Open University of Catalonia (UOC) Master's Degree in Data Science

MASTER'S THESIS

Machine Learning and Computer Vision in Healthcare and Medical Applications

Assessing a Novel Approach to Cut-off Estimation Using Bootstrapping

A Simulation-Based Study

Author: Lucia Blanc Velázquez

Tutor: Cristian Tebé Cordomí

Professor: Laia Subirats Maté

Barcelona, May 27, 2025

Credits/Copyright

FINAL PROJECT RECORD

Title of the project:	Assessing a Novel Approach to Cut-off Estimation Using Bootstrapping: A Simulation-Based Study
Author's name:	Lucia Blanc Velázquez
Collaborating teacher's name:	Cristian Tebé Cordomí
PRA's name:	Laia Subirats Maté
Delivery date (mm/yyyy):	05/2025
Degree or program:	Data Science Master
Final Project area:	Machine Learning and Computer Vision in Healthcare and Medical Applications
Language of the project:	English
Keywords	Biological Cut-off Optimization, Bootstrapping Resampling, Akaike Information Criterion

Quote

. . . wherever nature draws unclear boundaries, humans are happy to curate.

— Alice Dreger, Galileo's Middle Finger

Acknowledgements

First of all, I would like to thank my supervisors, Dr. Cristian Tebé and Dr. Natàlia Pallarès, for their guidance, patience, and insightful feedback throughout this project. Their support has been fundamental to my learning and growth.

I am also grateful to my colleagues, João and Pau, for their collaboration and help during the process.

Lastly, I want to thank my family for always being by my side. For their presence, understanding, and trust in every decision I make.

Glossary

- **AIC**: Akaike's information criterion.
- **AUC**: Area under the curve.
- **CFN**: Cost false negative.
- **CFP**: Cost false positive.
- **DOR**: Diagnostic odds ratio.
- **fm**: Cost index.
- **FN**: False negative.
- **FP**: False positive.
- **GDPR**: General Data Protection Regulation.
- **IU**: Index of union.
- **KDE**: Kernel density estimation.
- **LR$^+$**:Positive likelihood ratio.
- **LR$^-$**: Negative likelihood ratio.
- **MCT**: Misclassification cost term.
- **MSE**: Mean squared error.
- **NPV**: Negative predictive value.
- **RR**: Relative risk.
- **PPV**: Positive predictive value.
- **ROC**: Receiver operating characteristic.
- **Se**: Sensitivity.

- **Sp**: Specificity.
- **SDG**: Sustainable development goal.
- **TN**: True negative.
- **TP**: True positive.

Abstract

Finding a cut-off for biological markers is essential for clinical diagnosis and prognosis in many pathologies. It allows population stratification and provides an objective criterion for risk assessment and treatment selection. However, establishing reliable and reproducible cut-offs remains challenging due to the uncertainty about the appropriateness of dichotomizing continuous variables and the inherent dependence on specific data, which limits generalizability across populations.

Classical methods often introduce biases, such as the Youden index assumption of a single optimal threshold, leading to inconsistent results and overestimation of the biological marker effect, compromising statistical precision and power, resulting in a significant loss of information.

To overcome these challenges, we propose a cut-off point selection method that explores the full range of a biological marker values and evaluates their discriminatory ability using a log-binomial model. The process is then repeated using bootstrap resampling to estimate the empirical distribution of the possible cut-off point, improving robustness and reliability by prioritizing model plausibility over p-values using Akaike's information criterion. The resulting empirical distribution of bootstrapped cut-off points is visualized to identify optimal and stable thresholds.

To evaluate the proposed method, a simulation study is performed under three scenarios (null, small and moderate effect), considering a normal distribution with all analyses implemented and evaluated using R software.

This approach provides a statistical alternative to classical methods, and future research should explore its extension to more complex distributions and its applicability in real clinical settings.

Keywords: Biological cut-off points optimization, bootstrapping resampling, Akaike Information Criterion, biostatistics.

Resumen

Definir puntos de corte en biomarcadores continuos es fundamental para el diagnóstico clínico y la estratificación de pacientes, ya que, permite establecer criterios objetivos para evaluar riesgos y orientar decisiones terapéuticas. Sin embargo, seleccionar umbrales fiables y reproducibles sigue siendo un desafío, debido a la incertidumbre sobre la validez de dicotomizar variables continuas, y a la dependencia de datos específicos que limita la generalización.

Los métodos clásicos, como el índice de Youden, suelen asumir la existencia de un único umbral óptimo, lo que puede introducir sesgos, sobreestimar la asociación marcador-resultado, y reducir la precisión y potencia estadística. Para abordar estas limitaciones, proponemos un nuevo enfoque basado en modelos log-binomiales y selección mediante el Criterio de Información de Akaike, que escanea todo el rango de valores del biomarcador. Este procedimiento se integra en un esquema de remuestreo bootstrap, lo que permite estimar la distribución empírica de los puntos de corte y evaluar su estabilidad.

Para evaluar el método propuesto, se realiza un estudio de simulación bajo tres escenarios (efecto nulo, pequeño y moderado), considerando una distribución normal, con todos los análisis implementados y evaluados utilizando el software R.

El enfoque propuesto proporciona una estrategia más robusta y adaptable para la selección de puntos de corte, al incorporar la variabilidad de los datos y ofrecer estimaciones acompañadas de medidas de incertidumbre. Futuros estudios deberían evaluar su aplicabilidad a distribuciones más complejas y su desempeño en datos clínicos reales.

Palabras clave: Optimización de puntos de corte biológicos, remuestreo bootstrapping, Criterio de Información de Akaike, bioestadística.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and motivation

In recent years, the use of biological markers in biomedical research and clinical practice has increased significantly. These biological markers play a crucial role in the diagnosis, prognosis, and monitoring of diseases. However, the precise determination of cut-off points remains a challenging aspect, although dichotomization of continuous variables can offer certain advantages in specific contexts, it may also lead to a loss of valuable information and an overestimation of the relationship between a biological marker and a clinical outcome, thus compromising the validity of medical decisions (Harrell Jr. [2025]; Weinberg [1995]). Furthermore, different approaches can lead to not consistent results and affect patient stratification (Chang et al. [2017]).

Classical cut-off point selection methods are based on statistical criteria, many of which are derived from the ROC curve. Among the most commonly used are the Youden's index, which maximizes the sum of sensitivity (Se) and specificity (Sp); the Euclidean distance to the ideal point (0, 1) in the ROC curve; and the minimum p-value approach, which identifies the threshold with the greatest statistical significance in group separation. However, these methods have been shown to introduce bias and compromise the robustness of the results (Polley and Dignam [2021]; Duarte [2021]), and do not always accurately reflect the clinical relevance of the selected threshold, limiting their applicability in medical practice (Weinberg [1995]; Vickers et al. [2008]). In other cases, thresholds are based on predefined medical standards or statistical rules that may not generalize adequately to specific data sets.

As the use of biological markers in clinical decision-making expands, well-defined and statistically robust thresholds are essential to maximize their utility and diagnostic precision. Therefore, it is crucial to explore alternative statistical strategies that improve both the reliability

and applicability of cut-off points.

In this study, we propose a novel strategy based on exploring the density distribution of a biological marker and selecting the optimal cut-off based on a bootstrap process, considering both the mean and the mode of the resampled distribution. The motivation lies in the need for more robust and adaptive strategies that enhance clinical usefulness.

Ultimately, this work aims to develop a statistically robust and computationally efficient methodological approach for the determination of biological marker cut-off points. This would contribute to more reliable and adaptive tools that improve diagnostic accuracy and support better clinical outcomes.

1.2 Goals

1.2.1 Main goal

The main objective of this study is to evaluate a novel methodological approach to determine cut-off points in continuous biological markers associated with dichotomous outcomes, aiming to improve statistical robustness and clinical applicability compared to classical methods.

1.2.2 Specific goals

To achieve this, the study will address the following specific goals:

- **Review and contextualize the state of the art**: To conduct a review of classic cut-off estimation methods, identify their strengths and limitations, and provide a basis for methodological comparison.

- **Simulation of data**: To develop a simulated dataset of a continuous biological marker with a dichotomous outcome under different scenarios to assess the robustness and reliability of the proposed method.

- **Compare and evaluate the results**: To apply the new approach to the simulated data and compare its performance with classical methods identified in the review.

1.3 Sustainability, diversity, and ethical/social challenges

Scientific and technological advancements must be assessed not only for their effectiveness but also for their sustainability, equity, and ethical implications. This section examines the potential environmental, social, and ethical impacts of this project, considering its role in healthcare decision-making and its implications for data security, accessibility, and professional practice.

1.3.1 Sustainability

This project has an indirect environmental impact because it is based on simulation data and statistical analysis without significant consumption of physical resources. However, although it is developed in a digital environment, computational processes and server usage contribute to energy consumption. To mitigate this impact, strategies such as code optimization and the use of energy-efficient servers can be implemented.

Furthermore, optimizing cut-off points for biological markers enhances disease diagnosis and treatment, which can help to avoid unnecessary use of medical resources, such as redundant tests and ineffective treatments. This, in turn, can reduce the environmental footprint of the healthcare sector.

This project aligns with Sustainable Development Goal (SDG) 3: Health and Well-being by improving the accuracy of disease detection through optimized biological markers, thereby promoting more efficient and sustainable healthcare.

1.3.2 Diversity

It is important to evaluate whether optimized cut-off points work equally well in different populations. Some biological markers may vary by sex, age, race or social status, which could lead to diagnostic bias if it is not taken into account.

From an accessibility and diversity perspective, these tools should be implemented in health systems in a way that ensures that they benefit the entire population, not just those with access to advanced medical services.

1.3.3 Ethical/Social challenges

While this project uses simulated data, its application in real-world environments would involve the handling of sensitive biomedical data, which introduces security risks. Therefore, robust measures must be in place to preserve patient privacy and ensure data integrity, in compliance

with regulations such as the General Data Protection Regulation (GDPR) in the European Union (Regulation (EU) 2016/679) (European Union [2016]).

The estimation of cut-off points can significantly influence critical medical decisions. Therefore it is essential that methods used are both reproducible and explainable to minimize bias in result interpretation.

Although these tools can improve the efficiency of medical decision-making, their automation could reduce the need for certain diagnostic procedures by healthcare professionals. However, their use is intended as an additional tool to improve diagnostic accuracy without replacing clinical judgment.

1.4 Approach and methodology

To develop and assess the proposed approach, a data simulation study will be conducted using a continuous variable as a biological marker and a dichotomous variable to indicate the presence or absence of a disease. Different scenarios will be evaluated to explore its potential applicability and robustness.

1.4.1 Study design

The simulation will generate a continuous biological marker using a normal distribution $N(\mu, \sigma^2)$. This type of distribution is commonly used in simulations because it is easy to interpret and many biological measures follow this distribution in real populations, and it is also a good starting point for evaluating new methods under controlled and ideal conditions.

The methodology will be tested in different scenarios to evaluate its ability to manage diverse real-world conditions, including:

- **Null scenario**: No separation between groups. When there is no real distinction between the groups, the methodology is expected to fail to identify a valid cut-off point, indicating no effect. This test will ensure that the approach does not force the identification of a cut-off when none should exist.
- **Small discrimination scenario:** Minimal separation between groups. Classical methods, such as Youden's index, might still detect a cut-off in these cases, even if it is statistically insignificant or clinically irrelevant. The proposed bootstrapping-based method is expected to demonstrate improved stability and prevent the false identification of cut-off

points in low-separation contexts, ensuring thresholds will be only selected when they reflect meaningful differences.

- **Moderate discrimination scenario:** Partial but noticeable separation between groups. This scenario will test whether the proposed methodology can correctly identify meaningful and stable cut-off points when a true and clinically relevant difference exists. It will also help demonstrate the method's ability to discriminate between relevant and insignificant scenarios.

To determine the required sample sizes for statistically reliable results, a power analysis will be conducted using the *Power and Sample Size* software (PS [2025]). All scenarios will assume a prevalence of 10%, with sample size N varying depending on the scenario:

- Null scenario $\rightarrow N = 2000$
- Small scenario $\rightarrow N = 1000$
- Moderate scenario $\rightarrow N = 500$

Each scenario will be simulated **50** times using different seeds to ensure robustness and consistency. This will help assess the stability and generalizability of the proposed method across repeated samples.

1.4.2 Methodology for determining optimal cut-off points

In each simulated scenario, we will implement the following process:

1. **Biological marker distribution:** Examine the biological marker distribution to assess whether it is appropriate to establish a cut-off point.
2. **Scan biological marker:** Assess all possible values of each biological marker by dichotomizing the variable and ensuring that the tails of the distribution contain a minimum number of events before starting the estimations to avoid biased or null results.
3. **Cut-off estimation:** For each candidate cut-off point, the biological marker is dichotomized and a log-binomial model is fitted as follows:

$$\log(P(D = 1|\mathbb{I}(T \geq \text{cut-off})) = \beta_0 + \beta_1 \cdot \mathbb{I}(T \geq \text{cut-off})$$

where D indicates the presence or absence of a disease as a binary outcome (D = 0 for non-disease and D = 1 for disease population), and T represents the diagnostic test measure, which is the continuous biological marker (positive T^+ when $T \geq$ cutoff, and

negative T^- when $T <$ cutoff). This assumes that higher values are associated with a higher risk of disease.

The optimal cut-off point will be selected as the one with the lowest Akaike Information Criterion (AIC), a metric that balances model fit and complexity (Akaike [1974]), prioritizing model plausibility.

4. **Bootstrapping procedure:** Perform bootstrap resampling to assess the stability and variability of the selected cut-off points. For each resample, steps 2 and 3 are repeated, generating a distribution of optimal cut-offs across multiple iterations.

5. **Assess suitability and stability:** Evaluate the shape of the bootstrapped density function to check if the selected cut-off point is reliable. The analysis will look for signs of variability, multiple peaks, or irregularities that may suggest uncertainty. To summarize the distribution, two values will be considered: (i) the expected value (mean) and (ii) the most frequent (mode). Comparing these two will help decide which cut-off point is more reliable according to previous literature and diagnostic metrics.

6. **Validation:** Validate key diagnostic properties, such as Se, Sp, positive predictive value (PPV), negative predictive value (NPV), and accuracy, to assess the performance of the selected cut-off points in classification and decision-making.

All of these analyses will be performed using R software (R Core Team [2024]).

1.4.3 Challenges and advantages of the proposed methodology

The dichotomization of continuous variables can often oversimplify complex relationships, leading to a loss of valuable information. The dependence on specific data and the variability of cut-off points may lead to unreliable and inconsistent results.

The proposed methodology addresses these challenges by using bootstrapping to reduce data dependence, which enhances the robustness of the cut-off point estimation process. Furthermore, a critical evaluation of the density function distribution allows for an anticipatory assessment of whether dichotomizing the variable, including determining whether a single cut-off point is sufficient or if multiple thresholds are necessary.

Nevertheless, certain limitations must be acknowledged. A key aspect is the extension to more complex distributions; while the normal distribution is considered, real-world biological markers may follow more heterogeneous patterns, requiring further methodological adaptations. Furthermore, as this study is based on simulated data, future research should evaluate the

performance of the proposed methodology using real clinical datasets to confirm its practical utility.

1.4.4 Validation of methodology results

Once the optimal cut-off points are determined, the results will be compared with classical cut-off point estimation methods, including:

1. Youden's J statistic
2. Euclidean distance to the ideal ROC point (0,1)
3. Maximum product of Se and Sp
4. Index of union (IU)
5. Cost-based approach
6. Misclassification cost
7. Diagnostic odds ratio (DOR)
8. Minimum p-value method

Each method will be applied to the same simulated scenarios and their results will be contrasted with those of the proposed methodology to evaluate differences in cut-off point variability, robustness and clinical interpretability.

The analysis will focus on assessing the performance between the proposed approach and existing methods for clinical applications, highlighting their strengths, limitations, and suitability in real-world diagnostic scenarios. The interpretation and validation of the results will be carried out by evaluating key properties. This will provide a comprehensive understanding of the method's diagnostic performance.

1.5 Schedule

The tasks are organized according to the modules outlined in the learning plan. Each module includes subtasks designed to achieve specific goals. Delivering these subtasks marks significant milestones, with each task's end date serving as a partial milestone within its respective module. The most critical task is the model evaluation, which entails advancing through the various components that make up the model.

The project's timeline starts on 13 January 2025, marking the initiation of its definition, and concludes on 27 June 2025, the scheduled date for its public defense. *Figure 1.1* provides a visual summary of the project timeline and its milestones.

