150	2.687	3.031	9.864	3.673

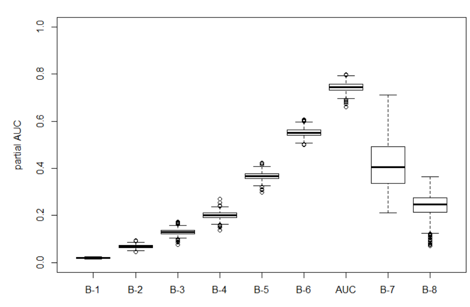

Figure A.25: Experiment 4 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method B

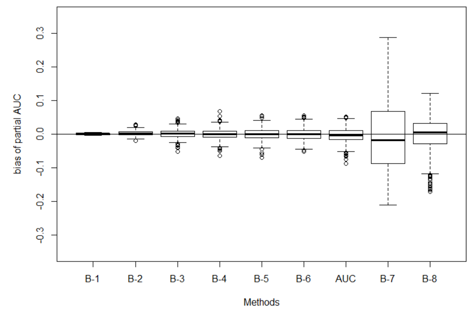

Figure A.26: Experiment 4 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method B

Table A.16: Experiment 4 - Properties of $\hat{a}$ (log-odds ratio) under method B

	a1	a2	a3	a4	a5	a6	a	a7	a8
Mean	1.288	1.690	1.641	1.631	1.613	1.613	1.588	1.616	1.610
Bias	0.761	0.281	0.097	0.046	0.023	0.020	0.023	0.021	0.026
SD	0.872	0.530	0.311	0.214	0.152	0.143	0.151	0.144	0.160

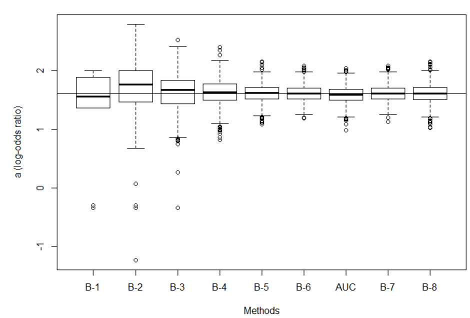

Figure A.27: Experiment 4 - Boxplot showing the distribution of $\hat{a}$ (log-odds ratio) under method B

Table A.17: Experiment 4 - Properties of $\hat{b}$ (heterogeneity) under method B

	b1	b2	b3	b4	b5	b6	b	b7	b8
Mean	-0.086	0.012	0.011	0.019	0.016	0.016	-0.007	0.016	0.013
Bias	0.046	0.036	0.022	0.016	0.009	0.007	0.007	0.007	0.010
SD	0.215	0.189	0.148	0.126	0.094	0.086	0.082	0.086	0.099

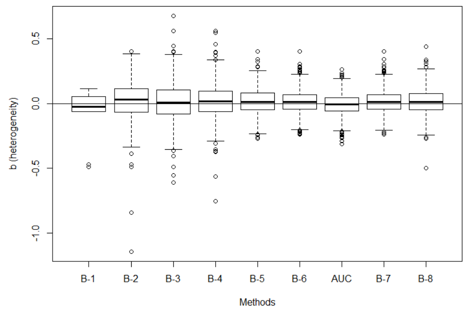

Figure A.28: Experiment 4 - Boxplot showing the distribution of $\hat{b}$ (heterogeneity) under method B

A.3.4 $\widehat{AUC}^*_{0,s}$ under method B

Table A.18: Experiment 4 - Properties of $\widehat{AUC}^*_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.199	0.332	0.429	0.504	0.613	0.689	0.747	0.634	0.533
Mean	0.205	0.344	0.432	0.503	0.611	0.689	0.743	0.633	0.530
bias ($\times 10^{-3}$)	6.472	12.664	3.398	-0.494	-1.495	-0.599	-3.725	-0.944	-3.130
SD ($\times 10^{-2}$)	2.631	4.019	3.935	3.490	2.641	2.173	1.971	5.197	4.857

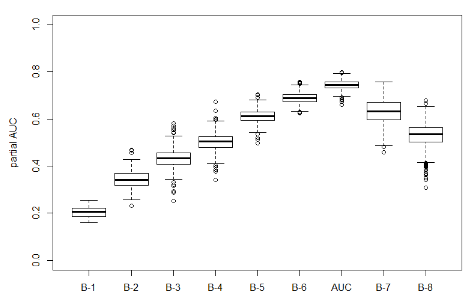

Figure A.29: Experiment 4 - Boxplot showing the distribution of $\widehat{AUC}^*_{0,s}$ under method B

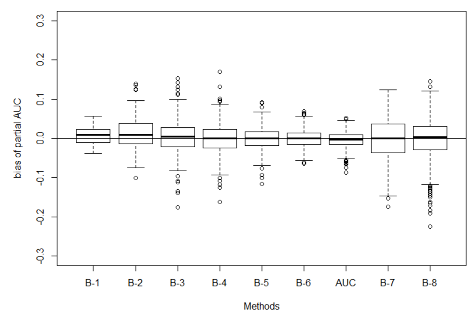

Figure A.30: Experiment 4 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method B

A.4 Experiment 5 - 50 cases and 200 non-cases per study, mixed uniform distribution of specificities, a=ln10, b=0

A.4.1 $\widehat{AUC}_{0,s}$ under method A

Table A.19: Experiment 5 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.032	0.095	0.172	0.256	0.437	0.629	0.827	0.500	0.303
Mean	0.032	0.095	0.170	0.254	0.434	0.624	0.821	0.495	0.301
Bias ($\times 10^{-3}$)	0.456	-0.371	-1.312	-2.215	-3.843	-5.129	-5.811	-4.602	-2.723
SD($\times 10^{-2}$)	0.696	1.133	1.338	1.441	1.559	1.646	1.687	11.478	5.126

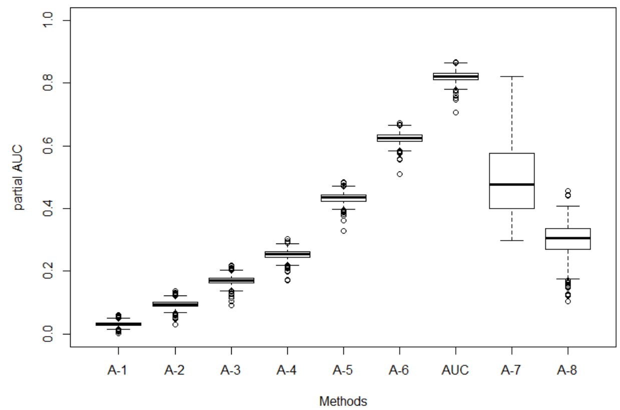

Figure A.31: Experiment 5 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method A

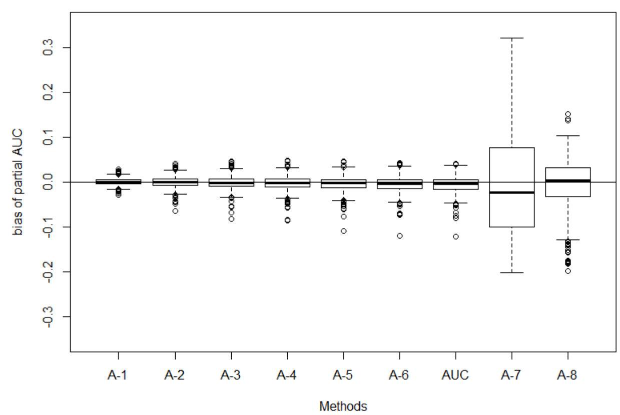

Figure A.32: Experiment 5 - Boxplot showing the distribution of bias $(\widehat{AUC}_{0,s})$ under method A