1.5.1 Tasks

- **M1 - Definition and planification**
  a. Definition of the idea: 13/01/25 to 24/01/25
  b. Previous research: 25/01/25 to 02/02/25
  c. Collection and analysis of publications: 03/02/25 to 15/02/25
  d. Development of the introduction: 16/02/25 to 27/02/25
  e. Writing PAC1: 28/02/25 to 09/03/25
  f. PAC1 Submission: 09/03/25 (DEADLINE)
- **M2 - State of the art**
  a. Research of data simulation: 03/02/25 to 27/02/25
  b. Documentation of the state of the art: 03/02/25 to 15/03/25
  c. Documentation of materials and methods: 03/02/25 to 15/03/25
  d. Layout and management of the bibliography: 14/02/25 to 09/05/25
  e. Writing PAC2: 21/02/25 to 30/03/25
  f. PAC2 submission: 30/03/25 (DEADLINE)
- **M3 - Implementation**
  a. Testing available models: 20/02/25 to 30/03/25
  b. Model selection: 16/02/25 to 30/03/25
  c. Feature extraction: 31/03/25 to 26/04/25
  d. Training: 05/03/25 to 26/04/25
  e. Model evaluation: 10/03/25 to 30/04/25
  f. Documentation of the results: 05/04/25 to 09/05/25
  g. Writing PAC3: 10/04/25 to 04/05/25

h. PAC3 submission: 04/05/25 (DEADLINE)

- **M4 - Writing the report**

  a. Documentation of the conclusions: 15/02/25 to 09/05/25

  b. Writing preliminary submission PAC4.1: 22/04/25 to 09/05/25

  c. PAC4.1 submission: 09/05/25 (DEADLINE)

  d. Writing final submission PAC4.2: 01/05/25 to 25/05/25

  e. PAC4.2 submission: 25/05/25 (DEADLINE)

  f. Preparation of the final presentation: 26/05/25 to 06/06/25

  g. Submission of the presentation: 03/06/25 (DEADLINE)

- **M5 - Public defense of the work**

  a. Preparation for submission of documentation to the committee: 01/06/25 to 06/06/25

  b. Submission of documentation to the committee: 06/06/25 (DEADLINE)

  c. Rehearsal and preparation for the oral presentation: 03/06/25 to 27/06/25

  d. Public defense of the work: 27/06/25 (DEADLINE)

1.5.2 Gantt diagram

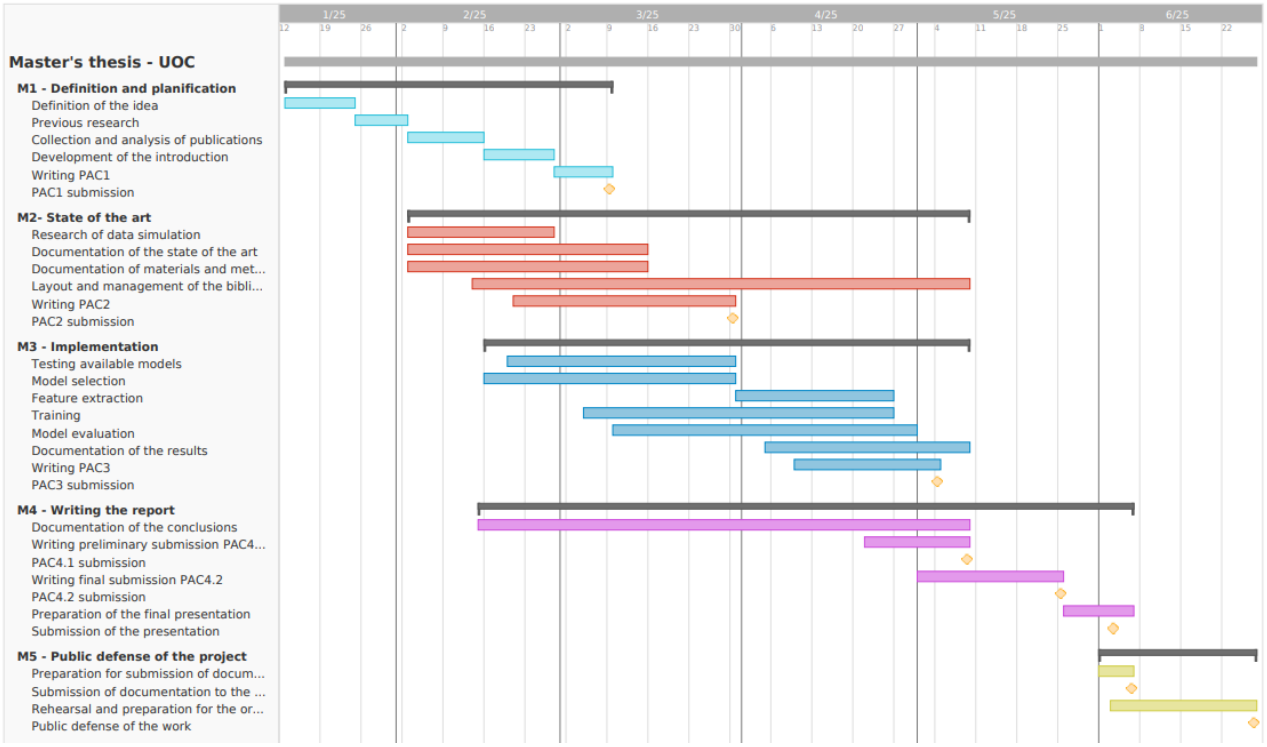


Figure 1.1: Gantt diagram of project timeline.

1.6 Summary of the outputs of the project

This project aims to develop and evaluate a novel method for determining optimal cut-off points by integrating bootstrapping resampling and AIC within a log-binomial model, addressing biases commonly associated with classical dichotomization techniques. The method will be tested on a simulated dataset, ensuring a robust statistical evaluation. By leveraging bootstrapping and AIC for cut-off point estimation, this study seeks to improve its accuracy and reliability.

1. **Algorithm development:** Implementation of a cut-off estimation method based on log-binomial models, AIC minimization, and bootstrapping to assess stability.

2. **Simulation-based evaluation:** The method will be tested across multiple scenarios using simulated datasets that represent different levels of discrimination between groups to assess its robustness and reliability.

3. **Comparison with classical methods:** The results of the proposed method will be compared with those obtained using classical cut-off point estimation techniques to evaluate

performance and interpretability.

1.7 Brief description of the remaining chapters of the report

- **State of the Art:** A review of classical methodologies for optimal cut-off estimation and discussing their application in biological marker-based clinical prediction.
- **Materials and Methods:** Detailed description of the data simulation process, the statistical techniques employed, and the validation metrics used in this study.
- **Results:** Presentation of findings, including the implementation process of the proposed method and the evaluation.
- **Conclusions:** A summary of the key insights derived from the study, along with implications for clinical practice , as well as the identification of its strengths and limitations.
- **References:** A list of the bibliographic sources cited in the previous sections.

Chapter 2

State of the Art

2.1 Diagnostic performance metrics

Quantitative diagnostic tests play a crucial role in both clinical research and decision-making, especially when continuous biological markers are evaluated to diagnose disease or assess risk.

The evaluation of diagnostic tests in clinical practice is fundamental, as it motivates the search for robust statistical methods when standard reference procedures are not feasible. As a result, new diagnostic tools based on laboratory biomarkers are becoming increasingly important, mainly when their results are expressed on a continuous scale (Hassanzad and Hajian-Tilaki [2024]).

In diagnostics with a binary outcome (disease or non-disease), let T represent the value of the continuous biological marker and c the chosen cut-off point. A subject is classified as **test positive** if $T \geq c$, and **test negative** otherwise. This classification leads to a confusion matrix structured as follows:

True Condition	**Test Positive** $(\boldsymbol{T \geq c})$	**Test Negative** $(\boldsymbol{T < c})$
Diseased (D)	True Positive TP	False Negative FN
Non-Diseased $(\bar{D})$	False Positive FP	True Negative TN

Table 2.1: Classification of continuous test results by true condition status.

From this matrix, two key performance metrics dependent on the cut-off point c can be derived:

- **Sensitivity (Se):** Probability of correctly identifying diseased individuals.

$$\mathrm{Se}(c) = P(T \geq c \mid D),$$

- **Specificity (Sp):** Probability of correctly identifying non-diseased individuals.

$$\mathrm{Sp}(c) = P(T < c \mid \bar{D})$$

We assume, without loss of generality, that higher values of biological markers increase the likelihood of disease, and the distributions of these values in diseased and non-diseased individuals are sufficiently different to justify the use of a threshold. This approach facilitates clinical workflows by providing clear decision rules for diagnosis or risk assessment. In addition to Se and Sp, clinicians are also interested in PPV and NPV, which express the probability of the presence or absence of disease given a specific test result. These are calculated as:

- **Positive predictive value (PPV):** Probability of the presence of disease given a specific test result.

$$\mathrm{PPV} = \frac{TP}{TP + FP},$$

- **Negative predictive value (NPV):** Probability of the absence of disease given a specific test result.

$$\mathrm{NPV} = \frac{TN}{TN + FN}$$

These measures are known to be affected by the prevalence of the disease in the target population (Hassanzad and Hajian-Tilaki [2024]).

Nevertheless, the dichotomization of continuous variables remains a controversial practice, as it imposes an artificial boundary on biological processes that are inherently continuous and rarely exhibit abrupt transitions. The estimation of an optimal cut-off is therefore critical, as a reasonable threshold improves patient stratification and clinical outcomes, whereas a poor choice increases the risk of misclassification and reduces diagnostic accuracy. Moreover, several authors have emphasized that dichotomization may distort analyses and compromise predictive validity, often leading to increased false-positive rates and weakened inferential power (Weinberg [1995]; Irwin and McClelland [2003]). As an alternative, some have recommended modeling continuous relationships directly, without categorization, to preserve the richness of the data and enhance interpretability (Vickers et al. [2008]; Hassanzad and Hajian-Tilaki [2024]). Nonetheless, in

clinical practice, binary decision thresholds are often necessary for actionability and communication. Clinicians often require a single threshold value to support decision-making and guide the selection of appropriate clinical management strategies.

This clinical need for clear and actionable decisions has driven the development of several strategies aimed at defining optimal thresholds, each focusing on identifying the point on the continuous scale that best separates the outcome of interest. Early approaches focused primarily on maximizing diagnostic accuracy metrics, such as Se and Sp, while more recent methods include statistical models, cost considerations, and information-based criteria.

In all cases, the objective is to select a threshold that is both statistically sound and clinically meaningful. However, each method relies on specific assumptions and exhibits strengths and limitations that may affect the generalizability of the selected cut-off. For these reasons, a critical evaluation of conventional cut-off determination techniques is necessary to understand their respective strengths, limitations, and clinical applicability.

2.1.1 Receiver Operating Characteristic (ROC) Curves

Many of the diagnostic biomarkers in modern medicine are quantitative, allowing various cut-off points to be established across their range of values. For each potential threshold, corresponding Se and Sp values can be derived. The graphical representation of the balance between Se and Sp across possible cut-offs is known as the Receiver Operating Characteristic (ROC) curve, a tool used to evaluate the diagnostic accuracy of tests (Metz [1978]; Zweig and Campbell [1993]; Pepe [2003]).

The ROC curve is plotted as a unit square, where the y-axis represents TP rate (Se) and the x-axis represents the FP rate (1 - Sp). Each point on the curve corresponds to a different cut-off value. When the threshold is set higher, fewer individuals test positive, increasing specificity but reducing sensitivity. Lower thresholds do the opposite, increasing sensitivity but lowering specificity.

An ideal curve slopes towards the upper left corner, indicating high discrimination with many true detections and few false detections. In contrast, a curve close to the diagonal indicates random classification and a curve below the diagonal reflects reverse or deceptive testing behavior *(Figure 2.1)*.

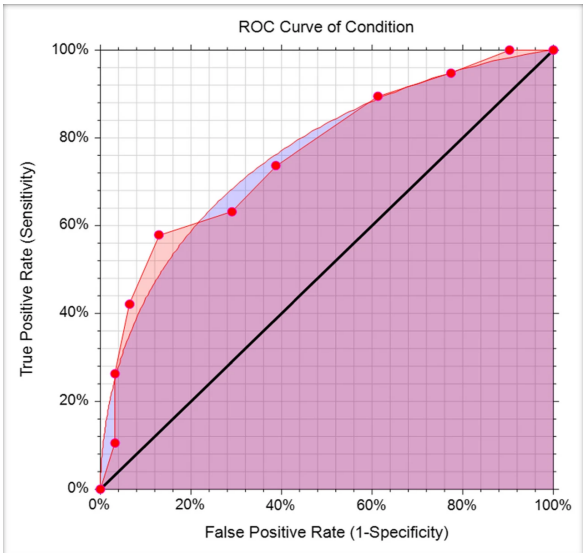


Figure 2.1: Example of a ROC curve.
Image reproduced from Hassanzad and Hajian-Tilaki [2024]

The area under the curve (AUC) is a central summary measure derived from the ROC curve that represents the probability that a randomly selected diseased individual will score higher on the test than a randomly selected non-diseased individual. It can be calculated parametrically using a binomial model or non-parametrically using empirical methods such as the Wilcoxon statistic (Hanley and McNeil [1982]; Bandos et al. [2017]). While the binomial model provides a smooth estimate of the curve and allows for a closed-form expression of the AUC, it may not hold in real-world data, especially when distributions are skewed or sample sizes are small. In such cases, non-parametric methods offer a more robust and flexible alternative.

A key property of the AUC is its invariance under monotone transformations of the test values, which makes it applicable across different measurement scales. However, despite being a useful metric to compare the overall performance of different tests, it does not directly provide the optimal cut-off point for clinical decisions. Its global nature may overlook relevant differences in specific clinical ranges, particularly when evaluating thresholds in regions where the balance between Se and Sp is clinically significant, or when one type of misclassification error is more important than the other (Perkins and Schisterman [2006]). In these situations, examining particular sections of the curve or focusing on clinically meaningful thresholds, such as partial AUCs, may offer more relevant insights.

Under the binomial assumption, the AUC can be calculated using the following closed-form expression:

$$\text{AUC} = \Phi\left(\frac{\mu_1 - \mu_0}{\sqrt{\sigma_1^2 + \sigma_0^2}}\right)$$

Where:

- μ_1, μ_0: means of the test results in the diseased and non-diseased populations, respectively.
- σ_1, σ_0: standard deviations in the diseased and non-diseased populations, respectively.
- Φ: the cumulative distribution function of the normal distribution $N(\mu, \sigma^2)$.

Although ROC curves and AUC provide a comprehensive summary of diagnostic accuracy, they do not directly indicate the optimal threshold needed for clinical decisions. Their role is to assess and compare the overall performance of diagnostic tests, but they do not constitute a method for selecting the optimal threshold. This limitation has led to the development of various methods specifically designed to identify optimal cut-off points.

2.2 Classical methods for cut-off estimation

The following conventional approaches rely on different criteria to define the optimal cut-off point that best separates individuals with and without the target condition. Most are based on maximizing diagnostic accuracy metrics such as Se, Sp, or combinations of both, while others introduce distance-based measures. Although widely used, these methods have several limitations. They often make assumptions about the data, depending on the distribution of test results, and may lack robustness when applied to different populations.

The most commonly used classical methods for selecting the cut-off point are presented below, together with their rationale, calculation, strengths and known limitations.

2.2.1 Youden's J statistic

The Youden index is one of the most widely used methods for determining an optimal cut-off point in the evaluation of diagnostic tests (Youden [1950]). It aims to identify the cut-off point that maximizes the sum of Se and Sp minus one, i.e:

$$J = \text{Se} + \text{Sp} - 1$$

The method provides the best balance between correctly detecting positive and negative cases. The index value ranges from 0 to 1, with higher values indicating better discriminatory perfor-

mance. This optimal point corresponds to the maximum vertical distance between the ROC curve and the random classification diagonal (Youden [1950]; Hilden and Glasziou [1996]) and is based on the simultaneous minimization of type I and type II errors.

Although Youden's index is simple and useful, it assumes that FP and FN have equal weight, without considering the prevalence of the disease or the clinical impact of classification errors. This can be problematic in contexts where one of the errors is more critical than the other, such as when avoiding FN is a priority. In addition, this index may produce unstable cut-off points in situations with flat ROC curves or small sample sizes (Nahm [2022]).

2.2.2 Euclidean distance

The Euclidean distance method selects the cut-off point that is geometrically closest to the ideal point (0, 1) in ROC space, which represents perfect classification (Se = 1, Sp = 1). The distance is calculated as:

$$\mathrm{d} = \sqrt{(1 - \mathrm{Se})^2 + (1 - \mathrm{Sp})^2}$$

The optimal threshold is the one that minimizes this distance, as it is the closest point to perfect diagnostic performance (Hajian-Tilaki [2018]). This method treats the trade-off between true and false classifications as a spatial problem, combining both Se and Sp into a single performance metric.

However, this method assumes equal importance for FP and FN, without considering disease prevalence and clinical consequences. Furthermore, it doesn't incorporate costs and may become unstable with outliers. In practice, the optimal point it identifies often coincides with that of the Youden index (Nahm [2022]); but may not be the most appropriate in clinical settings with unbalanced risks.

2.2.3 Maximum product of sensitivity and specificity

This method identifies the threshold that maximizes the product of Se and Sp, also known as the *geometric mean*, defined as:

$$\text{Product} = \max(\mathrm{Se} \cdot \mathrm{Sp})$$

The rationale is that Se and Sp must be simultaneously high to yield a high product, which is a stricter criterion than simply maximizing their sum.

It is easy to calculate and can give similar cut-offs to the Youden index, although it is more severe in terms of bias. Its main limitations include the assumption of equal importance of Se and Sp and the lack of consideration of clinical consequences or prevalence (Liu [2012]; Hajian-Tilaki [2018]).

2.2.4 Index of union (IU)

The IU (Unal [2017]) identifies the cut-off where both Se and Sp are as close as possible to the AUC. It selects the threshold that minimizes the difference between each of these metrics and the AUC, aiming for a balance that best reflects the test's global accuracy.

This is expressed by a function that minimizes the combined deviation from AUC and the absolute difference between Se and Sp.

$$\text{IU} = |\text{Se} - \text{AUC}| + |\text{Sp} - \text{AUC}|$$

The aim is to find the threshold that best reflects the overall diagnostic performance while maintaining a balance between Se and Sp. However, this method does not consider disease prevalence or unequal misclassification costs. Nevertheless, simulation studies have shown that IU can perform comparably or even better than other approaches under certain data conditions (Unal [2017]; Hassanzad and Hajian-Tilaki [2024]).

2.2.5 Cost approach

Cost-based methods aim to select the cut-off point that minimizes the total expected cost of misclassification, taking into account the prevalence of the disease and the different costs associated with FP, FN, TP and TN. A general formula for expected cost is:

$$\text{Cost} = C_{\text{FN}}(1 - \text{Se}) \cdot Pr + C_{\text{FP}}(1 - \text{Sp}) \cdot (1 - Pr) + C_{\text{TP}} \cdot \text{Se} \cdot Pr + C_{\text{TN}} \cdot \text{Sp} \cdot (1 - \text{Se})$$

where C_{FN}, C_{FP}, C_{TP} and C_{TN} represent the costs of FN, FP, TP and TN respectively, and Pr is disease prevalence.

This work uses the default cost definition, which sets the cost ratio at 5, implying that is five times more costly predicting an FN than predicting FP.

Another way to determine the optimal cutoff value is to use the cost index (fm), where maxi-

mizing fm corresponds to minimizing the average cost. Alternatively, the ***misclassification cost term*** **(MCT)** is a commonly used method that focuses only on the prevalence of the disease, C_{FP} and C_{FN}, and defines the optimal threshold as the point that minimizes the total expected cost of misclassification (Hintze [2007]):

$$\text{MCT} = \frac{C_{\text{FN}}}{C_{\text{FP}}} \cdot Pr \cdot (1 - \text{Se}) + (1 - Pr) \cdot (1 - \text{Sp})$$

This approach is clinically relevant, especially when the consequences of misclassification are different. However, all cost-based methods depend on the availability and reliability of cost estimates, which are often subjective, meaning that incorrect assumptions can lead to misleading results (López-Ratón et al. [2014]; Nahm [2022]).

2.2.6 Diagnosis Odds Ratio (DOR)

The DOR is a single measure calculated by diving the LR+ by the LR- and summarizes how effective a test is by combining Se and Sp. It is calculated as:

$$\text{DOR} = \frac{\text{LR+}}{\text{LR-}} = \frac{\text{Se} \cdot \text{Sp}}{(1 - \text{Se}) \cdot (1 - \text{Sp})}$$

By maximizing the LR^+ and minimizing the LR^+, the optimal cut-off point can be determined. The DOR reflects the probability of a positive test result in individuals with the disease compared to those without it, ranging from 0 to $+\infty$. A value of 1 means that the test cannot discriminate between diseased and non-diseased individuals, and values further from 1 indicate better test performance.

This method becomes unstable when any value in the confusion matrix is close to zero, especially when FP or FN are rare. This can lead to extremely high or low DOR values, which may not be clinically meaningful, especially when test results are not normally distributed (Glas et al. [2003]; Hajian-Tilaki [2018]; Hassanzad and Hajian-Tilaki [2024]).

Also, while the logarithm of the DOR [log(DOR)] tends to follow a normal distribution, its estimation becomes unreliable when FP or FN counts are very low, and confidence intervals based on log(DOR) may be inaccurate in such cases.

2.2.7 Min P-Value

This approach selects the threshold that produces the most statistically significant group difference, usually by using the p-value from a chi-squared or t-test. The optimal cut-off is the one with the smallest p-value.

Although this method is attractive and is widely used for exploratory analysis, it suffers from the problems of multiple testing and a high risk of type I error. It often lacks validation and can lead to overfitting unless corrections are applied (Rota and Antolini [2014]; Unal [2017]; Hilsenbeck and Clark [1996]).

2.3 Implementation of classical methods in R

Several packages have been developed to implement cut-off estimation methods in R. However, this study focuses exclusively on the `OptimalCutpoints` package, which provides a comprehensive and flexible framework for computing and comparing classical methods in continuous biomarkers.

2.3.0.1 OptimalCutpoints R package

In this work, `OptimalCutpoints` (López-Ratón et al. [2014]) is used to compute the set of classical methods for cut-off estimation presented previously. These methods serve as benchmarks for comparison with the proposed approach.

The `OptimalCutpoints` package in R calculates optimal cut-off points for diagnostic tests or continuous biomarkers to discriminate between diseased and healthy populations. It implements a set of methods that select cut-off points based on cost-benefit analysis, measures of diagnostic accuracy (such as Se, Sp, PPV, NPV and diagnostic likelihood ratios (LR+ and LR-)), and prevalence considerations. The package provides numerical and graphical results, facilitating a comprehensive assessment of diagnostic performance.

The main function `optimal.cutpoints()` allows the estimation of the method or criterion for determining the optimal threshold, the definition of the positive and negative classes, the inclusion of covariates, and the specification of parameters such as confidence intervals for the precision measures.

After calculating optimal cut-off points for each classical method, the package summarizes key performance metrics at the selected threshold, including Se, Sp, PPV, NPV, and LR diagnostics. It also allows visualization using ROC curves with optimal cut-off points.

In addition, `OptimalCutpoints` supports advanced settings such as cost-benefit adjustments, generalized indices, and flexible calculation of confidence intervals.

Example of usage:

```
library(OptimalCutpoints)
data(elas)

# Optimal cutpoint using Cost-Benefit method
optimal_cutpoint_cb <- optimal.cutpoints(
    X = "elas",
    status = "status",
    tag.healthy = 0,
    methods = "CB",
    data = elas,
    control = control.cutpoints(CFP = 1, CFN = 5),
    ci.fit = TRUE,
    conf.level = 0.95
)

# X: name of the diagnostic test variable.
# status: variable indicating disease status (0 = healthy, 1 = diseased).
# tag.healthy: value that represents healthy individuals in 'status'.
# methods: method used: "CB" = Cost-Benefit method.
# data: dataset containing all variables.
# control: cost ratio, Cost of False Positive = 1, False Negative = 5.
# ci.fit: compute confidence.intervals for accuracy metrics.
# conf.level: confidence level for the intervals (95%).
```

Chapter 3

Methods and Materials

This study proposes an alternative methodology developed and validated by data simulation. This approach is based on a log-binomial model that incorporates bootstrapping resampling to improve the robustness, stability and reliability of cut-off point estimation. The plausibility of the model is prioritized using AIC over p-values, and two estimates of the optimal cut-off point are selected: the expected value and the mode of the distribution generated by the bootstrap process.

3.1 Data simulation

To evaluate the performance of the proposed method, a simulation study was performed to reflect various clinical scenarios. A continuous biological marker was simulated under the assumption of a normal distributions $D \sim \mathcal{N}(\mu, 1)$ in two groups: non-diseased individuals ($D = 0$) and diseased individuals ($D = 1$). The prevalence of the event (diseased group) was set at 10%, reflecting a low frequency. These simulation conditions are consistent with previous literature on biological marker behavior under both symmetric and asymmetric conditions (Pugachev [1984]; Rota and Antolini [2014]).

Three scenarios were considered based on the effect size of the biological marker:

- **Null scenario (N = 2000):** no difference between groups, with shared mean $\mu_0 = \mu_1 = 0$.
- **Small effect scenario (N = 1000):** the diseased group has a slight mean shift $\mu_1 = 0.5$.
- **Moderate effect scenario (N = 500):** the diseased group presents a more evident shift $\mu_1 = 1$.

The number of observations for each scenario was adjusted to ensure sufficient statistical power. A power analysis was performed using the *Power and Sample Size* software (PS [2025]), and the final sample sizes were chosen to balance estimation precision and computational feasibility for a fixed prevalence rate of 10%.

For each scenario, 50 independent simulations were generated by varying the seed in each iteration using `set.seed`. This procedure allows a systematic assessment of the stability and variability of the estimated cut-off points. For each simulated dataset, diagnostic performance metrics were computed for the best cut-off, and their corresponding confidence intervals were estimated.

All simulations were implemented in R (version 4.4.0), allowing control over data generation, reproducibility and comparison of methods.

3.2 Proposed approach for cut-off determination

To improve the robustness, clinical interpretability and stability of the cut-off point for continuous biomarkers, we propose a methodology based on preliminary density and ROC curve analyses, log-binomial modeling, AIC-based selection and bootstrap resampling. *Figure 3.1* outlines the workflow of the proposed methodology for cut-off point estimation.

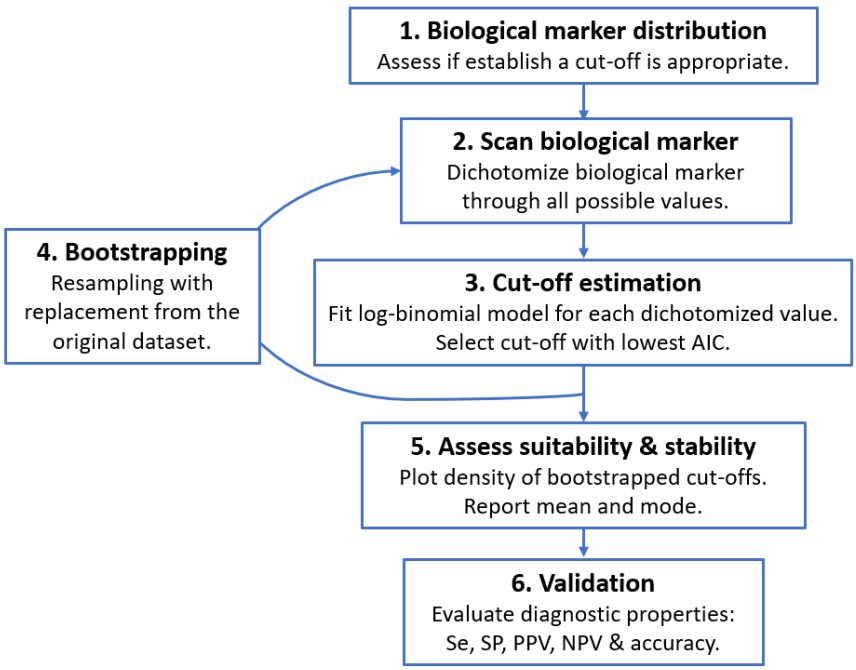


Figure 3.1: Workflow of the proposed method for cut-off point estimation.

3.2.1 Exploratory analysis: density distribution and ROC curve

To assess whether it is appropriate to set a meaningful cut-off point in the data, an exploratory graphical analysis is performed for each of the simulated scenarios. This includes:

- Density plots: biological marker density distribution by group (diseased vs. non-diseased) to visually assess overlap and separation.
- ROC curve and AUC: to quantify discriminatory performance.

This visual exploration provides an intuitive understanding of the potential of the marker to discriminate individuals. *Figure 3.2* displays the density distribution plots and the corresponding ROC curves across the simulated scenarios.

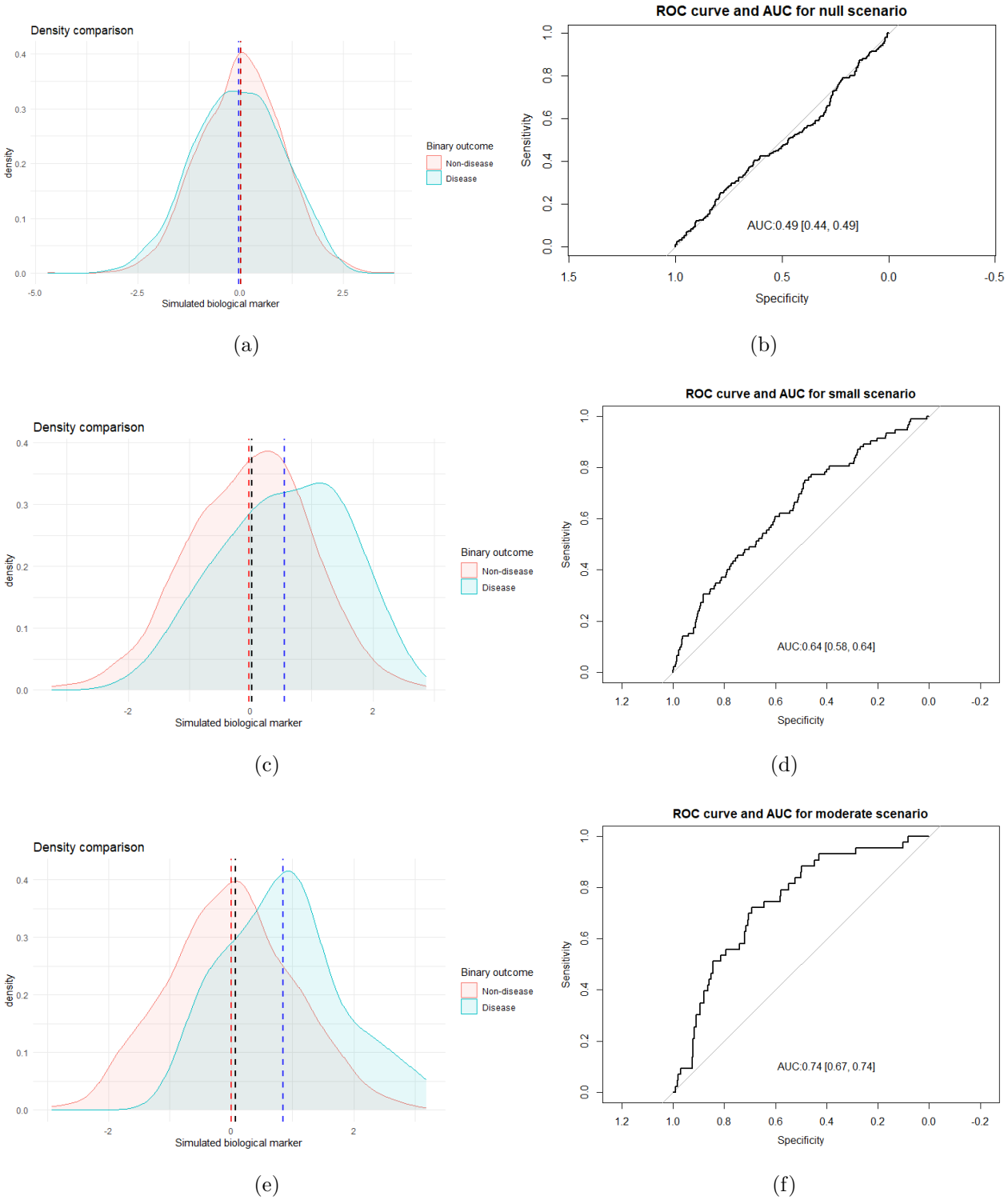


Figure 3.2: Exploratory analysis for each scenario.
(a, c, e) Density plot of the continuous biological marker for the disease and non-disease groups in null, small and moderate scenarios respectively. (b, d, f) ROC curve and AUC summarizing the overall discriminative ability in null, small and moderate scenarios, respectively. The black line indicates the mean of the simulated biological marker; the blue line represents the mean of the biological marker in the disease group; and the red line the mean of the biological marker in the non-disease group.

3.2.2 Bootstrap-based estimation with log-binomial modeling

To improve the robustness and generalizability of the cut-off point estimation, we applied a bootstrap-based estimation strategy combined with log-binomial modeling and AIC-based model selection.

A log-binomial model is a generalized linear model for binary outcomes that uses the logarithm as the link function and assumes binomial distribution. It estimates the relative risk (RR) or risk ratio, which quantifies how the probability of the outcome changes between groups (in our case, the groups defined by the candidate cut-off point) (Zhang and Yu [1998]), thereby making the results more clinically interpretable (Zhou et al. [2011]).

For each simulated dataset, the following steps were performed:

1. A range of candidate cut-off values were scanned across the distribution of the biomarker.

2. For each candidate cut-off c, the biological marker was dichotomized as $I(T > c)$, ensuring that the extreme tails of the distribution retained a sufficient number of events to avoid unstable or biased estimates.

3. A log-binomial regression model was fitted for each candidate cut-off, with the following form:
$$\log(P(D = 1)|\mathbb{I}(T \geq \text{cut-off})) = \beta_0 + \beta_1 \cdot \mathbb{I}(T \geq \text{cut-off})$$

4. For each model, the AIC was computed and the cut-off point with the lowest value was selected as the optimal threshold for that bootstrap sample, prioritizing model plausibility over traditional significance-based approaches.

To capture uncertainty and ensure robustness, this procedure was embedded in a bootstrap resampling framework. Bootstrapping is a non-parametric technique that estimates the distribution of a statistic by repeated sampling with replacement from the original dataset (Efron and Tibshirani [1994]). In this context, it allows to quantify the variability of threshold estimates and to construct confidence intervals (Habibzadeh et al. [2016]). In this study, 75 bootstrap samples were generated for each simulated dataset. Ideally, a larger number of bootstrap samples would be preferable, but the computational time and resource demands were too high for a personal computer. For each iteration, the optimal cut-off was estimated according to the above procedure, resulting in a distribution of optimal cut-offs rather than a single one.

3.2.3 Estimation of optimal cut-off: Mean and Mode

Two summary statistics are extracted from the bootstrap distribution of optimal cut-off points obtained for each simulated dataset to represent the optimal threshold derived from the proposed methodology:

- **Mean (Expected value):** the arithmetic mean of the 75 bootstrapped cut-off points. This value offers a robust central estimate by averaging across all replicates.
- **Mode (maximum density point):** corresponds to the most frequently selected cut-off, identified as the value with the highest peak in a kernel density estimation (KDE) of the bootstrap distribution. This represents the most probable or recurrent threshold.

Figure 3.3 shows an example of a density function obtained after running a simulation, illustrating how the expected value and mode are derived from the bootstrap distribution. This visual representation highlights the central tendency (mean) and the most frequent estimate (mode), as well as other local maxima of high density.

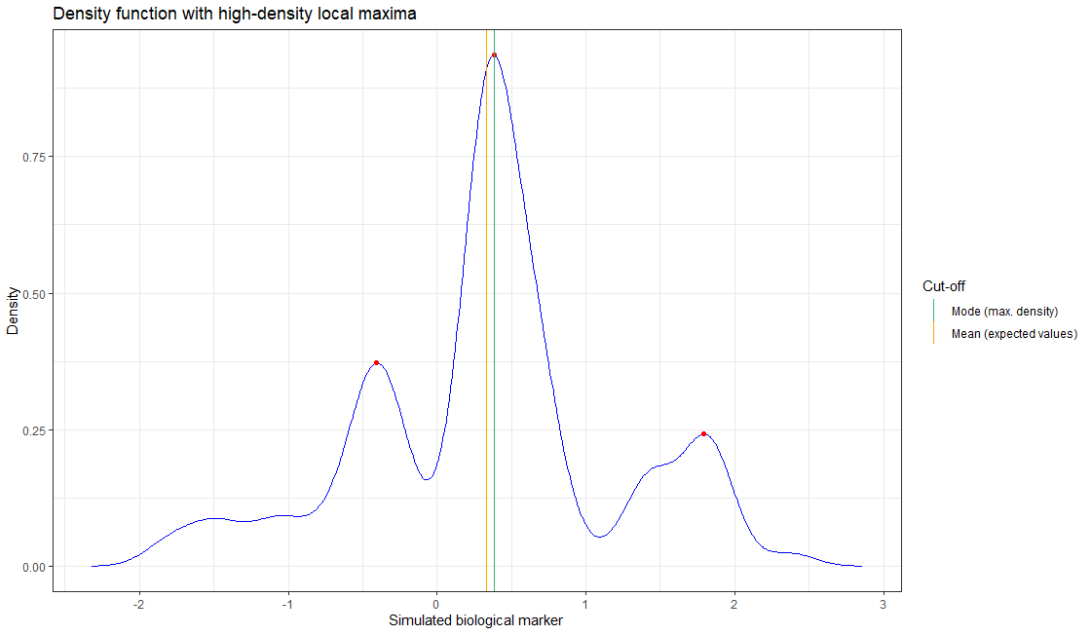


Figure 3.3: Density function of the bootstrapped optimal cut-off points.
The orange line indicates the mean (expected value), and the green line represents the mode (maximum density point). Red dots indicate local density maxima.

These two results are representative of the proposed method and will be used for further comparison with classical methods. The use of both the expected value and the mode allows a more complete characterization of the bootstrap distribution, reflecting the central tendency and the most frequent estimate respectively.

In this way, the proposed approach avoids reliance on a single point estimate from the original sample and incorporates model variability and plausibility through the combined use of bootstrap, log-binomial modeling and AIC.

3.3 Comparison and evaluation of diagnostic methods

The classical methods for optimal cut-off estimation presented earlier in the State of the Art section were applied to each simulated dataset to compare their performance with the proposed method. All these approaches were implemented in R: the Youden index, maximum product of Se and Sp, cost approach, misclassification cost, DOR, and minimum p-value methods were computed using the `OptimalCutpoints` package (López-Ratón et al. [2014]). In contrast, the IU and Euclidean distance methods were implemented manually using base R code.

For each simulation replica and scenario, the optimal cut-off determined by each method was stored along with all the simulated datasets. After all simulations were completed, the stored cut-offs for each classical method were averaged and then reapplied to the corresponding datasets to classify individuals as diseased or non-diseased based on whether their biological marker value exceeded the selected threshold. This allowed contingency tables to be constructed and standard diagnostic performance metrics to be calculated, including: Se, Sp, PPV, NPV and overall accuracy, allowing robust summary measures of the performance of each method to be estimated for each dataset.

As these metrics were estimated across all simulation replicates ($n = 50$ per scenario), 95% confidence intervals for the cut-off values and diagnostic metrics were computed using empirical percentiles derived from the simulation distributions. With these confidence intervals, we aimed to reflect the variability and uncertainty of each method's performance.