A.4.2 $\widehat{AUC}^*_{0,s}$ under method A

Table A.20: Experiment 5 - Properties of $\widehat{AUC}^*_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.319	0.476	0.573	0.640	0.729	0.786	0.827	0.746	0.666
Mean	0.323	0.474	0.568	0.635	0.723	0.780	0.821	0.739	0.660
bias ($\times 10^{-3}$)	4.558	-1.857	-4.372	-5.539	-6.405	-6.412	-5.811	-6.686	-6.013
SD($\times 10^{-2}$)	6.963	5.665	4.460	3.602	2.598	2.058	1.687	4.267	4.458

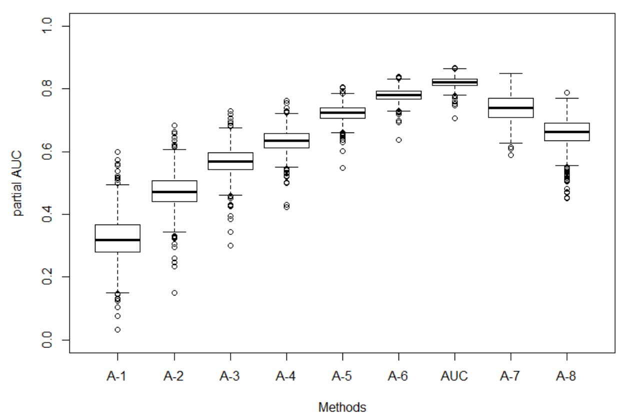

Figure A.33: Experiment 5 - Boxplot showing the distribution of ($\widehat{AUC}^*_{0,s}$) under method A

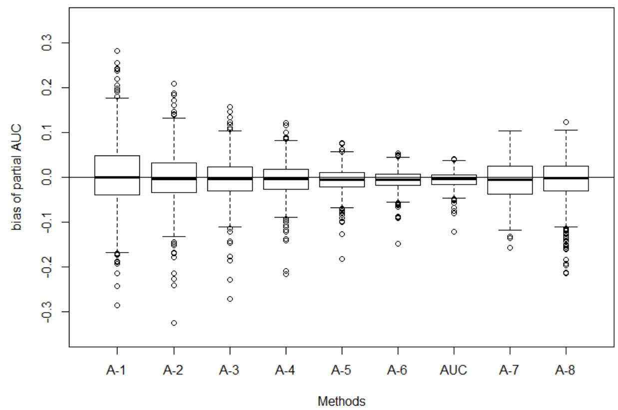

Figure A.34: Experiment 5 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method A

A.4.3 $\widehat{AUC}_{0,s}$ under method B

Table A.21: Experiment 5 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
Valid	9	162	603	946	1000	1000	1000	1000	1000
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.032	0.095	0.172	0.256	0.437	0.629	0.827	0.500	0.303
Mean	0.031	0.095	0.171	0.254	0.435	0.626	0.821	0.496	0.300
Bias ($\times 10^{-3}$)	-0.608	-0.587	-0.622	-1.709	-2.885	-3.383	-5.811	-3.713	-3.239
SD($\times 10^{-2}$)	0.689	1.082	1.225	1.454	1.537	1.605	1.687	11.590	5.081

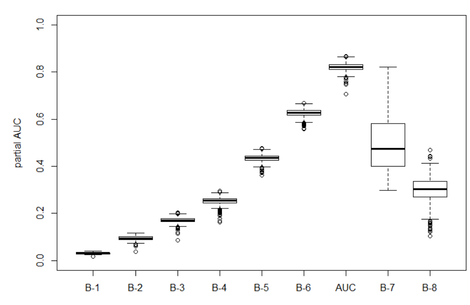

Figure A.35: Experiment 5 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method B

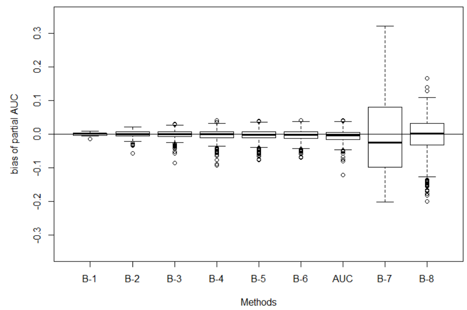

Figure A.36: Experiment 5 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method B

Table A.22: Experiment 5 - Properties of $\hat{a}$ (log-odds ratio) under method B

	a1	a2	a3	a4	a5	a6	a	a7	a8
Mean	2.687	2.460	2.348	2.322	2.293	2.282	2.258	2.281	2.296
Bias	0.398	0.144	0.058	0.034	0.025	0.025	0.026	0.025	0.025
SD	0.631	0.380	0.241	0.184	0.159	0.159	0.163	0.158	0.159

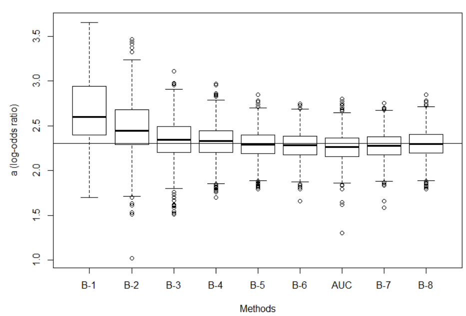

Figure A.37: Experiment 5 - Boxplot showing the distribution of $\hat{a}$ (log-odds ratio) under method B

Table A.23: Experiment 5 - Properties of $\hat{b}$ (heterogeneity) under method B

	b1	b2	b3	b4	b5	b6	b	b7	b8
Mean	0.119	0.077	0.034	0.033	0.028	0.022	-0.019	0.016	0.028
Bias	0.058	0.035	0.021	0.016	0.010	0.009	0.009	0.009	0.011
SD	0.242	0.187	0.146	0.127	0.101	0.094	0.093	0.093	0.103

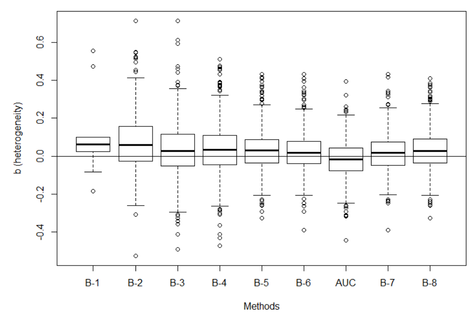

Figure A.38: Experiment 5 - Boxplot showing the distribution of $\hat{b}$ (heterogeneity) under method B

A.4.4 $\widehat{AUC}^*_{0,s}$ under method B

Table A.24: Experiment 5 - Properties of $\widehat{AUC}^*_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.319	0.476	0.573	0.640	0.729	0.786	0.827	0.746	0.666
Mean	0.313	0.473	0.571	0.636	0.724	0.782	0.821	0.740	0.659
bias ($\times 10^{-3}$)	-6.080	-2.933	-2.074	-4.271	-4.809	-4.229	-5.811	-5.616	-7.017
SD ($\times 10^{-2}$)	6.886	5.409	4.084	3.634	2.562	2.006	1.687	4.369	4.296