Comparative tables were generated, stratified by simulation scenario (null, small, and moderate effect), summarizing the mean performance and confidence intervals for each method. Additionally, forest plots were produced to visually represent both the cut-off estimates and for diagnostic performance metrics, facilitating a clearer assessment of precision, variability, and consistency across methods.

This robust evaluation framework enables a systematic and reproducible comparison between classical methods and the proposed approach, capturing both point estimates and the stability of each method under different data conditions. An example of the R code used for each method is provided in *Appendix A*, where the methodological procedures were standardized to ensure comparability across simulations.

3.4 Materials and software

All statistical analyses and simulations were performed using R version 4.4.0 (2024-06-14) (R Core Team [2024]) for Windows. The following R packages were used throughout the project:

- **Data manipulation and processing:** `dplyr`, `tidyr`, `tibble`, `purrr`, `stringr`
- **Summary tables and formatting:** `gtsummary`, `gt`, `kableExtra`
- **ROC and cut-off analysis:** `OptimalCutpoints`, `ThresholdROC`, `pROC`, `DescTools`
- **Diagnostic test evaluation:** `epiR`, `caret`
- **Graphics and visualization:** `ggplot2`

All results were obtained from Quarto reports rendered in RStudio (Posit team [2025]) to ensure reproducibility and structured documentation of each simulation scenario.

Chapter 4

Results

The comparative results of the classical cut-off point estimation methods and the new proposed approach, applied to the three simulated scenarios: null, small and moderate, are presented below.

For each scenario, the performance of the estimated cut-off points and their diagnostic ability are evaluated. Tables summarizing the mean diagnostic metrics obtained by the simulations for each method are included, as well as forest plots illustrating the distribution of the estimated cut-off points with their corresponding confidence intervals. This presentation allows a visual comparison of the variability, stability and consistency of each method.

4.1 Null Scenario

In the null scenario, where there are no real differences between disease and non-disease groups, the new proposed method with its two estimation approaches (mean and mode of bootstrapped cut-off points distribution) showed a similar behavior to the minimum p-value method.

The estimated cut-off points were close to the center of the distribution: –0.07 for the expected value, –0.12 for the maximum density, and –0.15 for the minimum p-value. Accuracy values were comparable across the three methods (ranging from 0.45 to 0.48).

Most classical methods returned thresholds near the center the distribution: the Youden index (cut-off = 0.23; accuracy = 0.57), Euclidean distance (cut-off = 0.03; accuracy = 0.51), maximum product of Se and Sp (cut-off = 0.02, accuracy = 0.51), and IU (cut-off = -0.00; accuracy = 0.50). These methods report relatively balanced Se and Sp values, but overall performance remained modest.

On the other hand, the cost-based methods and DOR produced thresholds in more extreme cut-off values: 3.49 for the cost approach, 3.40 for the misclassification cost method, and 2.59 for the DOR. These thresholds were associated with very low Se values (0.00 to 0.01) and high Sp (approximately 1.00). In all three methods, overall accuracy was around 0.90. *Table 4.1 and figure 4.1* summarize the diagnostic performance metrics and the estimated cut-off points with their 95% confidence intervals for each method under the null scenario.

Comparison of Cut-off Methods: Null scenario

Method	Cut-off point [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	PPV [95% CI]	NPV [95% CI]	Accuracy [95% CI]
New approach: Expected value	-0.07 [-1.02; 0.72]	0.53 [0.47; 0.60]	0.47 [0.45; 0.49]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.47 [0.46; 0.49]
New approach: Maximum density	-0.12 [-1.99; 1.86]	0.54 [0.49; 0.61]	0.45 [0.43; 0.47]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.46 [0.44; 0.48]
Min p-value	-0.15 [-2.63; 2.49]	0.56 [0.50; 0.62]	0.44 [0.42; 0.45]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.45 [0.43; 0.47]
Youden's J statistic	0.23 [-1.80; 1.66]	0.41 [0.35; 0.49]	0.59 [0.57; 0.61]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.57 [0.56; 0.59]
Euclidean distance	0.03 [-0.23; 0.36]	0.48 [0.42; 0.55]	0.51 [0.50; 0.53]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.51 [0.49; 0.53]
Max. product Se*Sp	0.02 [-0.24; 0.36]	0.49 [0.43; 0.55]	0.51 [0.49; 0.53]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.51 [0.49; 0.52]
Index of Union	-0.00 [-0.09; 0.09]	0.50 [0.44; 0.57]	0.50 [0.48; 0.52]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.50 [0.48; 0.51]
Cost approach	3.49 [2.83; 4.22]	0.00 [0.00; 0.00]	1.00 [1.00; 1.00]	0.02 [0.00; 0.19]	0.90 [0.89; 0.91]	0.90 [0.89; 0.91]
Misclassification cost	3.40 [2.79; 4.22]	0.00 [0.00; 0.00]	1.00 [1.00; 1.00]	0.08 [0.00; 0.60]	0.90 [0.89; 0.91]	0.90 [0.89; 0.91]
log(DOR)	2.59 [-2.16; 4.21]	0.00 [0.00; 0.01]	0.99 [0.99; 1.00]	0.09 [0.00; 0.27]	0.90 [0.89; 0.91]	0.90 [0.88; 0.91]

Table 4.1: Diagnostic performance metrics for each method under the null scenario.

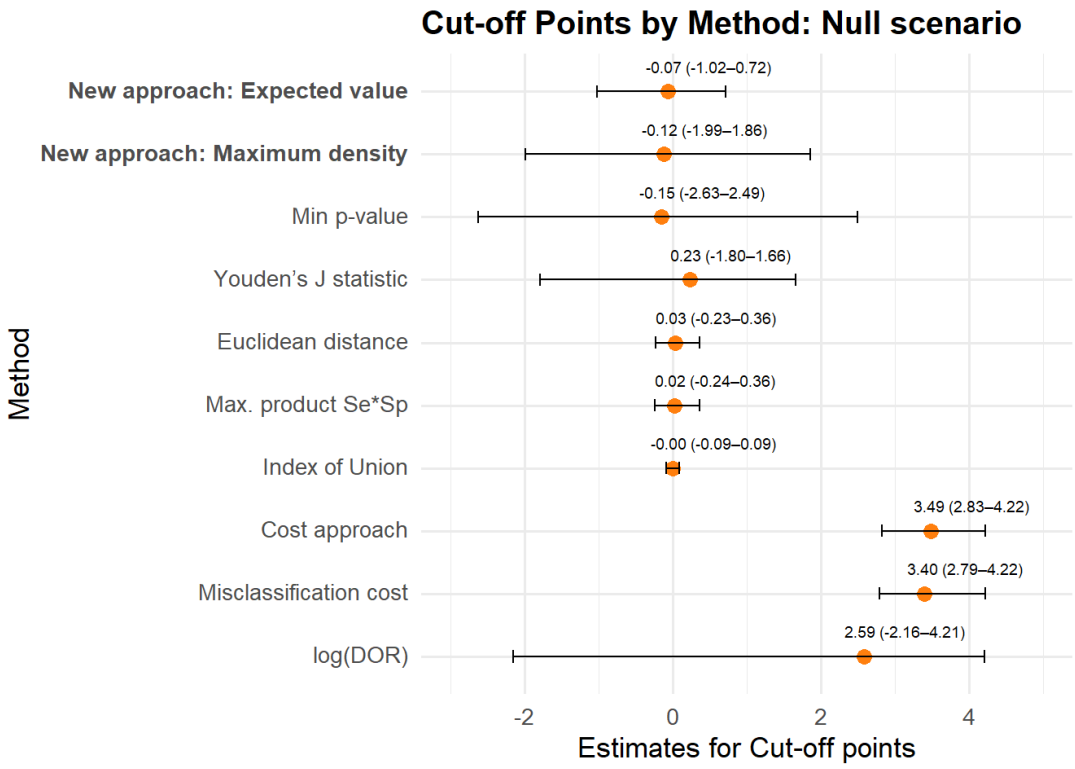


Figure 4.1: Estimated cut-off points and 95% CI by method: Null scenario.

4.2 Small Scenario

In the small scenario, where the disease group has a modest mean shift ($\mu = 0.5$), the two variants of the proposed method based on the expected value (cut-off = 0.88) and the maximum density (cut-off = 0.93) of the bootstrapped cut-off distribution, yield similar results to the minimum p-value method (cut-off = 0.89). Accuracy values for these methods ranged from 0.77 to 0.78, with Sp values between 0.81 and 0.82 and Se between 0.35 and 0.37.

The classical methods (Youden index, Euclidean distance, maximum product of Se and Sp, and IU), returned lower cut-offs (ranging from 0.29 to 0.34) with Se values ranging from 0.57 to 0.59 and Sp values between 0.61 and 0.63. Overall accuracy for these methods was slightly lower (between 0.61 and 0.63).

Once again, the cost-based approaches and DOR produced higher cut-offs: 3.21 (cost), 2.98 (misclassification cost) and 1.54 (DOR). Se was low (0.01 for cost and misclassification cost, and 0.15 for DOR), while Sp remained high (1.00 for cost-based methods and 0.94 for DOR). The accuracy was 0.90 for the cost-based methods and 0.86 for DOR. *Table 4.2 and figure 4.2* summarize the diagnostic performance metrics and the estimated cut-off points with their 95%

confidence intervals for each method under the small scenario.

Comparison of Cut-off Methods: Small scenario

Method	Cut-off point [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	PPV [95% CI]	NPV [95% CI]	Accuracy [95% CI]
New approach: Expected value	0.88 [0.10; 1.57]	0.37 [0.28; 0.47]	0.81 [0.80; 0.84]	0.18 [0.13; 0.23]	0.92 [0.90; 0.94]	0.77 [0.74; 0.79]
New approach: Maximum density	0.93 [-0.28; 2.10]	0.35 [0.27; 0.46]	0.82 [0.80; 0.85]	0.18 [0.14; 0.23]	0.92 [0.90; 0.94]	0.78 [0.75; 0.80]
Min p-value	0.89 [-0.34; 2.17]	0.37 [0.28; 0.47]	0.81 [0.80; 0.84]	0.18 [0.13; 0.24]	0.92 [0.90; 0.94]	0.77 [0.75; 0.80]
Youden's J statistic	0.32 [-0.48; 0.90]	0.58 [0.48; 0.66]	0.62 [0.59; 0.64]	0.14 [0.11; 0.17]	0.93 [0.91; 0.94]	0.62 [0.60; 0.64]
Euclidean distance	0.32 [0.04; 0.61]	0.58 [0.48; 0.66]	0.62 [0.60; 0.64]	0.14 [0.11; 0.17]	0.93 [0.91; 0.94]	0.62 [0.60; 0.64]
Max. product Se*Sp	0.34 [0.04; 0.71]	0.57 [0.48; 0.65]	0.63 [0.61; 0.65]	0.14 [0.11; 0.17]	0.93 [0.91; 0.95]	0.62 [0.61; 0.64]
Index of Union	0.29 [0.10; 0.49]	0.59 [0.50; 0.68]	0.61 [0.58; 0.63]	0.14 [0.11; 0.17]	0.93 [0.91; 0.95]	0.61 [0.59; 0.63]
Cost approach	3.21 [2.75; 3.88]	0.01 [0.00; 0.02]	1.00 [1.00; 1.00]	0.45 [0.00; 1.00]	0.90 [0.88; 0.92]	0.90 [0.88; 0.92]
Misclassification cost	2.98 [2.46; 3.70]	0.01 [0.00; 0.02]	1.00 [1.00; 1.00]	0.37 [0.00; 1.00]	0.90 [0.88; 0.92]	0.90 [0.88; 0.92]
log(DOR)	1.54 [-1.72; 3.70]	0.15 [0.09; 0.21]	0.94 [0.92; 0.95]	0.22 [0.13; 0.30]	0.91 [0.89; 0.93]	0.86 [0.83; 0.88]

Table 4.2: Diagnostic performance metrics for each method under the small scenario.

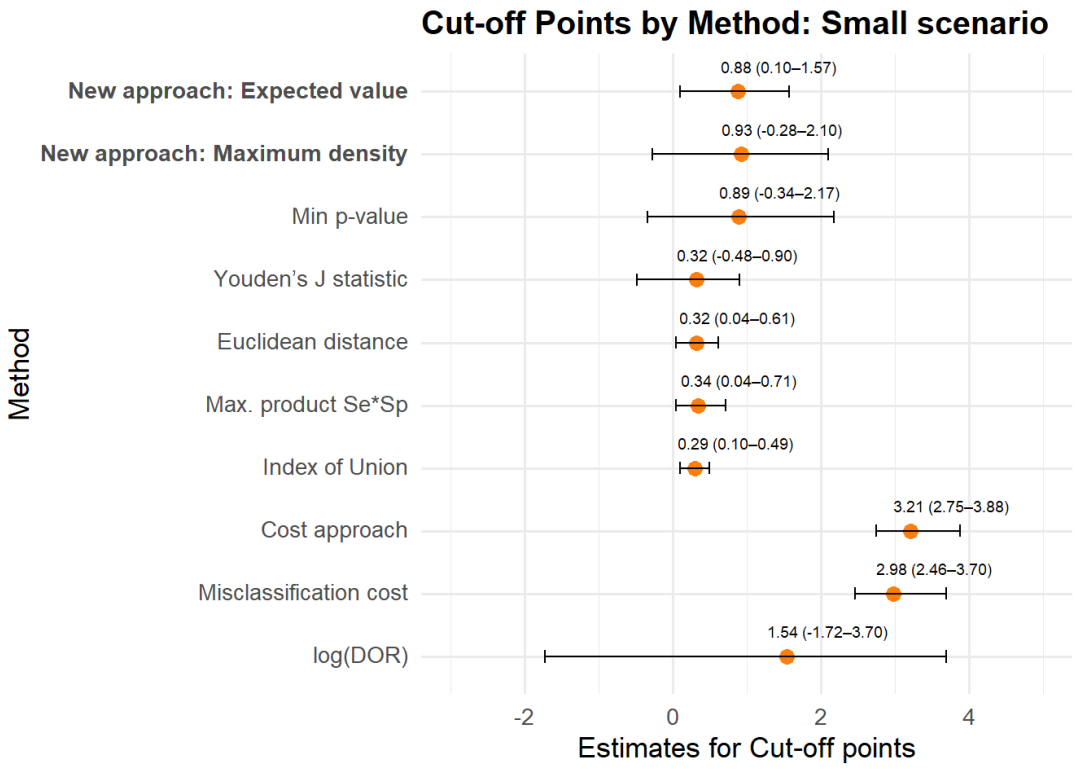


Figure 4.2: Estimated cut-off points and 95% CI by method: Small scenario.

4.3 Moderate Scenario

In the moderate scenario, which simulates a more substantial separation between the disease and non-disease groups ($\mu = 1$), the proposed method produced higher accuracy values that in previous scenarios. The expected value approach yielded a cut-off of 1.14 and an accuracy of 0.83, while the maximum density approach estimated a cut-off of 1.07 with an accuracy of 0.82. However, Se values were between 0.44 and 0.47, and Sp were between 0.86 and 0.88.

The minimum p-value method gave a cut-off of 0.83, with higher Se (0.58), lower Sp (0.80) and accuracy of 0.78.

Classical methods such as the Youden index, Euclidean distance, maximum product of Se and Sp, and IU, produced similar thresholds, ranging from 0.55 to 0.57. Their diagnostic metrics were also comparable, with accuracy around 0.71, and balanced Se/Sp values (ranging from 0.68 to 0.71, respectively).

The cost-based methods continued to select higher thresholds: 2.99 for the cost approach, and 2.61 for the misclassification cost. These methods reported high Sp (up to 1.00), low Se (ranging from 0.03 to 0.6), and accuracy values around 0.90.

The DOR method, with its slightly lower threshold of 2.08, yielded a higher Se (0.14) and a PPV of 0.45, while maintaining high Sp (0.98) and an overall accuracy of 0.90. *Table 4.3 and figure 4.3* summarize the diagnostic performance metrics and the estimated cut-off points with their 95% confidence intervals for each method under the moderate scenario.

Comparison of Cut-off Methods: Moderate scenario

Method	Cut-off point [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	PPV [95% CI]	NPV [95% CI]	Accuracy [95% CI]
New approach: Expected value	1.14 [0.67; 1.57]	0.43 [0.30; 0.58]	0.87 [0.84; 0.91]	0.28 [0.19; 0.40]	0.93 [0.91; 0.96]	0.83 [0.80; 0.86]
New approach: Maximum density	1.07 [0.39; 1.95]	0.46 [0.33; 0.62]	0.86 [0.83; 0.89]	0.27 [0.18; 0.38]	0.94 [0.91; 0.96]	0.82 [0.78; 0.85]
Min p-value	0.83 [0.18; 2.00]	0.57 [0.44; 0.70]	0.80 [0.77; 0.83]	0.24 [0.19; 0.32]	0.94 [0.91; 0.96]	0.77 [0.74; 0.81]
Youden's J statistic	0.56 [0.16; 1.01]	0.68 [0.55; 0.82]	0.71 [0.68; 0.75]	0.21 [0.17; 0.28]	0.95 [0.92; 0.97]	0.71 [0.68; 0.74]
Euclidean distance	0.55 [0.27; 0.84]	0.69 [0.57; 0.82]	0.71 [0.67; 0.74]	0.21 [0.17; 0.28]	0.95 [0.92; 0.97]	0.71 [0.68; 0.74]
Max. product Se*Sp	0.57 [0.18; 1.00]	0.68 [0.55; 0.82]	0.72 [0.68; 0.75]	0.21 [0.17; 0.28]	0.95 [0.92; 0.97]	0.71 [0.68; 0.74]
Index of Union	0.56 [0.31; 0.80]	0.68 [0.55; 0.82]	0.71 [0.68; 0.75]	0.21 [0.17; 0.28]	0.95 [0.92; 0.97]	0.71 [0.68; 0.74]
Cost approach	2.99 [2.48; 3.71]	0.03 [0.00; 0.07]	1.00 [0.99; 1.00]	0.70 [0.00; 1.00]	0.90 [0.87; 0.92]	0.90 [0.87; 0.92]
Misclassification cost	2.61 [1.95; 3.46]	0.06 [0.00; 0.12]	1.00 [0.99; 1.00]	0.59 [0.00; 1.00]	0.90 [0.87; 0.92]	0.90 [0.87; 0.92]
log(DOR)	2.08 [-0.90; 3.46]	0.14 [0.04; 0.24]	0.98 [0.97; 0.99]	0.45 [0.20; 0.67]	0.91 [0.88; 0.93]	0.90 [0.87; 0.92]

Table 4.3: Diagnostic performance metrics for each method under the moderate scenario.