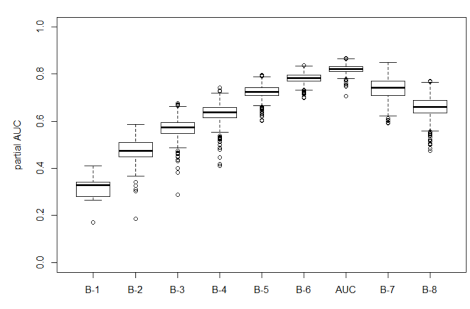

Figure A.39: Experiment 5 - Boxplot showing the distribution of $\widehat{AUC}^*_{0,s}$ under method B

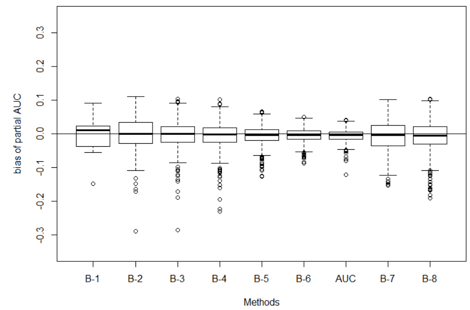

Figure A.40: Experiment 5 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method B

A.5 Experiment 6 - 50 cases and 200 non-cases per study, mixed uniform distribution of specificities, a=ln20, b=0

A.5.1 $\widehat{AUC}_{0,s}$ under method A

Table A.25: Experiment 6 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.046	0.124	0.210	0.302	0.492	0.688	0.887	0.551	0.348
Mean	0.047	0.123	0.209	0.299	0.488	0.682	0.880	0.546	0.345
bias ($\times 10^{-3}$)	0.722	-0.428	-1.591	-2.659	-4.466	-5.786	-6.445	-5.290	-3.108
SD($\times 10^{-2}$)	0.907	1.273	1.402	1.448	1.468	1.472	1.472	11.745	5.661

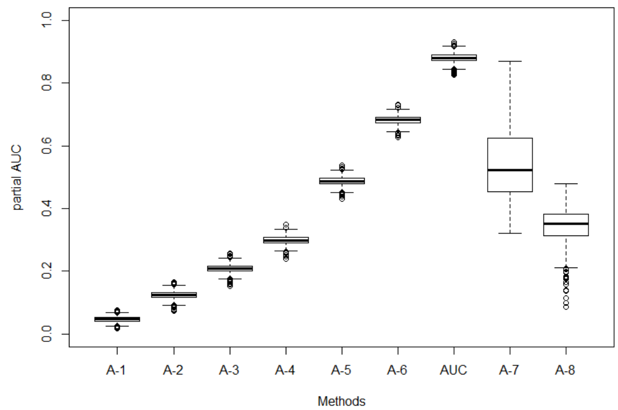

Figure A.41: Experiment 6 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method A

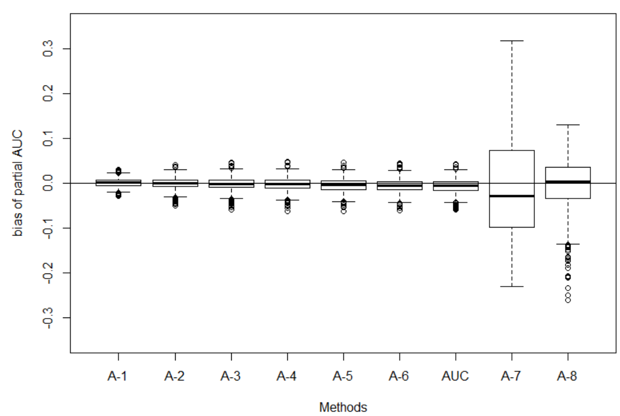

Figure A.42: Experiment 6 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method A

A.5.2 $\widehat{AUC}^*_{0,s}$ under method A

Table A.26: Experiment 6 - Properties of $\widehat{AUC}^*_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.463	0.618	0.701	0.755	0.820	0.860	0.887	0.831	0.772
Mean	0.470	0.616	0.696	0.748	0.813	0.853	0.880	0.823	0.765
Bias ($\times 10^{-3}$)	7.223	-2.139	-5.305	-6.649	-7.443	-7.233	-6.445	-8.144	-6.560
SD($\times 10^{-2}$)	9.068	6.363	4.675	3.619	2.446	1.840	1.472	3.473	4.031

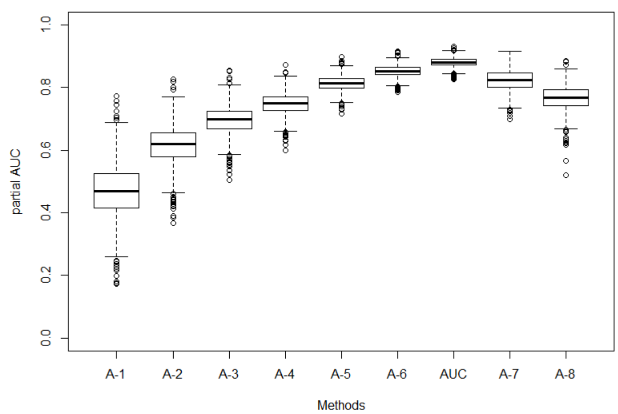

Figure A.43: Experiment 6 - Boxplot showing the distribution of ($\widehat{AUC}^*_{0,s}$) under method A

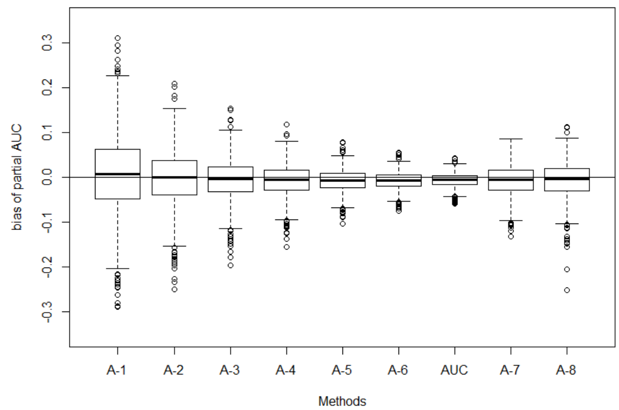

Figure A.44: Experiment 6 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method A

A.5.3 $\widehat{AUC}_{0,s}$ under method B

Table A.27: Experiment 6 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
Valid	7	193	612	935	1000	1000	1000	1000	1000
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.046	0.124	0.210	0.302	0.492	0.688	0.887	0.551	0.348
Mean	0.048	0.122	0.207	0.297	0.486	0.681	0.880	0.545	0.342
Bias ($\times 10^{-3}$)	1.701	-1.982	-3.345	-4.862	-6.390	-6.385	-6.445	-6.296	-6.570
SD ($\times 10^{-2}$)	0.726	1.037	1.338	1.531	1.542	1.426	1.472	11.832	5.688

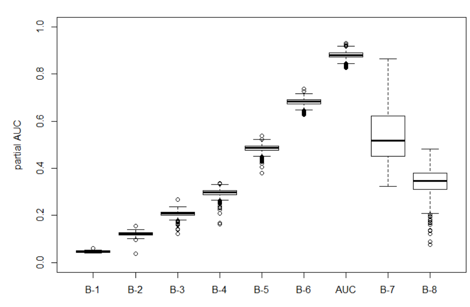

Figure A.45: Experiment 6 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method B

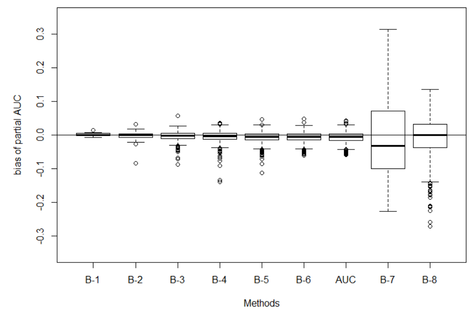

Figure A.46: Experiment 6 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method B

Table A.28: Experiment 6 - Properties of $\hat{a}$ (log-odds ratio) under method B

	a1	a2	a3	a4	a5	a6	a	a7	a8
Mean	3.027	3.059	3.018	2.978	2.941	2.931	2.934	2.935	2.945
Bias	0.552	0.098	0.049	0.040	0.038	0.035	0.041	0.036	0.038
SD	0.743	0.313	0.221	0.201	0.195	0.187	0.201	0.190	0.195

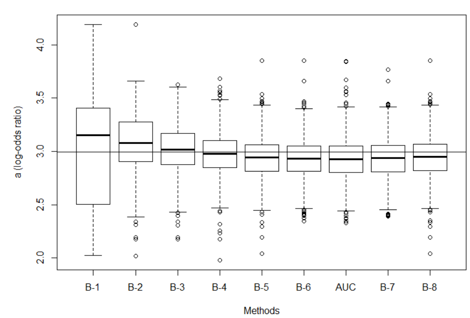

Figure A.47: Experiment 6 - Boxplot showing the distribution of $\hat{a}$ (log-odds ratio) under method B

Table A.29: Experiment 6 - Properties of $\hat{b}$ (heterogeneity) under method B

	b1	b2	b3	b4	b5	b6	b	b7	b8
Mean	-0.021	0.051	0.059	0.049	0.045	0.029	-0.038	0.019	0.044
Bias	0.105	0.043	0.029	0.020	0.013	0.011	0.012	0.012	0.015
SD	0.324	0.208	0.171	0.140	0.116	0.104	0.111	0.109	0.121

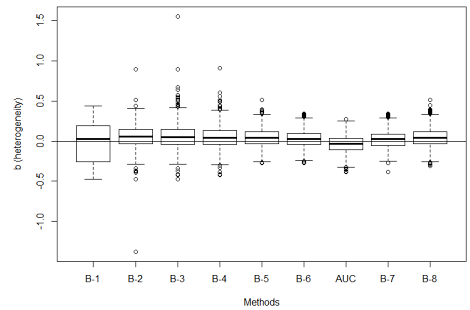

Figure A.48: Experiment 6 - Boxplot showing the distribution of $\hat{b}$ (heterogeneity) under method B

A.5.4 $\widehat{AUC}^{*}_{0,s}$ under method B

Table A.30: Experiment 6 - Properties of $\widehat{AUC}^{*}_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.463	0.618	0.701	0.755	0.820	0.860	0.887	0.831	0.772
Mean	0.480	0.608	0.690	0.742	0.810	0.852	0.880	0.821	0.757
bias ($\times 10^{-3}$)	17.012	-9.910	-11.149	-12.155	-10.650	-7.982	-6.445	-9.924	-14.435
SD($\times 10^{-2}$)	7.263	5.183	4.460	3.827	2.569	1.782	1.472	3.568	4.255

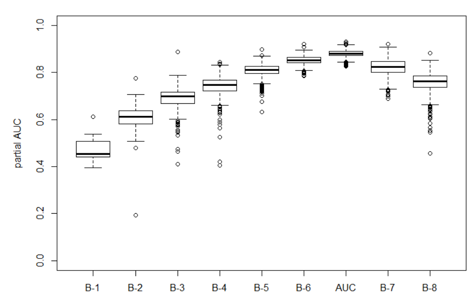

Figure A.49: Experiment 6 - Boxplot showing the distribution of $\widehat{AUC}^{*}_{0,s}$ under method B

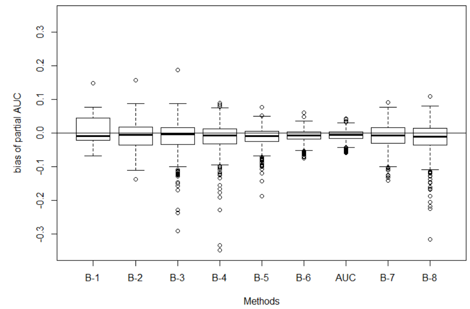

Figure A.50: Experiment 6 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method B

A.6 Experiment 7 - 50 cases and 200 non-cases per study, beta (5,2) distribution of specificities, a=ln2, b=0

A.6.1 $\widehat{AUC}_{0,s}$ under method A

Table A.31: Experiment 7 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.418	0.157
Mean	0.010	0.036	0.076	0.128	0.262	0.426	0.615	0.420	0.158
bias ($\times 10^{-3}$)	0.561	0.919	1.115	1.280	1.578	1.560	1.136	1.488	1.277
SD($\times 10^{-2}$)	0.366	0.880	1.322	1.648	2.008	2.169	2.229	3.031	2.285

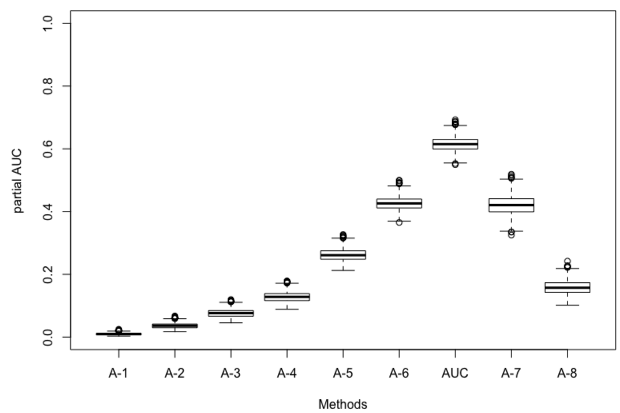

Figure A.51: Experiment 7 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method A

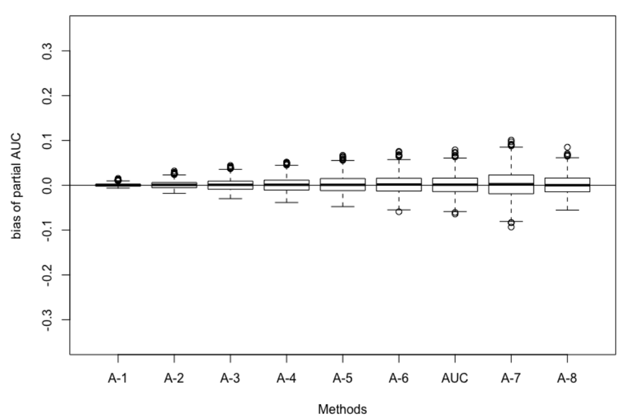

Figure A.52: Experiment 7 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method A

A.6.2 $\widehat{AUC}^*_{0,s}$ under method A

Table A.32: Experiment 7 - Properties of $\widehat{AUC}^*_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.527	0.348
Mean	0.099	0.181	0.255	0.321	0.436	0.532	0.615	0.529	0.351
Bias ($\times 10^{-3}$)	5.612	4.595	3.716	3.199	2.631	1.950	1.136	1.987	3.097
SD($\times 10^{-2}$)	3.664	4.400	4.408	4.120	3.346	2.711	2.229	2.861	4.075