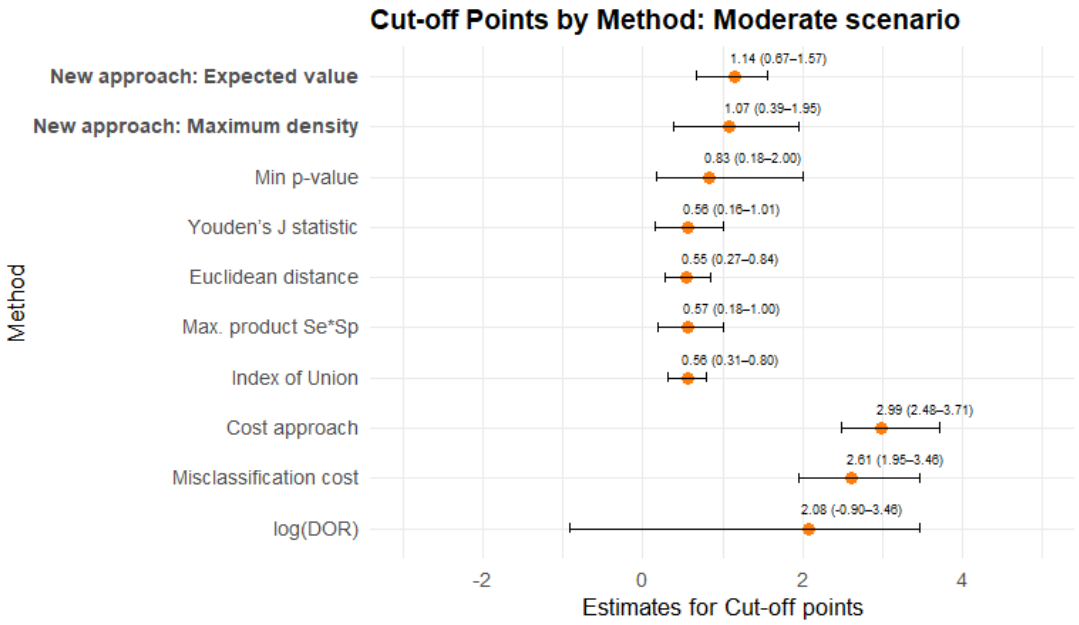


Figure 4.3: Estimated cut-off points and 95% CI by method: Moderate scenario.

Chapter 5

Discussion

In medicine, the identification of biological markers is a key objective, particularly for assessing disease risk or aiding in diagnosis. Defining an optimal cut-off point can improve clinical decision-making; however, this process is often complex, and commonly used methods aim to balance Se and Sp to determine this threshold.

In contrast to classical methods, the proposed approach allows for the incorporation of data variability, which provides greater stability and a clearer interpretation of the selected cut-off points. This advantage is particularly evident when observing the behavior of the different methods in the three scenarios (null, small and moderate), suggesting that a method's usefulness depends as much on its adaptability to the data context as on its theoretical foundation (Hassanzad and Hajian-Tilaki [2024]; Duarte [2021]). The new method does not force artificial separations when there is no real effect and it reacts progressively as the effect becomes more evident.

In the null scenario, where there is no real difference between the groups, both the new approach and the method based on the minimum p-value identified cut-off points close to the center of the distribution, resulting in low diagnostic accuracy, consistent with the lack of signal. In contrast, other methods, such as those based on cost or DOR, selected very extreme cut-offs. While this choice artificially increased the overall accuracy of the test, it was mainly due to the predominance of true negatives in a low-prevalence setting. This result highlights the risk of relying solely on global metrics such as accuracy without considering disease distribution and clinical context (Habibzadeh et al. [2016]; Polley and Dignam [2021]).

When a small difference between groups was introduced (small effect scenario), the new method maintained a conservative behavior, estimating cut-off points slightly shifted towards the tail of the non-diseased group. This choice allowed improved Sp at the cost of reduced Se, which

may be desirable in clinical situations where it is preferable to avoid false positives (Coste and Pouchot [2003]). In contrast, other methods trying to balance Se and Sp tended to choose lower thresholds, increasing Se but reducing overall accuracy. The stability of the new approach in this intermediate scenario suggests that it may be useful, particularly in clinical settings where there is uncertainty or low discriminative power (Hilden and Glasziou [1996]).

In the moderate effect scenario, all methods improved their performance, but the proposed approach continued to show advantages. While the classical methods achieved an acceptable balance between Se and Sp, their overall accuracy remained around 70%. In contrast, the new method achieved an accuracy of over 80%, along with narrower confidence intervals, which reinforces its reliability (Efron and Tibshirani [1994]; Royston and Sauerbrei [2003]).

The use of bootstrapping in this sense not only improves the robustness of the procedure by reducing the dependence on a single sample but also enables a more critical assessment of the biological marker's distribution. This helps determine whether dichotomization is appropriate and, if so, which threshold is most representative, potentially leading to better clinical decisions (Irwin and McClelland [2003]; Perkins and Schisterman [2006]).

This approach also promotes a shift in the way cut-off estimation is understood. Instead of relying on rigid, predefined rules, it encourages a flexible and realistic strategy that considers the uncertainty and complexity of real clinical data (Vickers [2008]; Zhou et al. [2011]). By doing so, it supports decisions that are not only statistically sound but also better aligned with clinical needs (Pepe [2003]; Chang et al. [2017]).

Nonetheless, this flexibility comes with some limitations. The method is computationally intensive, as it requires multiple iterations and repeated model fitting. This may hinder its application in settings with limited computational capacity or when dealing with large datasets. In this study, all simulations and analyses were conducted on a standard laptop, which restricted both the number of bootstrap iterations and the complexity of the scenarios explored. At the study level, the simulations were based on simplified assumptions, including fixed prevalence and normally distributed biological markers, which may not fully reflect real-world variability. Additionally, the cut-off was estimated using a log-binomial model, though future studies could explore alternatives, such as logistic regression to improve adaptability. A default cost ratio of 5 was used due to the lack of clinical input, but ideally, cost definitions should be guided by expert knowledge to reflect real consequences.

For these reasons, future research should apply this method to real clinical datasets, particularly those involving class imbalance, missing values, or asymmetric distributions. Increasing the number of iterations and implementing the method in high-performance computing or cloud-

based environments would also be valuable, as this would improve scalability. Expanding the simulation framework to include a wider range of prevalence and distributional shapes would allow the performance of the method to be tested under more statistical scenarios. Furthermore, future studies could explore integrating cost or clinical utility measures directly into the cut-off estimation process.

Overall, this work proposes a robust and flexible strategy for selecting cut-off points in continuous biological markers. It addresses several strengths and limitations of classical methods, and provides a more realistic alternative based on data variability, result stability, and clinical relevance. Rather than aiming for a single optimal solution, it emphasizes the importance of understanding how the data behaves before making a decision.

Chapter 6

Conclusions

This study presents a novel, data-driven approach to selecting optimal cut-off points for continuous biological markers. This approach combines log-binomial modeling, bootstrap resampling and AIC-based selection. The method produced stable and consistent results across all simulated scenarios. Notably, in the null scenario, it correctly identified central values with limited discriminative power, reflecting the absence of a true effect. In small and moderate scenarios, it provided reliable thresholds and achieved greater overall accuracy than classical methods, particularly in situations with a weak signal, where other approaches tended to overfit or produce extreme values.

A key strength of this approach is its ability to incorporate sampling variability through bootstrapping, which improves the robustness and interpretability of the results. Using both the mean and the mode of the bootstrapped cut-off distribution provides greater flexibility, enabling the method to adapt to different clinical and statistical contexts. This contributes to more transparent and informed decision-making, particularly when dealing with diagnostic uncertainty or class imbalance.

Overall, the proposed method addresses the key limitations of classical methods by emphasizing model stability, adaptability, and clinical relevance. It shifts the concept of dichotomization from a static threshold to a dynamic process based on data variability and contextual understanding. Future research should focus on applying this method to real-world clinical datasets to further evaluate its practical impact and integration into diagnostic protocols.

Bibliography

Frank E. Harrell Jr. *Biostatistics for Biomedical Research*. Vanderbilt University, 2025. URL https://hbiostat.org/bbr/info.

C.R. Weinberg. How bad is categorization? *Epidemiology*, 6:345–347, 1995. URL https://www.jstor.org/stable/3702077.

C. Chang, M.K. Hsieh, W.Y. Chang, A.J. Chiang, and J. Chen. Determining the optimal number and location of cutoff points with application to data of cervical cancer. *PLoS One*, 12:e0176231, 2017. URL https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0176231&type=printable/.

Mei-Yin C. Polley and James J. Dignam. Statistical considerations in the evaluation of continuous biomarkers. *Journal of Nuclear Medicine*, 62(5):605–611, 2021. ISSN 0161-5505. doi: 10.2967/jnumed.120.251520. URL https://jnm.snmjournals.org/content/62/5/605.

Paulo Schiavom Duarte. Cut point identification of continuous biomarkers: A challenge that goes beyond statistical aspects. *Journal of Nuclear Medicine*, 62(12):1833, 2021. doi: 10.2967/jnumed.121.262905. URL https://jnm.snmjournals.org/content/62/12/1833.

A. J. Vickers, E. Basch, and M. W. Kattan. Against diagnosis. *Annals of Internal Medicine*, 149 (3):200–203, 2008. doi: 10.7326/0003-4819-149-3-200808050-00010. URL https://pubmed.ncbi.nlm.nih.gov/18678847/.

European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504. Accessed: 2025-05-23.

PS. *Power and Sample Size Program*. Department of Biostatistics, Vanderbilt University, Vanderbilt, EEUU, 2025. URL https://cqsclinical.app.vumc.org/ps/.

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974. doi: 10.1109/TAC.1974.1100705.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL https://www.r-project.org/.

M. Hassanzad and K. Hajian-Tilaki. Methods of determining optimal cut-point of diagnostic biomarkers with application of clinical data in roc analysis: an update review. *BMC Medical Research Methodology*, 24:84, 2024. doi: 10.1186/s12874-024-02198-2. URL https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-024-02198-2.

J.R. Irwin and G.H. McClelland. Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, 40:366–371, 2003. URL https://www.researchgate.net/publication/228162485_Negative_Consequences_of_Dichotomizing_Continuous_Predictor_Variables/.

C. E. Metz. Basic principles of roc analysis. *Seminars in Nuclear Medicine*, 8(4):283–298, 1978. doi: 10.1016/s0001-2998(78)80014-2. URL https://doi.org/10.1016/s0001-2998(78)80014-2.

M. H. Zweig and G. Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4):561–577, 1993. URL https://pubmed.ncbi.nlm.nih.gov/8472349/. Erratum in: Clin Chem 1993 Aug;39(8):1589.

Margaret Sullivan Pepe. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, 2003. doi: 10.1093/oso/9780198509844.001.0001. URL https://doi.org/10.1093/oso/9780198509844.001.0001. Online edn, Oxford Academic, 31 Oct. 2023, accessed 11 Mar. 2025.

J.A. Hanley and B.J. McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143:29–36, 1982. doi: 10.1148/radiology.143.1.7063747. URL https://pubmed.ncbi.nlm.nih.gov/7063747/.

A. I. Bandos, B. Guo, and D. Gur. Estimating the area under roc curve when the fitted binormal curves demonstrate improper shape. *Academic Radiology*, 24(2):209–219, 2017. doi: 10.1016/j.acra.2016.09.020. URL https://doi.org/10.1016/j.acra.2016.09.020.

N.J. Perkins and E.F. Schisterman. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American Journal of Epidemiology*, 163(7):670–675, 2006. doi: 10.1093/aje/kwj063. URL https://pubmed.ncbi.nlm.nih.gov/16410346/.

W.J. Youden. Index for rating diagnostic tests. *Cancer*, 3:32–35, 1950. URL https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/1097-0142(1950)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3.

J. Hilden and P. Glasziou. Regret graphs, diagnostic uncertainty and youden's index. *Statistics in Medicine*, 15:969–986, 1996. URL https://pubmed.ncbi.nlm.nih.gov/8783436/.

Francis Sahngun Nahm. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1):25–36, 2022. doi: 10.4097/kja.21209. URL https://pubmed.ncbi.nlm.nih.gov/35124947/. Erratum in: Korean J Anesthesiol. 2022 Apr;75(2):167.

K. Hajian-Tilaki. The choice of methods in determining the optimal cut-off value for quantitative diagnostic test evaluation. *Statistical methods in medical research*, 27(8):2374–2383, 2018. doi: 10.1177/0962280216680383. URL https://doi.org/10.1177/0962280216680383.

X. Liu. Classification accuracy and cut point selection. *Statistics in Medicine*, 31(23):2676–2686, 2012. doi: 10.1002/sim.4509.

Iker Unal. Defining an optimal cut-point value in roc analysis: An alternative approach. *Computational and mathematical methods in medicine*, Epub3762651, 2017. doi: 10.1155/2017/3762651. URL https://doi.org/10.1155/2017/3762651.

Daniel J. Hintze. *NCSS Documentation: One ROC Curve and Cutoff Analysis - Chapter 546*. NCSS, Kaysville, 2007.

M. López-Ratón, M. X. Rodríguez-Álvarez, C. C. Suárez, and F. G. Sampedro. Optimalcutpoints: An r package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software*, 61(8):1–36, 2014. doi: 10.18637/jss.v061.i08. URL https://doi.org/10.18637/jss.v061.i08.

Afina S. Glas, Jeroen G. Lijmer, Martin H. Prins, Gouke J. Bonsel, and Patrick M.M. Bossuyt. The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology*, 56(11):1129–1135, 2003. ISSN 0895-4356. doi: https://doi.org/10.1016/S0895-4356(03)00177-X. URL https://www.sciencedirect.com/science/article/pii/S089543560300177X.

Matteo Rota and Laura Antolini. Finding the optimal cut-point for gaussian and gamma distributed biomarkers. *Computational Statistics and Data Analysis*, 69:1–14, 01 2014. doi: 10.1016/j.csda.2013.07.015.

S.G. Hilsenbeck and G.M. Clark. Practical p-value adjustment for optimally selected cutpoints. *Statistics in Medicine*, 15(1):103–112, 1996. doi: 10.1002/(SICI)1097-0258(19960115)15:1⟨103::AID-SIM156⟩3.0.CO;2-Y. URL https://pubmed.ncbi.nlm.nih.gov/8614741/.

V.S. Pugachev. Chapter 8 - estimation of distributions. In V.S. Pugachev, editor, *Probability Theory and Mathematical Statistics for Engineers*, pages 273–310. Pergamon, 1984. ISBN 978-0-08-029148-2. doi: https://doi.org/10.1016/B978-0-08-029148-2.50012-4. URL https://www.sciencedirect.com/science/article/pii/B9780080291482500124.

Jun Zhang and Kai F Yu. What's the relative risk? a method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*, 280(19):1690–1691, 1998. doi: 10.1001/jama.280.19.1690. URL https://doi.org/10.1001/jama.280.19.1690.

Xiao-Hua Zhou, Nancy A Obuchowski, and Donna K McClish. *Statistical Methods in Diagnostic Medicine*. Wiley, 2011.

Bradley Efron and Robert J Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1994. URL https://doi.org/10.1201/9780429246593.

Farrokh Habibzadeh, Parham Habibzadeh, and Mohammad Yadollahie. On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochemia Medica*, 26(3):297–307, 2016. doi: 10.11613/BM.2016.034. URL https://doi.org/10.11613/BM.2016.034.

Posit team. *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA, 2025. URL http://www.posit.co/.

J. Coste and J. Pouchot. A grey zone for quantitative diagnostic and screening tests. *International Journal of Epidemiology*, 32(2):304–313, 2003. doi: 10.1093/ije/dyg054. URL https://pubmed.ncbi.nlm.nih.gov/12714554/.

P. Royston and W. Sauerbrei. Stability of multivariable fractional polynomial models with selection of variables and transformations: a bootstrap investigation. *Statistics in Medicine*, 22:639–659, 2003. doi: 10.1002/sim.1310. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.1310.

A. J. Vickers. Decision analysis for the evaluation of diagnostic tests, prediction models and molecular markers. *The American Statistician*, 62(4):314–320, 2008. doi: 10.1198/000313008X370302. URL https://pubmed.ncbi.nlm.nih.gov/19132141/.

Appendix A

Simulation code for Cut-off evaluation

A.1 Data Simulation Scenarios

Null Scenario

```
set.seed(params$report_seed)
n <- 2000
mu <- 0
sigma <- 1
prev <- 0.1

x <- rbinom(n, 1, prev)
y <- rnorm(n, mean = 0, sd = sigma)

dades_n <- data.frame(y, x)
```

Small Scenario

```
set.seed(params$report_seed)
n <- 1000
mu <- 0.5
sigma <- 1
prev <- 0.1

x <- rbinom(n, 1, prev)
```

```
dades_ep <- tibble(x = x) %>%
  group_by(x) %>%
  mutate(
    y = case_when(
      x == 1 ~ rnorm(n(), mean = mu, sd = sigma),
      x == 0 ~ rnorm(n(), mean = 0, sd = 1)
    )
  )
```

Moderate Scenario

```
set.seed(params$report_seed)
n <- 500
mu <- 1
sigma <- 1
prev <- 0.1

x <- rbinom(n, 1, prev)

dades_em <- tibble(x = x) %>%
  group_by(x) %>%
  mutate(
    t = case_when(
      x == 1 ~ rnorm(n(), mean = mu, sd = sigma),
      x == 0 ~ rnorm(n(), mean = 0, sd = 1)
    )
  )
```

A.2 Example of Loop for Null Scenario

```
# Normal distribution: Null scenario
```{r}
nboot<- 75
```

```{r}
print(params$report_seed)
```

```{r echo=FALSE, message=FALSE, warning=FALSE}
#set.seed(1858)
set.seed(params$report_seed)

n <- 2000 # observacions
mu <- 0   # constant
sigma <- 1  # sd de l'error
prev<-0.1

# Variables
x <- rbinom(n, 1, prev)
y <- rnorm(n, mean = 0, sd= sigma)

# Dades simulades
dades_n <- data.frame(y, x)


# Save dataframes
# dades <- list()
# save(dades, file = "dades_loop_null_v2.Rdata")

load("dades_loop_null_v2.Rdata")

var_seed <- params$report_seed

dades <- dades |> purrr::list_assign(var_seed = dades_n)
```
```

```
names(dades)[names(dades) == "var_seed"] <- var_seed

save(dades, file = "dades_loop_null_v2.Rdata")


#head(dades_n)
#tapply(dades_n$y, dades_n$x, mean)
#tapply(dades_n$y, dades_n$x, sd)

```


`r tab_cap("Description of data simulation: Null scenario")`

```{r echo=FALSE, message=FALSE, warning=FALSE}
dades_n |>
  dplyr::select(x, y) |>
  gtsummary::tbl_summary(
    type = x ~"categorical",
        digits = all_continuous() ~ 2)


model <- lm(y ~ x,data=dades_n)

tab_model(model,show.se=T)
```


`r fig_cap("Density comparison plot")`
```{r echo=FALSE, warning=FALSE, message=FALSE, cache=FALSE}

dades_n$x_c<-factor(dades_n$x,
                    levels = c(0,1),
                    labels = c("Non-disease", "Disease"))
```

```
dades_n$y_pred<-dades_n$y

d <- dplyr::select(dades_n,c(y_pred, x_c, x))
d <- filter(d, x_c%in%c("Disease", "Non-disease"))
d$x_c<-factor(d$x_c, levels = c("Disease", "Non-disease"))


ggplot(d, aes(y_pred,fill=x_c,color=x_c)) +
 geom_density(alpha = 0.1) +
   geom_vline(xintercept = mean(d$y_pred), linetype = "dashed", color
      ↪ = "black", linewidth = 0.8) +
 geom_vline(xintercept = mean(d$y_pred[d$x_c=="Non-disease"]),
    ↪ linetype = "dashed", color = "red", linewidth = 0.8,alpha =
    ↪ 0.8) +
 geom_vline(xintercept = mean(d$y_pred[d$x_c=="Disease"]), linetype =
    ↪ "dashed", color = "blue", linewidth = 1,alpha = 0.8) +
 theme_minimal()+labs(x="Predictor (y)")+
  ggtitle("Density comparison")+
  guides(fill = guide_legend(title = "Binary outcome"),
         legend.background = element_rect(fill = "white"), colour =
           ↪ guide_legend(title = "Binary outcome"))

```


`r fig_cap("ROC Curve")`
```{r echo=FALSE, warning=FALSE, message=FALSE, cache=FALSE}

myroc <- roc(d$x, d$y_pred, ci=TRUE, direction="<")
plot(myroc, main="ROC curve and AUC")
text(0.4, 0.1, paste0("AUC:", round(auc(myroc), 2),
                      " [",round(myroc$ci[1],2), ", ",round(myroc$ci
                        ↪ [2],2),"]" ))
```



```
## Methods
### New approach

```{r echo=FALSE, message=FALSE, warning=FALSE}
d_cut <- data.frame(
  x = as.numeric(d$x),
  y_pred = as.numeric(d$y_pred)
)
```

Bootstrapping

```{r echo=FALSE, warning=FALSE, message=FALSE, cache=FALSE}

# set.seed(123)
boots<-seq(1:nrow(d))
p<-c()
for(j in 1:nboot){
  # print(j)
  llista_s <- sample(boots, replace=T)
  ds <- d[llista_s,]
  x_outcome <-quantile(d$y_pred,probs = seq(0,1,0.01),na.rm=T)

  ai <- c()
  bi <- c()
  val <- c()
  hr <- c()
  p1x <- c()
  p2x <- c()
  cind <- c()
  lr <- c()
  br <- c()
  pvalue <- c()
  voltes <- c()
  for(i in x_outcome){
    ds$y <- cut(ds$y_pred, breaks = c(-Inf,i, Inf))
    if(all(table(ds$y, ds$x)>1)){
      m <- glm(x~y,data=ds, family=poisson(link = "log"))
```
```

```
      l <- summary(m)
      pvalue <- c(pvalue, l$coefficients[2,4])
      ai <- c(ai,AIC(m))
      bi <- c(bi, BIC(m))
      val <- c(val,i)
      h <- exp(l$coefficients[2,1])
      p1 <- round(exp(confint(m))[2,1],2)
      p2 <- round(exp(confint(m))[2,1],2)
      hr <- c(hr,h)
      p1x <- c(p1x, p1)
      p2x <- c(p2x, p2)
      br <- c(br, BrierScore(m))
      voltes <- c(voltes, j)
    }
  }
  d3 <- data.frame(voltes, val, hr, p1x,p2x,pvalue,br)

  minim <- min(d3$pvalue[d3$voltes==j])
  coff <- min(d3[d3$pvalue==minim& d3$voltes==j, "val"])
   print(coff)
   p<-rbind(p,coff)
}

  p<-as.data.frame(p)

 # save(p, file = file.path("I:\CTebe\2_Projectes\2025_01TFMLBlanc\2_
    ↪ Dades\2_Analisi", "Null_scenario.Rda"))

# load(file="I:/CTebe/2_Projectes/2025_01TFMLBlanc/2_Dades/2_Analisi/
   ↪ Null/Null_scenario.Rda")

```


`r fig_cap("Density function with high-density local maxima")`

```{r echo=FALSE, warning=FALSE, message=FALSE, cache=FALSE}
funcio_densitat <- density(p$V1)
pics <- which(diff(sign(diff(funcio_densitat$y))) == -2) + 1
```

```
maximlocal_x <- funcio_densitat$x[pics]
maximlocal_y <- funcio_densitat$y[pics]

data.frame(valors=round(maximlocal_x,2),densitat=round(maximlocal_y,2)
   ↪ )%>%
  gt()


ggplot(data.frame(x = funcio_densitat$x, y = funcio_densitat$y), aes(x
   ↪  = x, y = y)) +
  geom_line(color = "blue") +
  geom_point(data = data.frame(x = maximlocal_x[maximlocal_y>mean(
     ↪ funcio_densitat$y)],
                               y = maximlocal_y[maximlocal_y>mean(
                                  ↪ funcio_densitat$y)]),
             aes(x = x, y = y), col = "red", pch = 16)+
  geom_vline(aes(xintercept=mean(p$V1[p$V1!="Inf"]), color="orange"),
     ↪ show_guide=TRUE)+
  labs(title = "Density function with high-density local maxima",
       x = "X-axis",y = "Density")+
  scale_color_manual("Cut-off", values=c(orange="orange"),labels=c( "E
     ↪ (values)"))+
  theme_bw()
```



```
##### Expected value
```{r echo=FALSE, warning=FALSE, message=FALSE, cache=FALSE}

xintercept=mean(p$V1[p$V1!="Inf"])

d$y_cat_b <- ifelse(d$y_pred>xintercept, "Positive", "Negative")

#construim la taula
tab_b <- table(d$y_cat_b, d$x)
tab_b <- tab_b[c(2,1),c(2,1)]
x <- diagnostic(tab_b)
```
```

```
```{r echo=FALSE, warning=FALSE, message=FALSE, cache=FALSE}
y <- epiR::epi.tests(tab_b)
y
```


The diagnostic indices taking `r round(xintercept,3)` as the cut-off
    ↪ point are as follows

```{r echo=FALSE, warning=FALSE, message=FALSE, cache=FALSE}
kable(round(x[c(1:4,9),],3), caption = "Predictor", align = "l") %>%
 kable_styling(bootstrap_options = c("striped","hover"))

### Confusion matrix (paquete caret)
d$y_catfn_b <- factor(d$y_cat_b,
                      levels=c("Positive", "Negative"),
                      labels=c("Non-disease", "Disease"))
machine_matriz <- caret::confusionMatrix(d$y_catfn_b, d$x_c)

metrics_boot_expect <- data.frame(
  method = "New approach: Expected value",
  optimal_cutpoint = round(xintercept, 3),
  sensitivity = x["Sensitivity", "Estim."],
  sensitivity_low = x["Sensitivity", "Low.lim(95%)"],
  sensitivity_high = x["Sensitivity", "Up.lim(95%)"],

  specificity = x["Specificity", "Estim."],
  specificity_low = x["Specificity", "Low.lim(95%)"],
  specificity_high = x["Specificity", "Up.lim(95%)"],

  ppv = x["Pos.Pred.Val.", "Estim."],
  ppv_low = x["Pos.Pred.Val.", "Low.lim(95%)"],
  ppv_high = x["Pos.Pred.Val.", "Up.lim(95%)"],

  npv = x["Neg.Pred.Val.", "Estim."],
  npv_low = x["Neg.Pred.Val.", "Low.lim(95%)"],
  npv_high = x["Neg.Pred.Val.", "Up.lim(95%)"],
```
```

```
  accuracy = x["Accuracy", "Estim."],
  accuracy_low = x["Accuracy", "Low.lim(95%)"],
  accuracy_high = x["Accuracy", "Up.lim(95%)"]
)
```


Maximum density

```{r echo=FALSE, warning=FALSE, message=FALSE, cache=FALSE}

# x con mayor densidad
xintercept <- maximlocal_x[which.max(maximlocal_y)]


#construct dichotomous
d$y_cat_b <- ifelse(d$y_pred>xintercept, "Positive", "Negative")

#construim la taula
tab_b <- table(d$y_cat_b, d$x)
tab_b <- tab_b[c(2,1),c(2,1)]
x <- diagnostic(tab_b)
```

```{r echo=FALSE, warning=FALSE, message=FALSE, cache=FALSE}
y <- epiR::epi.tests(tab_b)
y
```


The diagnostic indices taking `r round(xintercept,3)` as the cut-off point are as follows

```{r echo=FALSE, warning=FALSE, message=FALSE, cache=FALSE}
kable(round(x[c(1:4,9),],3), caption = "Predictor", align = "l") %>%
```

```
 kable_styling(bootstrap_options = c("striped","hover"))

### Confusion matrix (paquete caret) - https://rdrr.io/cran/caret/man/
    ↪ confusionMatrix.html
d$y_catfn_b <- factor(d$y_cat_b,
                      levels=c("Positive", "Negative"),
                      labels=c("Non-disease", "Disease"))
machine_matriz <- caret::confusionMatrix(d$y_catfn_b, d$x_c)


metrics_boot_maxdens <- data.frame(
  method = "New approach: Maximum density",
  optimal_cutpoint = round(xintercept, 3),
  sensitivity = x["Sensitivity", "Estim."],
  sensitivity_low = x["Sensitivity", "Low.lim(95%)"],
  sensitivity_high = x["Sensitivity", "Up.lim(95%)"],

  specificity = x["Specificity", "Estim."],
  specificity_low = x["Specificity", "Low.lim(95%)"],
  specificity_high = x["Specificity", "Up.lim(95%)"],

  ppv = x["Pos.Pred.Val.", "Estim."],
  ppv_low = x["Pos.Pred.Val.", "Low.lim(95%)"],
  ppv_high = x["Pos.Pred.Val.", "Up.lim(95%)"],

  npv = x["Neg.Pred.Val.", "Estim."],
  npv_low = x["Neg.Pred.Val.", "Low.lim(95%)"],
  npv_high = x["Neg.Pred.Val.", "Up.lim(95%)"],

  accuracy = x["Accuracy", "Estim."],
  accuracy_low = x["Accuracy", "Low.lim(95%)"],
  accuracy_high = x["Accuracy", "Up.lim(95%)"]
)
```



```
### Youden's J statistic

```{r echo=FALSE, message=FALSE, warning=FALSE}

youden_res <- optimal.cutpoints(
  X = "y_pred",
  status = "x",
  tag.healthy = 0,
  methods = "Youden",
  data = d_cut,
  ci.fit = TRUE,
  conf.level = 0.95,
  trace = FALSE
)


summary(youden_res)

# Metrics results
opt_youden <- summary(youden_res)$Youden$Global$optimal.cutoff

cutoff_youden <- opt_youden$cutoff

se_youden<-opt_youden$Se
sp_youden<-opt_youden$Sp
ppv_youden<-opt_youden$PPV
npv_youden<-opt_youden$NPV

# Calculation CI
sensitivity <- se_youden[1, "Value"]
sensitivity_low <-se_youden[1, "ll"]
sensitivity_high <- se_youden[1, "ul"]

specificity <- sp_youden[1, "Value"]
specificity_low <- sp_youden[1, "ll"]
specificity_high <- sp_youden[1, "ul"]

ppv <- ppv_youden[1, "Value"]
ppv_low <- ppv_youden[1, "ll"]
```
```

```
ppv_high <- ppv_youden[1, "ul"]

npv <- npv_youden[1, "Value"]
npv_low <- npv_youden[1, "ll"]
npv_high <- npv_youden[1, "ul"]



# Classification

# Confusion matrix
d$y_cat <- ifelse(d$y_pred >= cutoff_youden, "Predicted disease", "
   ↪ Predicted non-disease")
tab <- table(d$y_cat, d$x_c)
tab <- tab[,c(1,2)]

kable(tab, caption = "Contingency Table - Youden", align = "c") %>%
   kable_styling(bootstrap_options = c("striped", "hover", "condensed"
      ↪ , "responsive"),
                 full_width = F, position = "center")

# Diagnostic
diag <- epi.tests(tab)
diag


# Accuracy
correct<-sum(tab[1,1], tab[2,2])

total <- sum(tab)

acc_youden <- as.numeric(correct/total)
ci_acc <- binom.test(correct, total, conf.level = 0.95)$conf.int

# Metrics
metrics_youden <- data.frame(
  method = "Youden",
  optimal_cutpoint = cutoff_youden,
```

```
  sensitivity = sensitivity,
  sensitivity_low = sensitivity_low,
  sensitivity_high = sensitivity_high,

  specificity = specificity,
  specificity_low = specificity_low,
  specificity_high = specificity_high,

  ppv = ppv,
  ppv_low = ppv_low,
  ppv_high = ppv_high,

  npv = npv,
  npv_low = npv_low,
  npv_high = npv_high,

  accuracy = acc_youden,
  accuracy_low = ci_acc[1],
  accuracy_high = ci_acc[2]
)


metrics_t<-metrics_youden %>%
   select(method, optimal_cutpoint, sensitivity, sensitivity_low,
      ↪ sensitivity_high,
          specificity, specificity_low, specificity_high, ppv, ppv_low
             ↪ , ppv_high, npv, npv_low, npv_high, accuracy, accuracy
             ↪ _low, accuracy_high) %>%
   gt() %>%
   tab_header(title = "Optimal Cut-off Point Metrics - Youden") %>%
   fmt_number(
     columns = where(is.numeric),
     decimals = 3
   )
```

```

 
 
```

```
### Euclidean distance
```{r echo=FALSE, message=FALSE, warning=FALSE}
# Se and Sp
Se <- myroc$sensitivities
Sp <- myroc$specificities
thres <- myroc$thresholds

eucl_res <- sqrt((1 - Se)^2 + (1 - Sp)^2)
best_index <- which.min(eucl_res)

optimal_threshold <- thres[best_index]


# Classification

# Confusion matrix
d$y_cat <- ifelse(d$y_pred >=optimal_threshold,"Predicted disease", "
   ↪ Predicted non-disease")

tab <- table(d$y_cat, d$x_c)
tab <- tab[,c(1,2)]

kable(tab, caption = "Contingency Table - Euclidean", align = "c") %>%
   kable_styling(bootstrap_options = c("striped", "hover", "condensed"
      ↪ , "responsive"),
                 full_width = F, position = "center")

# Diagnostic
diag <- epi.tests(tab)
diag

metrics<-diag[["detail"]]


# Calculation IC
# PPV, NPV, Acc
se <- metrics[metrics$statistic == "se", ]
sp <- metrics[metrics$statistic == "sp", ]
ppv  <- metrics[metrics$statistic == "pv.pos", ]
```
```

```
npv  <- metrics[metrics$statistic == "pv.neg", ]
acc  <- metrics[metrics$statistic == "diag.ac", ]


# Metrics
metrics_eucl <- data.frame(
  method = "Euclidean distance",
  optimal_cutpoint = optimal_threshold,

  sensitivity = se$est,
  sensitivity_low = se$lower,
  sensitivity_high = se$upper,

  specificity = sp$est,
  specificity_low = sp$lower,
  specificity_high = sp$upper,

  ppv = ppv$est,
  ppv_low = ppv$lower,
  ppv_high = ppv$upper,

  npv = npv$est,
  npv_low = npv$lower,
  npv_high = npv$upper,

  accuracy = acc$est,
  accuracy_low = acc$lower,
  accuracy_high = acc$upper
)



metrics_t<-metrics_eucl %>%
   select(method, optimal_cutpoint, sensitivity, sensitivity_low,
      ↪ sensitivity_high,
          specificity, specificity_low, specificity_high, ppv, ppv_low
             ↪ , ppv_high, npv, npv_low, npv_high, accuracy, accuracy
             ↪ _low, accuracy_high) %>%
   gt() %>%
```

```
   tab_header(title = "Optimal Cut-off Point Metrics - Euclidean
      ↪ distance") %>%
   fmt_number(
     columns = where(is.numeric),
     decimals = 3
   )


```


Maximum product of Se and Sp

```{r echo=FALSE, message=FALSE, warning=FALSE}

mprod_res <- optimal.cutpoints(
  X = "y_pred",
  status = "x",
  tag.healthy = 0,
  methods = "MaxProdSpSe",
  data = d_cut,
  ci.fit = TRUE,
  conf.level = 0.95,
  trace = FALSE
)



summary(mprod_res)

# Results
opt_mprod <- summary(mprod_res)$MaxProdSpSe$Global$optimal.cutoff

cutoff_mprod <- opt_mprod$cutoff

se_mprod<-opt_mprod$Se
sp_mprod<-opt_mprod$Sp
ppv_mprod<-opt_mprod$PPV
```

```
npv_mprod<-opt_mprod$NPV

# Calculation IC
sensitivity <- se_mprod[1, "Value"]
sensitivity_low <-se_mprod[1, "ll"]
sensitivity_high <- se_mprod[1, "ul"]

specificity <- sp_mprod[1, "Value"]
specificity_low <- sp_mprod[1, "ll"]
specificity_high <- sp_mprod[1, "ul"]

ppv <- ppv_mprod[1, "Value"]
ppv_low <- ppv_mprod[1, "ll"]
ppv_high <- ppv_mprod[1, "ul"]

npv <- npv_mprod[1, "Value"]
npv_low <- npv_mprod[1, "ll"]
npv_high <- npv_mprod[1, "ul"]



# Classification

# Confusion matrix
d$y_cat <- ifelse(d$y_pred >= cutoff_mprod, "Predicted disease", "
   ↪ Predicted non-disease")


tab <- table(d$y_cat, d$x_c)
tab <- tab[,c(1,2)]

kable(tab, caption = "Contingency Table - Maximum product Se*Sp",
   ↪ align = "c") %>%
   kable_styling(bootstrap_options = c("striped", "hover", "condensed"
      ↪ , "responsive"),
                 full_width = F, position = "center")

# Diagnostic
diag <- epi.tests(tab)
```

```
diag


# Accuracy
correct<-sum(tab[1,1], tab[2,2])

total <- sum(tab)

acc_mprod <- as.numeric(correct/total)
ci_acc <- binom.test(correct, total, conf.level = 0.95)$conf.int

# Metrics
metrics_mprod <- data.frame(
  method = "Maximum product Se*Sp",
  optimal_cutpoint = cutoff_mprod,

  sensitivity = sensitivity,
  sensitivity_low = sensitivity_low,
  sensitivity_high = sensitivity_high,

  specificity = specificity,
  specificity_low = specificity_low,
  specificity_high = specificity_high,

  ppv = ppv,
  ppv_low = ppv_low,
  ppv_high = ppv_high,

  npv = npv,
  npv_low = npv_low,
  npv_high = npv_high,

  accuracy = acc_mprod,
  accuracy_low = ci_acc[1],
  accuracy_high = ci_acc[2]
)


metrics_t<-metrics_mprod %>%
```

```
    select(method, optimal_cutpoint, sensitivity, sensitivity_low,
       ↪ sensitivity_high,
           specificity, specificity_low, specificity_high, ppv, ppv_low
              ↪ , ppv_high, npv, npv_low, npv_high, accuracy, accuracy
              ↪ _low, accuracy_high) %>%
    gt() %>%
    tab_header(title = "Optimal Cut-off Point Metrics - Max. product Se
       ↪ *Sp") %>%
    fmt_number(
      columns = where(is.numeric),
      decimals = 3
    )