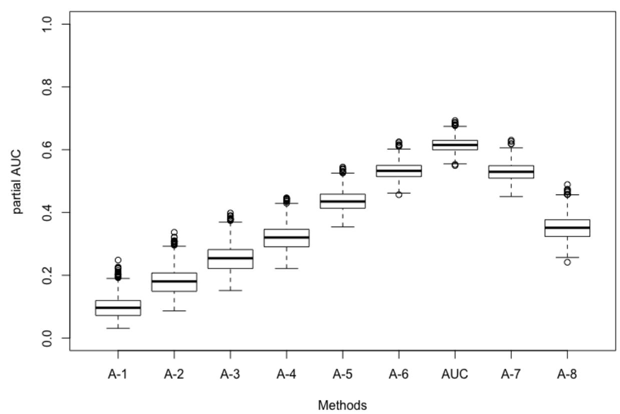

Figure A.53: Experiment 7 - Boxplot showing the distribution of ($\widehat{AUC}^*_{0,s}$) under method A

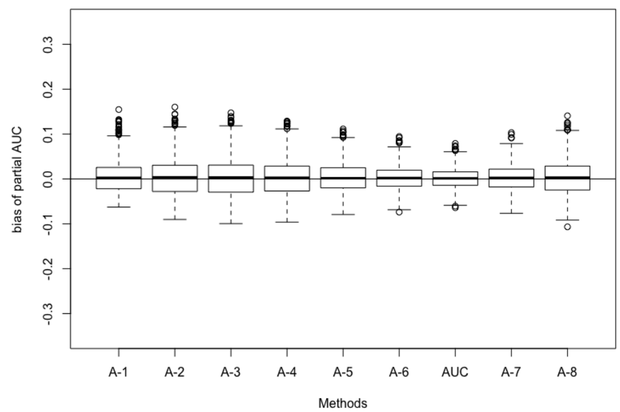

Figure A.54: Experiment 7 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method A

A.6.3 $\widehat{AUC}_{0,s}$ under method B

Table A.33: Experiment 7 - Properties of $\widehat{AUC}_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
Valid	0	0	2	922	1000	1000	1000	1000	1000
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.418	0.157
Mean	NaN	NaN	0.077	0.131	0.262	0.427	0.615	0.420	0.159
bias ($\times 10^{-3}$)	NaN	NaN	1.864	4.419	1.831	2.911	1.136	1.803	1.484
SD($\times 10^{-2}$)	NaN	NaN	0.258	1.361	1.561	1.955	2.229	2.845	1.997

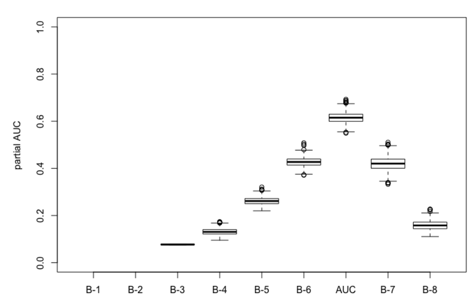

Figure A.55: Experiment 7 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method B

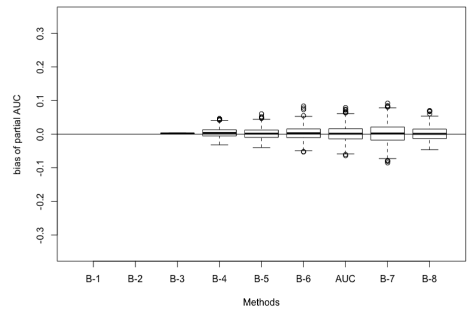

Figure A.56: Experiment 7 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method B

Table A.34: Experiment 7 - Properties of $\hat{a}$ (log-odds ratio) under method B

	a1	a2	a3	a4	a5	a6	a	a7	a8
Mean	NaN	NaN	0.804	0.736	0.696	0.711	0.703	0.704	0.696
Bias	NA	NA	0.042	0.045	0.023	0.019	0.020	0.018	0.023
SD	NA	NA	0.204	0.213	0.151	0.140	0.141	0.133	0.151

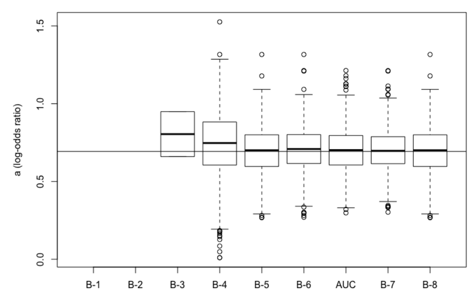

Figure A.57: Experiment 7 - Boxplot showing the distribution of $\hat{a}$ (log-odds ratio) under method B

Table A.35: Experiment 7 - Properties of $\hat{b}$ (heterogeneity) under method B

	b1	b2	b3	b4	b5	b6	b	b7	b8
Mean	NaN	NaN	0.031	-0.005	-0.009	-0.003	-0.003	-0.005	-0.009
Bias	NA	NA	0.004	0.015	0.013	0.010	0.006	0.009	0.013
SD	NA	NA	0.064	0.124	0.114	0.102	0.080	0.093	0.114

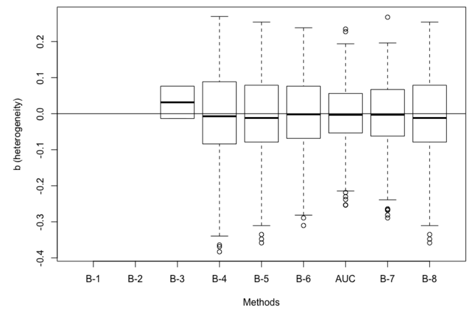

Figure A.58: Experiment 7 - Boxplot showing the distribution of $\hat{b}$ (heterogeneity) under method B

A.6.4 $\widehat{AUC}^*_{0,s}$ under method B

Table A.36: Experiment 7 - Properties of $\widehat{AUC}^*_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.527	0.348
Mean	NaN	NaN	0.257	0.329	0.436	0.534	0.615	0.530	0.352
bias ($\times 10^{-3}$)	NaN	NaN	6.215	11.048	3.051	3.639	1.136	2.401	3.619
SD($\times 10^{-2}$)	NaN	NaN	0.858	3.402	2.602	2.444	2.229	2.569	3.307

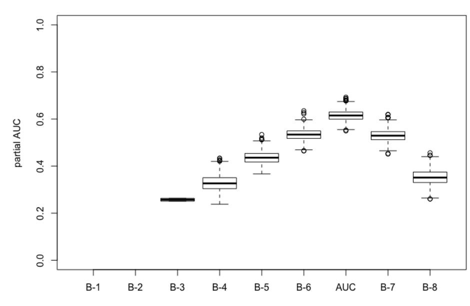

Figure A.59: Experiment 7 - Boxplot showing the distribution of $\widehat{AUC}^*_{0,s}$ under method B

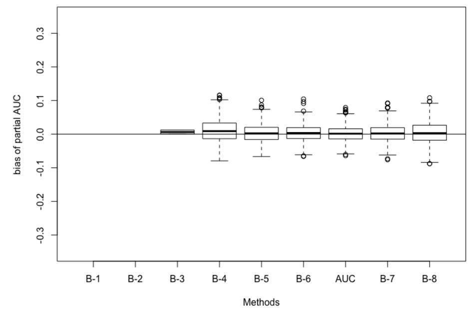

Figure A.60: Experiment 7 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method B