```


Index of union (IU)
```{r echo=FALSE, message=FALSE, warning=FALSE}

# Calculation AUC
auc_val <- as.numeric(auc(myroc))


roc_df <- data.frame(
  threshold = myroc$thresholds,
  sens = myroc$sensitivities,
  spec = myroc$specificities

)

# Calculation IU
roc_df$IU <- abs(roc_df$sens - auc_val) + abs(roc_df$spec - auc_val)
best_iu <- roc_df[which.min(roc_df$IU), ]



# Classification
```

```r
# Confusion matrix
d$y_cat <- ifelse(d$y_pred >= best_iu$threshold, "Predicted disease",
   ↪ "Predicted non-disease")


tab <- table(d$y_cat, d$x_c)
tab <- tab[,c(1,2)]

kable(tab, caption = "Contingency Table - Index of Union", align = "c"
   ↪ ) %>%
   kable_styling(bootstrap_options = c("striped", "hover", "condensed"
      ↪ , "responsive"),
                 full_width = F, position = "center")

# Diagnostic
diag <- epi.tests(tab)
diag


x_iu<-diagnostic(tab)

# Calculation metrics with IC
se <- x_iu["Sensitivity", ]
sp <- x_iu["Specificity", ]
ppv  <- x_iu["Pos.Pred.Val.", ]
npv  <- x_iu["Neg.Pred.Val.", ]
acc  <- x_iu["Accuracy", ]


# Metrics
metrics_iu <- data.frame(
  method = "Index of Union (IU)",
  optimal_cutpoint = best_iu$threshold,

  sensitivity = se[1],
  sensitivity_low = se[2],
  sensitivity_high = se[3],

  specificity = sp[1],
```

```
  specificity_low = sp[2],
  specificity_high = sp[3],

  ppv = ppv[1],
  ppv_low = ppv[2],
  ppv_high = ppv[3],

  npv = npv[1],
  npv_low = npv[2],
  npv_high = npv[3],

  accuracy = acc[1],
  accuracy_low = acc[2],
  accuracy_high = acc[3]
)


metrics_t<-metrics_iu %>%
   select(method, optimal_cutpoint, sensitivity, sensitivity_low,
      ↪ sensitivity_high,
          specificity, specificity_low, specificity_high, ppv, ppv_low
             ↪ , ppv_high, npv, npv_low, npv_high, accuracy, accuracy
             ↪ _low, accuracy_high) %>%
   gt() %>%
   tab_header(title = "Optimal Cut-off Point Metrics - Index of Union"
      ↪ ) %>%
   fmt_number(
     columns = where(is.numeric),
     decimals = 3
   )

```

```
### Cost approach
```{r echo=FALSE, message=FALSE, warning=FALSE}

cost_ap_res <- optimal.cutpoints(
  X = "y_pred",
  status = "x",
  tag.healthy = 0,
  methods = "CB",
  data = d_cut,
  control = control.cutpoints(
  costs.ratio = 5), # It costs 5 times more to predict a FN than a FP
  ci.fit = TRUE,
  conf.level = 0.95,
  trace = FALSE
)

summary(cost_ap_res)

# Results
opt_cb <- summary(cost_ap_res)$CB$Global$optimal.cutoff

cutoff_cb <- opt_cb$cutoff

se_cb<-opt_cb$Se
sp_cb<-opt_cb$Sp
ppv_cb<-opt_cb$PPV
npv_cb<-opt_cb$NPV

# Calculation CI
sensitivity <- se_cb[1, "Value"]
sensitivity_low <-se_cb[1, "ll"]
sensitivity_high <- se_cb[1, "ul"]

specificity <- sp_cb[1, "Value"]
specificity_low <- sp_cb[1, "ll"]
specificity_high <- sp_cb[1, "ul"]

ppv <- ppv_cb[1, "Value"]
ppv_low <- ppv_cb[1, "ll"]
```
```

```
ppv_high <- ppv_cb[1, "ul"]

npv <- npv_cb[1, "Value"]
npv_low <- npv_cb[1, "ll"]
npv_high <- npv_cb[1, "ul"]



# Classification

# Confusion matrix
d$y_cat <- ifelse(d$y_pred >= cutoff_cb, "Predicted disease", "
   ↪ Predicted non-disease")


tab <- table(d$y_cat, d$x_c)
tab <- tab[,c(1,2)]

kable(tab, caption = "Contingency Table - Cost approach", align = "c")
   ↪  %>%
   kable_styling(bootstrap_options = c("striped", "hover", "condensed"
      ↪ , "responsive"),
                 full_width = F, position = "center")

# Diagnostic
diag <- epi.tests(tab)
diag


# Accuracy
correct<-sum(tab[1,1], tab[2,2])

total <- sum(tab)

acc_cb <- as.numeric(correct/total)
ci_acc <- binom.test(correct, total, conf.level = 0.95)$conf.int

# Metrics
metrics_cost <- data.frame(
```

```
  method = "Cost approach",
  optimal_cutpoint = cutoff_cb,

  sensitivity = sensitivity,
  sensitivity_low = sensitivity_low,
  sensitivity_high = sensitivity_high,

  specificity = specificity,
  specificity_low = specificity_low,
  specificity_high = specificity_high,

  ppv = ppv,
  ppv_low = ppv_low,
  ppv_high = ppv_high,

  npv = npv,
  npv_low = npv_low,
  npv_high = npv_high,

  accuracy = acc_cb,
  accuracy_low = ci_acc[1],
  accuracy_high = ci_acc[2]
)

metrics_t<-metrics_cost %>%
   select(method, optimal_cutpoint, sensitivity, sensitivity_low,
      ↪ sensitivity_high,
          specificity, specificity_low, specificity_high, ppv, ppv_low
             ↪ , ppv_high, npv, npv_low, npv_high, accuracy, accuracy
             ↪ _low, accuracy_high) %>%
   gt() %>%
   tab_header(title = "Optimal Cut-off Point Metrics - Cost approach")
      ↪  %>%
   fmt_number(
     columns = where(is.numeric),
     decimals = 3
   )

```

```
 
 




### Misclassification cost

```{r echo=FALSE, message=FALSE, warning=FALSE}

misscost_res <- optimal.cutpoints(
  X = "y_pred",
  status = "x",
  tag.healthy = 0,
  methods = "MCT",
  data = d_cut,
  ci.fit = TRUE,
  conf.level = 0.95,
  trace = FALSE
)


summary(misscost_res)

# Results
opt_misscost <- summary(misscost_res)$MCT$Global$optimal.cutoff

cutoff_misscost <- opt_misscost$cutoff

se_misscost<-opt_misscost$Se
sp_misscost<-opt_misscost$Sp
ppv_misscost<-opt_misscost$PPV
npv_misscost<-opt_misscost$NPV

# Calculation CI
sensitivity <- se_misscost[1, "Value"]
```
```

```
sensitivity_low <-se_misscost[1, "ll"]
sensitivity_high <- se_misscost[1, "ul"]

specificity <- sp_misscost[1, "Value"]
specificity_low <- sp_misscost[1, "ll"]
specificity_high <- sp_misscost[1, "ul"]

ppv <- ppv_misscost[1, "Value"]
ppv_low <- ppv_misscost[1, "ll"]
ppv_high <- ppv_misscost[1, "ul"]

npv <- npv_misscost[1, "Value"]
npv_low <- npv_misscost[1, "ll"]
npv_high <- npv_misscost[1, "ul"]



# Classification

# Confusion matrix
d$y_cat <- ifelse(d$y_pred >= cutoff_misscost, "Predicted disease", "
   ↪ Predicted non-disease")


tab <- table(d$y_cat, d$x_c)
tab <- tab[,c(1,2)]

kable(tab, caption = "Contingency Table - Misclassification cost",
   ↪ align = "c") %>%
   kable_styling(bootstrap_options = c("striped", "hover", "condensed"
      ↪ , "responsive"),
                 full_width = F, position = "center")

# Diagnostic
diag <- epi.tests(tab)
diag



# Accuracy
```

```
correct<-sum(tab[1,1], tab[2,2])

total <- sum(tab)

acc_misscost <- as.numeric(correct/total)
ci_acc <- binom.test(correct, total, conf.level = 0.95)$conf.int

# Metrics
metrics_misscost <- data.frame(
  method = "Misclassification cost",
  optimal_cutpoint = cutoff_misscost[1],

  sensitivity = sensitivity,
  sensitivity_low = sensitivity_low,
  sensitivity_high = sensitivity_high,

  specificity = specificity,
  specificity_low = specificity_low,
  specificity_high = specificity_high,

  ppv = ppv,
  ppv_low = ppv_low,
  ppv_high = ppv_high,

  npv = npv,
  npv_low = npv_low,
  npv_high = npv_high,

  accuracy = acc_misscost,
  accuracy_low = ci_acc[1],
  accuracy_high = ci_acc[2]
)



metrics_t<-metrics_misscost %>%
   select(method, optimal_cutpoint, sensitivity, sensitivity_low,
      ↪ sensitivity_high,
```

```
          specificity, specificity_low, specificity_high, ppv, ppv_low
             ↪ , ppv_high, npv, npv_low, npv_high, accuracy, accuracy
             ↪ _low, accuracy_high) %>%
  gt() %>%
  tab_header(title = "Optimal Cut-off Point Metrics -
     ↪ Misclassification cost") %>%
  fmt_number(
    columns = where(is.numeric),
    decimals = 3
  )

```


Diagnosis Odds Ratio (DOR)

```{r echo=FALSE, message=FALSE, warning=FALSE}

dor_res <- optimal.cutpoints(
  X = "y_pred",
  status = "x",
  tag.healthy = 0,
  methods = "MaxDOR",
  data = d_cut,
  ci.fit = TRUE,
  conf.level = 0.95,
  trace = FALSE
)

summary(dor_res)

# Results
opt_dor <- summary(dor_res)$MaxDOR$Global$optimal.cutoff

cutoff_dor <- opt_dor$cutoff

se_dor<-opt_dor$Se
sp_dor<-opt_dor$Sp
```

```
ppv_dor<-opt_dor$PPV
npv_dor<-opt_dor$NPV

# Calculation CI
sensitivity <- se_dor[1, "Value"]
sensitivity_low <-se_dor[1, "ll"]
sensitivity_high <- se_dor[1, "ul"]

specificity <- sp_dor[1, "Value"]
specificity_low <- sp_dor[1, "ll"]
specificity_high <- sp_dor[1, "ul"]

ppv <- ppv_dor[1, "Value"]
ppv_low <- ppv_dor[1, "ll"]
ppv_high <- ppv_dor[1, "ul"]

npv <- npv_dor[1, "Value"]
npv_low <- npv_dor[1, "ll"]
npv_high <- npv_dor[1, "ul"]



# Classification

# Confusion matrix
d$y_cat <- ifelse(d$y_pred >= cutoff_dor, "Predicted disease", "
   ↪ Predicted non-disease")


tab <- table(d$y_cat, d$x_c)
tab <- tab[,c(1,2)]

kable(tab, caption = "Contingency Table - Diagnostic Odds Ratio",
   ↪ align = "c") %>%
   kable_styling(bootstrap_options = c("striped", "hover", "condensed"
      ↪ , "responsive"),
                 full_width = F, position = "center")

# Diagnostic
```

```
diag <- epi.tests(tab)
diag


# Accuracy
correct<-sum(tab[1,1], tab[2,2])

total <- sum(tab)

acc_dor <- as.numeric(correct/total)
ci_acc <- binom.test(correct, total, conf.level = 0.95)$conf.int

# Metrics
metrics_dor <- data.frame(
  method = "Diagnostic Odds Ratio",
  optimal_cutpoint = cutoff_dor,

  sensitivity = sensitivity,
  sensitivity_low = sensitivity_low,
  sensitivity_high = sensitivity_high,

  specificity = specificity,
  specificity_low = specificity_low,
  specificity_high = specificity_high,

  ppv = ppv,
  ppv_low = ppv_low,
  ppv_high = ppv_high,

  npv = npv,
  npv_low = npv_low,
  npv_high = npv_high,

  accuracy = acc_dor,
  accuracy_low = ci_acc[1],
  accuracy_high = ci_acc[2]
)
```

```
metrics_t<-metrics_dor %>%
   select(method, optimal_cutpoint, sensitivity, sensitivity_low,
      ↪ sensitivity_high,
          specificity, specificity_low, specificity_high, ppv, ppv_low
             ↪ , ppv_high, npv, npv_low, npv_high, accuracy, accuracy
             ↪ _low, accuracy_high) %>%
   gt() %>%
   tab_header(title = "Optimal Cut-off Point Metrics - Diagnostic Odds
      ↪  Ratio") %>%
   fmt_number(
     columns = where(is.numeric),
     decimals = 3
   )

```


Min p-value

```{r echo=FALSE, message=FALSE, warning=FALSE}
minpvalue_res <- optimal.cutpoints(
  X = "y_pred",
  status = "x",
  tag.healthy = 0,
  methods = "MinPvalue",
  data = d_cut,
  ci.fit = TRUE,
  conf.level = 0.95,
  trace = FALSE)


summary(minpvalue_res)

# Results
opt_minp <- summary(minpvalue_res)$MinPvalue$Global$optimal.cutoff

cutoff_minp <- opt_minp$cutoff
```

```
se_minp<-opt_minp$Se
sp_minp<-opt_minp$Sp
ppv_minp<-opt_minp$PPV
npv_minp<-opt_minp$NPV


# Calculation CI
sensitivity <- se_minp[1, "Value"]
sensitivity_low <-se_minp[1, "ll"]
sensitivity_high <- se_minp[1, "ul"]

specificity <- sp_minp[1, "Value"]
specificity_low <- sp_minp[1, "ll"]
specificity_high <- sp_minp[1, "ul"]

ppv <- ppv_minp[1, "Value"]
ppv_low <- ppv_minp[1, "ll"]
ppv_high <- ppv_minp[1, "ul"]

npv <- npv_minp[1, "Value"]
npv_low <- npv_minp[1, "ll"]
npv_high <- npv_minp[1, "ul"]


# Classification

# Confusion matrix
d$y_cat <- ifelse(d$y_pred >= cutoff_minp, "Predicted disease", "
   ↪ Predicted non-disease")


tab <- table(d$y_cat, d$x_c)
tab <- tab[,c(1,2)]

kable(tab, caption = "Contingency Table - Min p-value", align = "c")
   ↪ %>%
   kable_styling(bootstrap_options = c("striped", "hover", "condensed"
      ↪ , "responsive"),
                 full_width = F, position = "center")
```

```
# Diagnostic
diag <- epi.tests(tab)
diag


# Accuracy
correct<-sum(tab[1,1], tab[2,2])

total <- sum(tab)

acc <- as.numeric(correct/total)
ci_acc <- binom.test(correct, total, conf.level = 0.95)$conf.int



# Metrics
metrics_minp <- data.frame(
  method = "Min p-value",
  optimal_cutpoint = cutoff_minp,

  sensitivity = sensitivity,
  sensitivity_low = sensitivity_low,
  sensitivity_high = sensitivity_high,

  specificity = specificity,
  specificity_low = specificity_low,
  specificity_high = specificity_high,

  ppv = ppv,
  ppv_low = ppv_low,
  ppv_high = ppv_high,

  npv = npv,
  npv_low = npv_low,
  npv_high = npv_high,

  accuracy = acc,
  accuracy_low = ci_acc[1],
```

```
  accuracy_high = ci_acc[2]
)

metrics_t<-metrics_minp %>%
   select(method, optimal_cutpoint, sensitivity, sensitivity_low,
      ↪ sensitivity_high,
          specificity, specificity_low, specificity_high, ppv, ppv_low
             ↪ , ppv_high, npv, npv_low, npv_high, accuracy, accuracy
             ↪ _low, accuracy_high) %>%
   gt() %>%
   tab_header(title = "Optimal Cut-off Point Metrics - Min p-value")
      ↪ %>%
   fmt_number(
     columns = where(is.numeric),
     decimals = 3
   )

```


Comparison table

`r tab_cap("Comparison of Cut-off Methods: Null scenario")`

```{r echo=FALSE, message=FALSE, warning=FALSE}

tab_cutoffs <- dplyr::bind_rows(
  metrics_boot_expect |> mutate(method = "New approach: Expected value
     ↪ "),
  metrics_boot_maxdens |> mutate(method = "New approach: Maximum
     ↪ density"),
  metrics_youden |> mutate(method = "Youden's J statistic"),
  metrics_eucl |> mutate(method = "Euclidean distance"),
  metrics_mprod |> mutate(method = "Max. product Se*Sp"),
  metrics_iu |> mutate(method = "Index of Union"),
  metrics_cost |> mutate(method = "Cost approach"),
  metrics_misscost |> mutate(method = "Misclassification cost"),
```

```
  metrics_dor |> mutate(method = "DOR"),
  metrics_minp |> mutate(method = "Min p-value")
)


# Format CI
format_ci <- function(value, low, high) {
  if (is.na(low) | is.na(high)) {
    return(sprintf("%.2f", value))
  } else {
    return(sprintf("%.2f [%.2f; %.2f]", value, low, high))
  }
}

# Metrics [CI]
tab_cutoffs_formatted <- tab_cutoffs |>
  mutate(
    Cutoff = optimal_cutpoint,
    Sensitivity = mapply(format_ci, sensitivity, sensitivity_low,
      ↪ sensitivity_high),
    Specificity = mapply(format_ci, specificity, specificity_low,
      ↪ specificity_high),
    PPV = mapply(format_ci, ppv, ppv_low, ppv_high),
    NPV = mapply(format_ci, npv, npv_low, npv_high),
    Accuracy = mapply(format_ci, accuracy, accuracy_low, accuracy_high
  )) |>
  select(
    Method = method,
    Cutoff,
    Sensitivity,
    Specificity,
    PPV,
    NPV,
    Accuracy
  )


# Final comparison table
library(gt)
```

```
tab_cutoffs_formatted |>
  gt()|>
  tab_header(
    title = md("**Comparison of Cut-off Methods**")
  ) |>
  cols_label(
    Method = "Method",
    Cutoff = "Cut-off point",
    Sensitivity = "Sensitivity",
    Specificity = "Specificity",
    PPV = "PPV",
    NPV = "NPV",
    Accuracy = "Accuracy"
  ) |>
  tab_style(
    style = cell_text(weight = "bold"),
    locations = cells_column_labels(everything())
  ) |>
  tab_style(
    style = cell_text(weight = "bold"),
    locations = cells_body(columns = Method)
  )

# Save cutoffs
# method_df <- data.frame(Method = tab_cutoffs_formatted$Method)
# save(method_df, file = "method_df_null_v2.Rdata")

load("method_df_null_v2.Rdata")

df <- tab_cutoffs_formatted |> dplyr::distinct(Method, .keep_all = T)
   ↪ |> dplyr::select(Cutoff)

method_df <- method_df|> dplyr::bind_cols(df)
rownames(method_df)<-NULL
save(method_df, file = "method_df_null_v2.Rdata")
```