A.7 Experiment 8 - 50 cases and 200 non-cases per study, uniform distribution of specificities, a=ln2, b=0

A.7.1 $\widehat{AUC}_{0,s}$ under method A

Table A.37: Experiment 8 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.419	0.156
Mean	0.010	0.036	0.076	0.128	0.261	0.425	0.614	0.420	0.157
bias ($\times 10^{-3}$)	0.515	0.814	0.945	1.030	1.116	0.865	0.311	0.742	0.943
SD($\times 10^{-2}$)	0.343	0.819	1.222	1.515	1.856	2.059	2.163	3.002	2.277

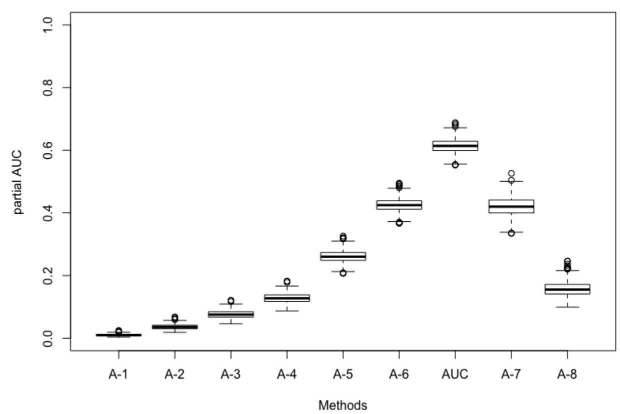

Figure A.61: Experiment 8 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method A

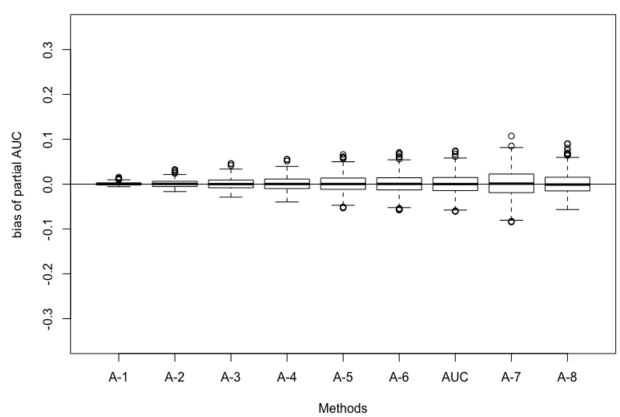

Figure A.62: Experiment 8 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method A

A.7.2 $\widehat{AUC}^*_{0,s}$ under method A

Table A.38: Experiment 8 - Properties of $\widehat{AUC}^*_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.527	0.348
Mean	0.099	0.181	0.254	0.320	0.435	0.532	0.614	0.528	0.350
Bias ($\times 10^{-3}$)	5.146	4.071	3.150	2.575	1.860	1.081	0.311	1.053	2.201
SD($\times 10^{-2}$)	3.431	4.094	4.072	3.788	3.094	2.573	2.163	2.733	3.890

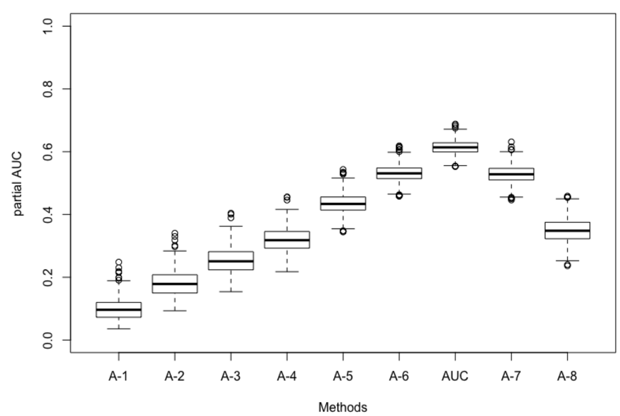

Figure A.63: Experiment 8 - Boxplot showing the distribution of ($\widehat{AUC}^*_{0,s}$) under method A

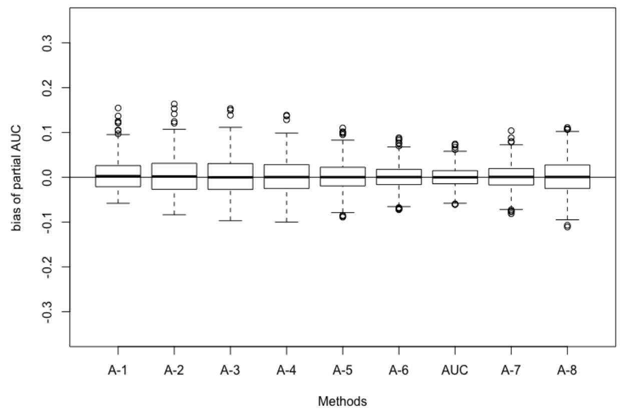

Figure A.64: Experiment 8 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method A

A.7.3 $\widehat{AUC}_{0,s}$ under method B

Table A.39: Experiment 8 - Properties of $\widehat{AUC}_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
Valid	0	0	3	917	1000	1000	1000	1000	1000
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.419	0.156
Mean	NaN	NaN	0.086	0.131	0.261	0.426	0.614	0.420	0.157
bias ($\times 10^{-3}$)	NaN	NaN	11.227	4.099	0.935	1.996	0.311	0.813	0.852
SD($\times 10^{-2}$)	NaN	NaN	0.991	1.262	1.439	1.964	2.163	2.863	1.998

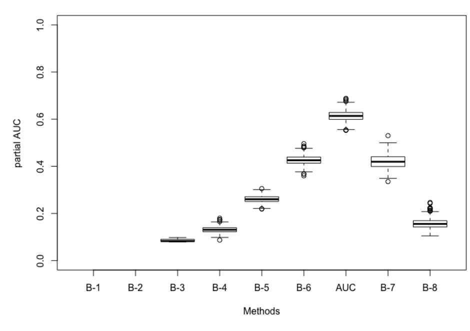

Figure A.65: Experiment 8 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method B

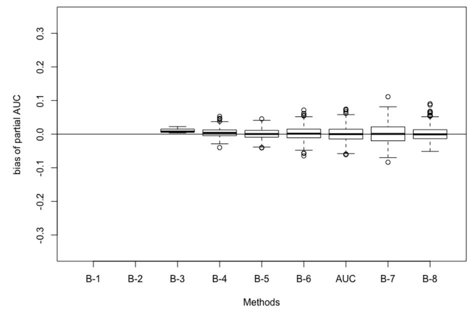

Figure A.66: Experiment 8 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method B

Table A.40: Experiment 8 - Properties of $\hat{a}$ (log-odds ratio) under method B

	a1	a2	a3	a4	a5	a6	a	a7	a8
Mean	NaN	NaN	0.930	0.734	0.686	0.704	0.698	0.697	0.686
Bias	NA	NA	0.017	0.047	0.023	0.021	0.019	0.017	0.023
SD	NA	NA	0.129	0.216	0.152	0.144	0.137	0.132	0.152

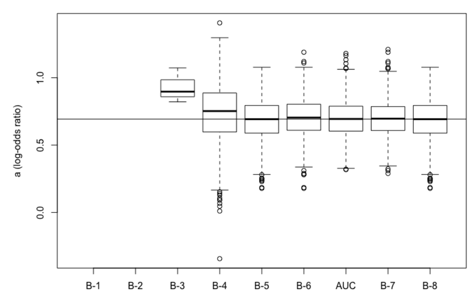

Figure A.67: Experiment 8 - Boxplot showing the distribution of $\hat{a}$ (log-odds ratio) under method B

Table A.41: Experiment 8 - Properties of $\hat{b}$ (heterogeneity) under method B

	b1	b2	b3	b4	b5	b6	b	b7	b8
Mean	NaN	NaN	0.012	-0.004	-0.010	-0.004	-0.004	-0.006	-0.010
Bias	NA	NA	0.000	0.015	0.012	0.010	0.006	0.008	0.012
SD	NA	NA	0.021	0.121	0.112	0.101	0.078	0.091	0.112

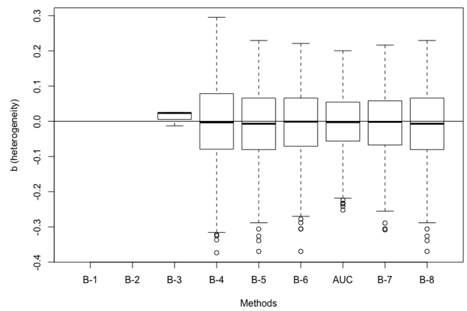

Figure A.68: Experiment 8 - Boxplot showing the distribution of $\hat{b}$ (heterogeneity) under method B

A.7.4 $\widehat{AUC}^*_{0,s}$ under method B

Table A.42: Experiment 8 - Properties of $\widehat{AUC}^*_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.527	0.348
Mean	NaN	NaN	0.288	0.328	0.435	0.533	0.614	0.529	0.350
bias ($\times 10^{-3}$)	NaN	NaN	37.423	10.248	1.559	2.495	0.311	1.160	2.093
SD($\times 10^{-2}$)	NaN	NaN	3.303	3.156	2.399	2.455	2.163	2.512	3.149

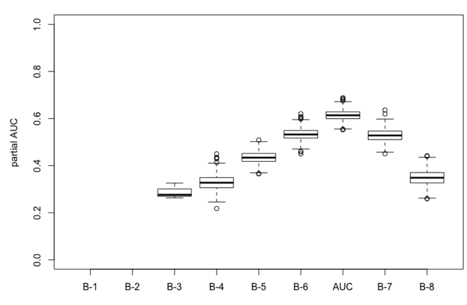

Figure A.69: Experiment 8 - Boxplot showing the distribution of $\widehat{AUC}^*_{0,s}$ under method B

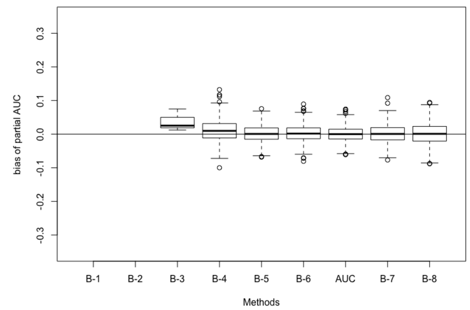

Figure A.70: Experiment 8 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method B

A.8 Experiment 9 - uniform (40,60) cases and uniform (160,240) non-cases per study, mixed uniform distribution of specificities, a=ln2, b=0

A.8.1 $\widehat{AUC}_{0,s}$ under method A

Table A.43: Experiment 9 - Properties of $\widehat{AUC}_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.319	0.160
Mean	0.010	0.036	0.076	0.127	0.259	0.423	0.611	0.317	0.160
bias ($\times 10^{-3}$)	0.331	0.383	0.255	0.052	-0.554	-1.524	-2.467	-1.639	-0.318
SD($\times 10^{-2}$)	0.282	0.656	0.968	1.218	1.677	2.151	2.425	9.649	3.517

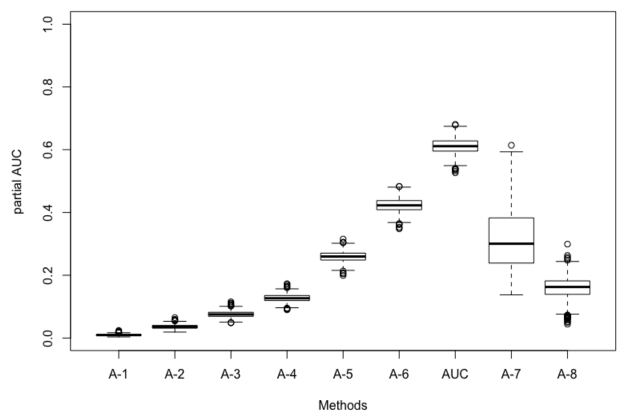

Figure A.71: Experiment 9 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method A

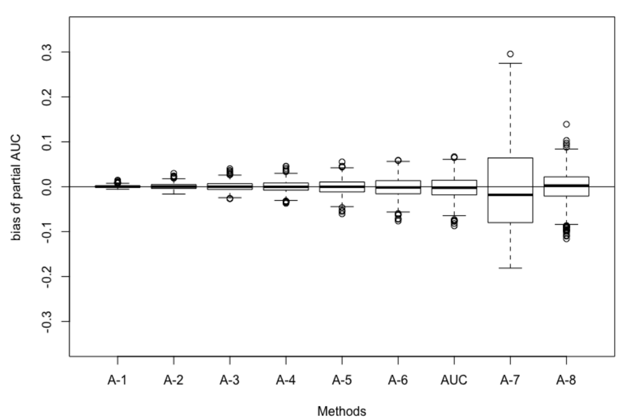

Figure A.72: Experiment 9 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method A

A.8.2 $\widehat{AUC}^*_{0,s}$ under method A

Table A.44: Experiment 9 - Properties of $\widehat{AUC}^*_{0,s}$ under method A

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.466	0.349
Mean	0.097	0.179	0.252	0.318	0.432	0.529	0.611	0.464	0.349
Bias ($\times 10^{-3}$)	3.307	1.914	0.850	0.130	-0.923	-1.906	-2.467	-1.706	-0.359
SD($\times 10^{-2}$)	2.816	3.279	3.227	3.045	2.794	2.689	2.425	6.112	4.327

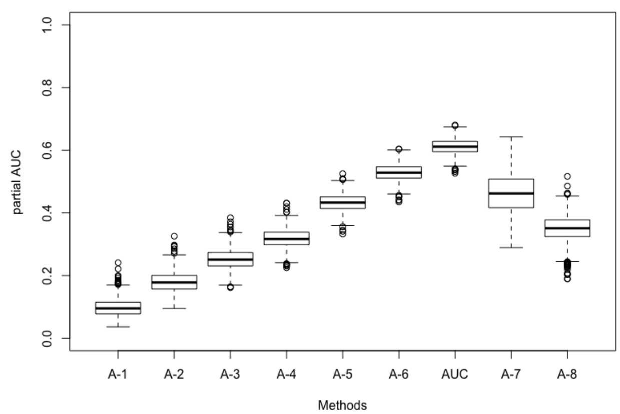

Figure A.73: Experiment 9 - Boxplot showing the distribution of $(\widehat{AUC}^*_{0,s})$ under method A

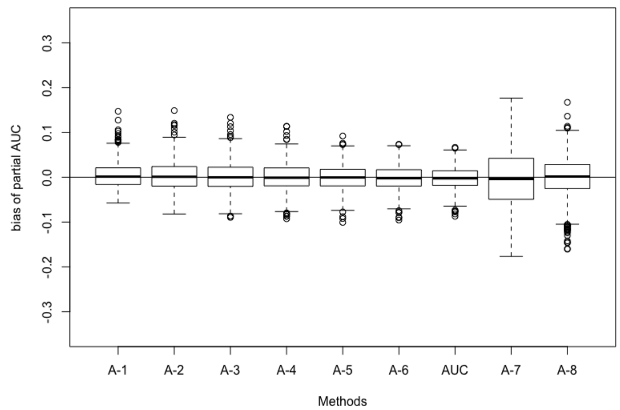

Figure A.74: Experiment 9 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method A

A.8.3 $\widehat{AUC}_{0,s}$ under method B

Table A.45: Experiment 9 - Properties of $\widehat{AUC}_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
Valid	10	176	583	944	1000	1000	1000	1000	1000
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.009	0.035	0.075	0.127	0.260	0.424	0.614	0.319	0.160
Mean	0.010	0.038	0.078	0.130	0.262	0.426	0.611	0.320	0.161
bias ($\times 10^{-3}$)	0.376	2.386	2.731	2.561	1.851	1.489	-2.467	0.971	0.371
SD($\times 10^{-2}$)	0.249	0.677	1.135	1.295	1.703	2.051	2.425	9.864	3.492

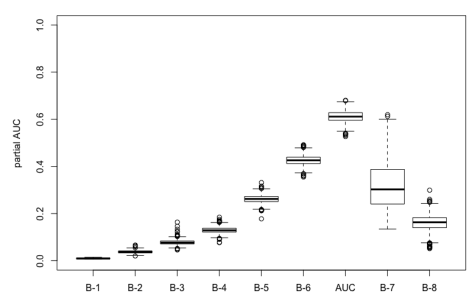

Figure A.75: Experiment 9 - Boxplot showing the distribution of $\widehat{AUC}_{0,s}$ under method B

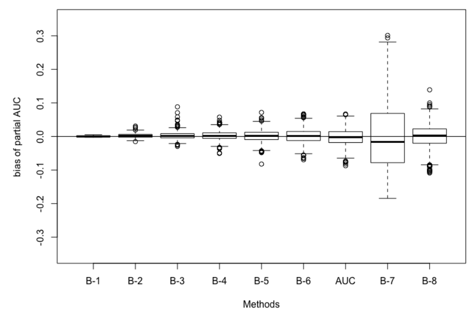

Figure A.76: Experiment 9 - Boxplot showing the distribution of bias ($\widehat{AUC}_{0,s}$) under method B

Table A.46: Experiment 9 - Properties of $\hat{a}$ (log-odds ratio) under method B

	a1	a2	a3	a4	a5	a6	a	a7	a8
Mean	0.880	0.881	0.798	0.753	0.714	0.704	0.680	0.706	0.714
Bias	1.069	0.704	0.260	0.098	0.034	0.022	0.023	0.024	0.041
SD	1.034	0.839	0.510	0.312	0.183	0.149	0.153	0.154	0.203

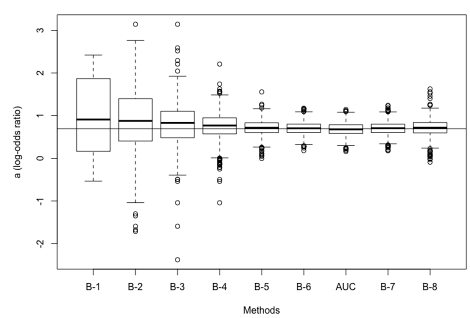

Figure A.77: Experiment 9 - Boxplot showing the distribution of $\hat{a}$ (log-odds ratio) under method B

Table A.47: Experiment 9 - Properties of $\hat{b}$ (heterogeneity) under method B

	b1	b2	b3	b4	b5	b6	b	b7	b8
Mean	0.039	0.048	0.033	0.022	0.012	0.007	-0.006	0.007	0.012
Bias	0.044	0.054	0.031	0.019	0.010	0.007	0.006	0.007	0.011
SD	0.209	0.233	0.176	0.138	0.099	0.085	0.076	0.084	0.106

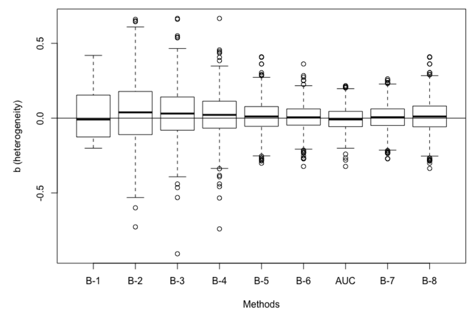

Figure A.78: Experiment 9 - Boxplot showing the distribution of $\hat{b}$ (heterogeneity) under method B

A.8.4 $\widehat{AUC}^*_{0,s}$ under method B

Table A.48: Experiment 9 - Properties of $\widehat{AUC}^*_{0,s}$ under method B

	p1	p2	p3	p4	p5	p6	AUC	p7	p8
s	0.1	0.2	0.3	0.4	0.6	0.8	1		
Theoretical	0.094	0.177	0.251	0.318	0.433	0.531	0.614	0.466	0.349
Mean	0.098	0.189	0.260	0.324	0.436	0.532	0.611	0.468	0.351
bias ($\times 10^{-3}$)	3.761	11.929	9.105	6.403	3.085	1.861	-2.467	1.763	1.212
SD	0.025	0.034	0.038	0.032	0.028	0.026	0.024	0.063	0.042
SD($\times 10^{-2}$)	2.485	3.387	3.785	3.238	2.839	2.563	2.425	6.282	4.181

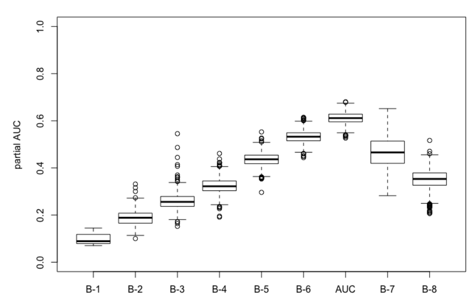

Figure A.79: Experiment 9 - Boxplot showing the distribution of $\widehat{AUC}^*_{0,s}$ under method B

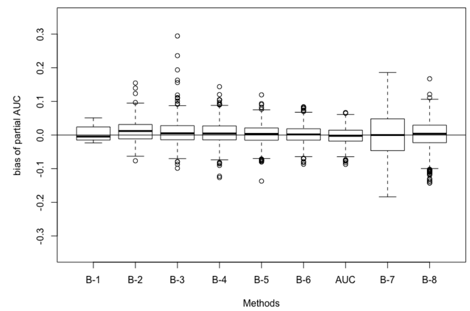

Figure A.80: Experiment 9 - Boxplot showing the distribution of bias ($\widehat{AUC}^*_{0,s}$) under method B

Bibliography

Fleiss, J. L., Levin, B., and Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*. John Wiley & Sons, Inc., 3 edition.

Irwig, L., Glasziou, P., and Fahey, M. (1993). Meta-analytic methods for disgnostic test accuracy. *Journal of Clinical Epidemiology*, **48**(1), 119–130.

Moses, L. E. and Shapiro, D. (1993). Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Statistics in Medicine*, **12**, 1293–1316.

Walter, S. D. (2002). Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Statistics in Medicine*, **21**, 1237–1256.

Walter, S. D. (2005). The partial area under the summary ROC curve. *Statistics in Medicine*, **25**, 2025–2040.