Appendix B

Simulation reports

Statistical report TFM Lucia Blanc

AUTHORS
Research team - Lucia Blanc
UOC
Product - Statistical report
Biostatistics Research Unit HUGTiP-IGTP - Lucia Blanc

PUBLISHED
May 23, 2025

1 Results for Null scenario

1.1 Comparison table

Table 1: Comparison of Cut-off Methods: Null scenario

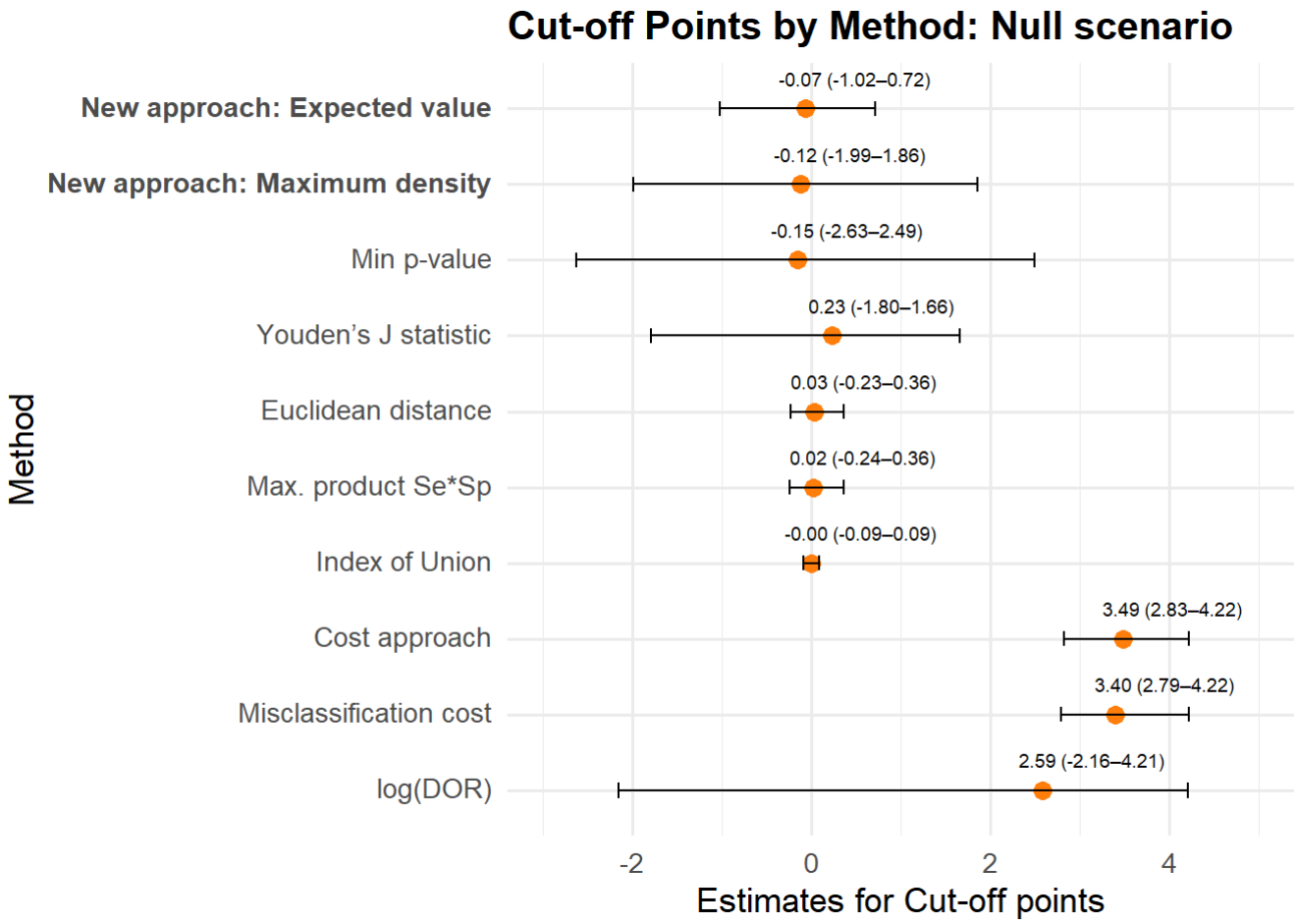
Comparison of Cut-off Methods: Null scenario

Method	Cut-off point [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	PPV [95% CI]	NPV [95% CI]	Accuracy [95% CI]
New approach: Expected value	-0.07 [-1.02; 0.72]	0.53 [0.47; 0.60]	0.47 [0.45; 0.49]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.47 [0.46; 0.49]
New approach: Maximum density	-0.12 [-1.99; 1.86]	0.54 [0.49; 0.61]	0.45 [0.43; 0.47]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.46 [0.44; 0.48]
Min p-value	-0.15 [-2.63; 2.49]	0.56 [0.50; 0.62]	0.44 [0.42; 0.45]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.45 [0.43; 0.47]
Youden's J statistic	0.23 [-1.80; 1.66]	0.41 [0.35; 0.49]	0.59 [0.57; 0.61]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.57 [0.56; 0.59]
Euclidean distance	0.03 [-0.23; 0.36]	0.48 [0.42; 0.55]	0.51 [0.50; 0.53]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.51 [0.49; 0.53]
Max. product Se*Sp	0.02 [-0.24; 0.36]	0.49 [0.43; 0.55]	0.51 [0.49; 0.53]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.51 [0.49; 0.52]
Index of Union	-0.00 [-0.09; 0.09]	0.50 [0.44; 0.57]	0.50 [0.48; 0.52]	0.10 [0.08; 0.12]	0.90 [0.88; 0.92]	0.50 [0.48; 0.51]
Cost approach	3.49 [2.83; 4.22]	0.00 [0.00; 0.00]	1.00 [1.00; 1.00]	0.02 [0.00; 0.19]	0.90 [0.89; 0.91]	0.90 [0.89; 0.91]
Misclassification cost	3.40 [2.79; 4.22]	0.00 [0.00; 0.00]	1.00 [1.00; 1.00]	0.08 [0.00; 0.60]	0.90 [0.89; 0.91]	0.90 [0.89; 0.91]
log(DOR)	2.59 [-2.16; 4.21]	0.00 [0.00; 0.01]	0.99 [0.99; 1.00]	0.09 [0.00; 0.27]	0.90 [0.89; 0.91]	0.90 [0.88; 0.91]

1.2 Comparison plots

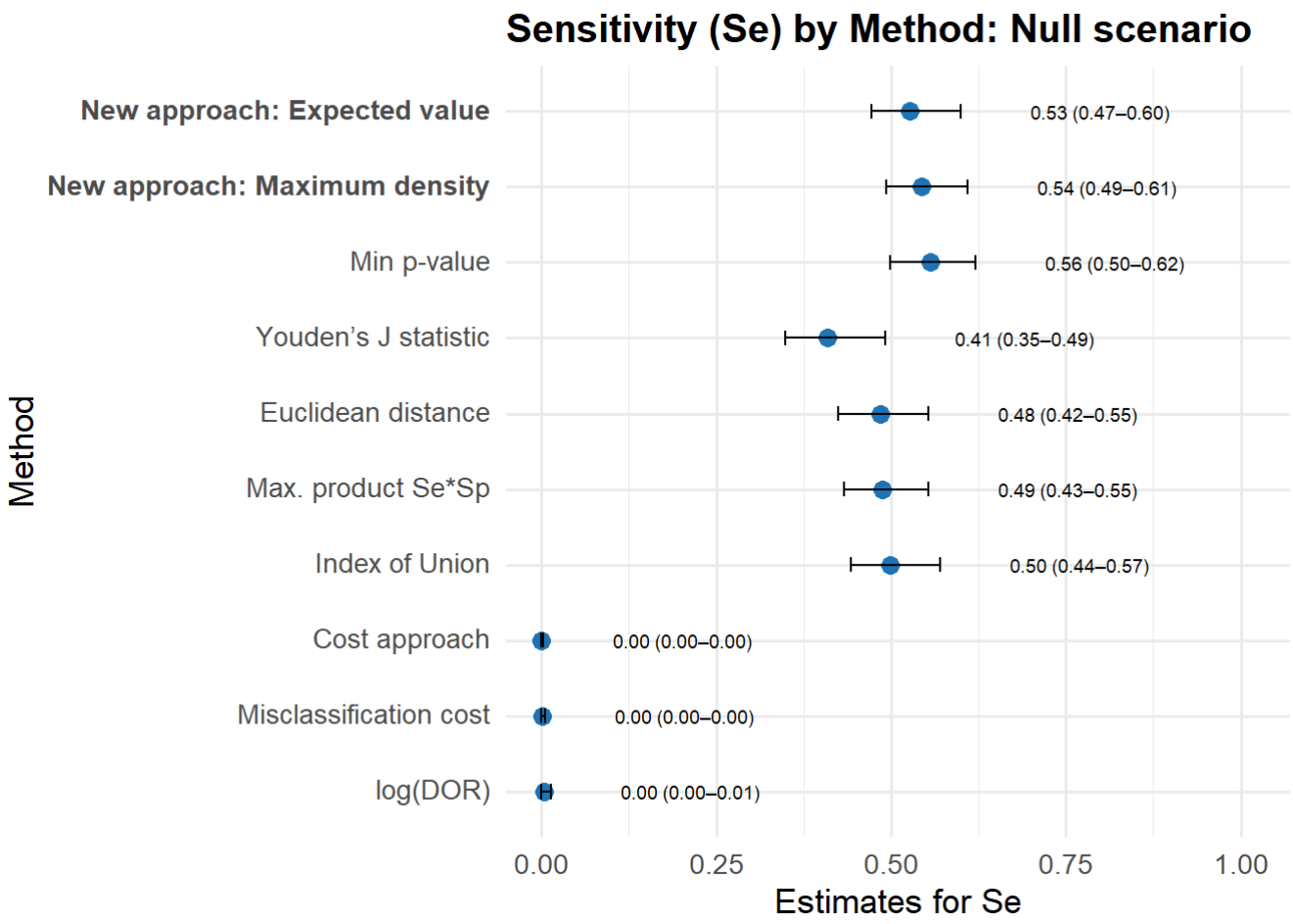
1.2.1 Cut-off points

Figure 1: Comparison of Cut-off point for each method



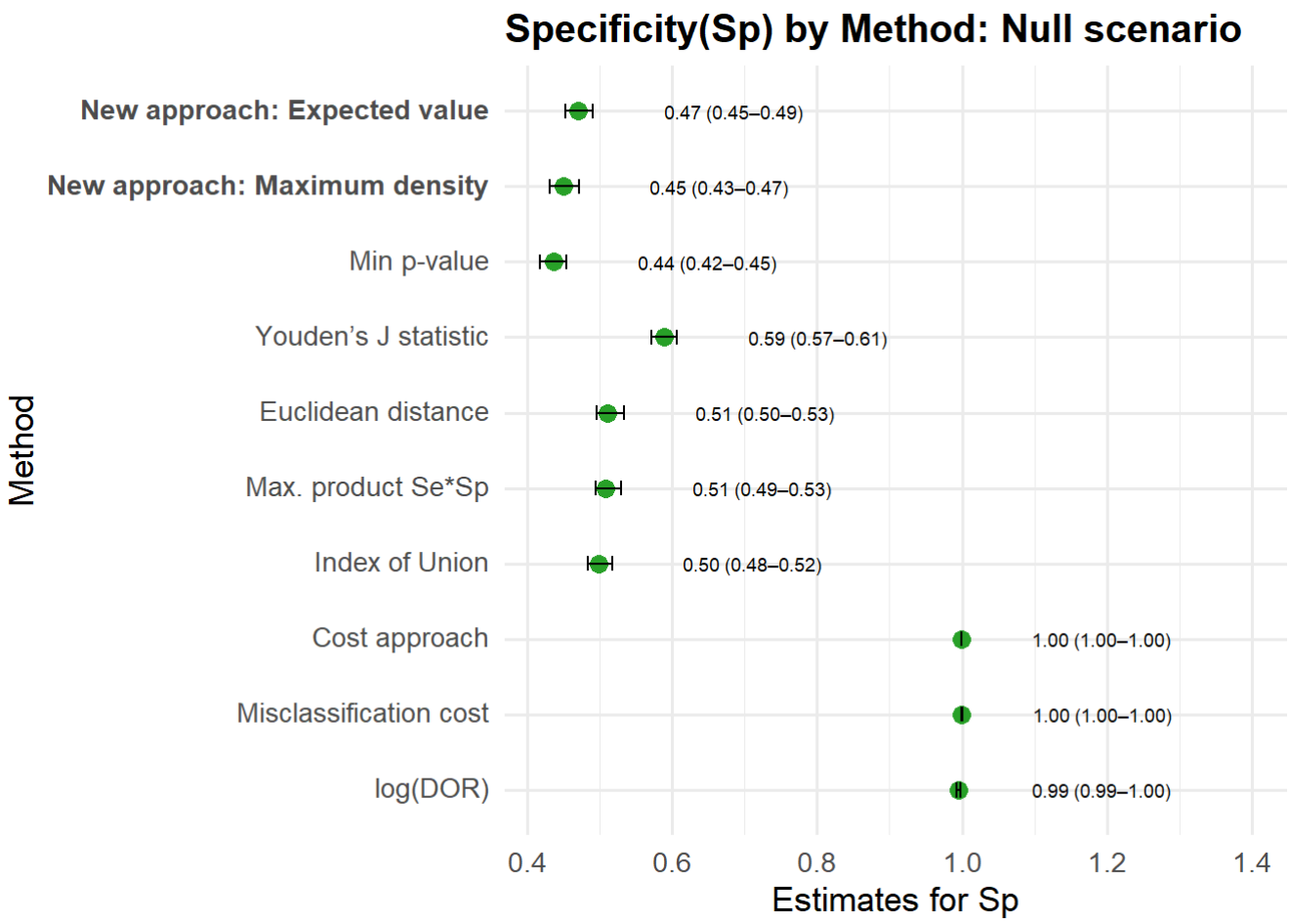
1.2.2 Sensitivity

Figure 2: Comparison of Sensitivity for each method



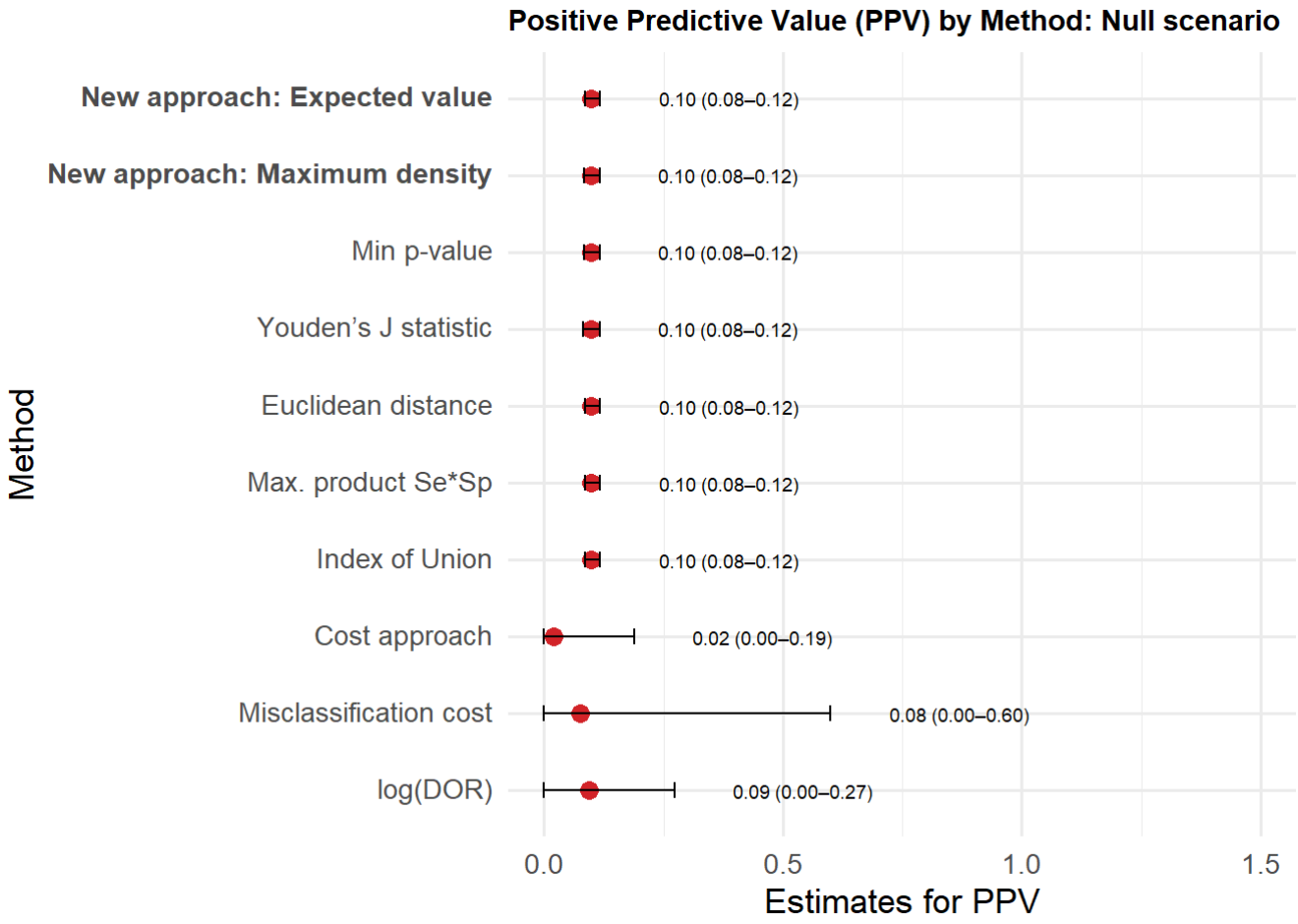
1.2.3 Specificity

Figure 3: Comparison of Specificity for each method



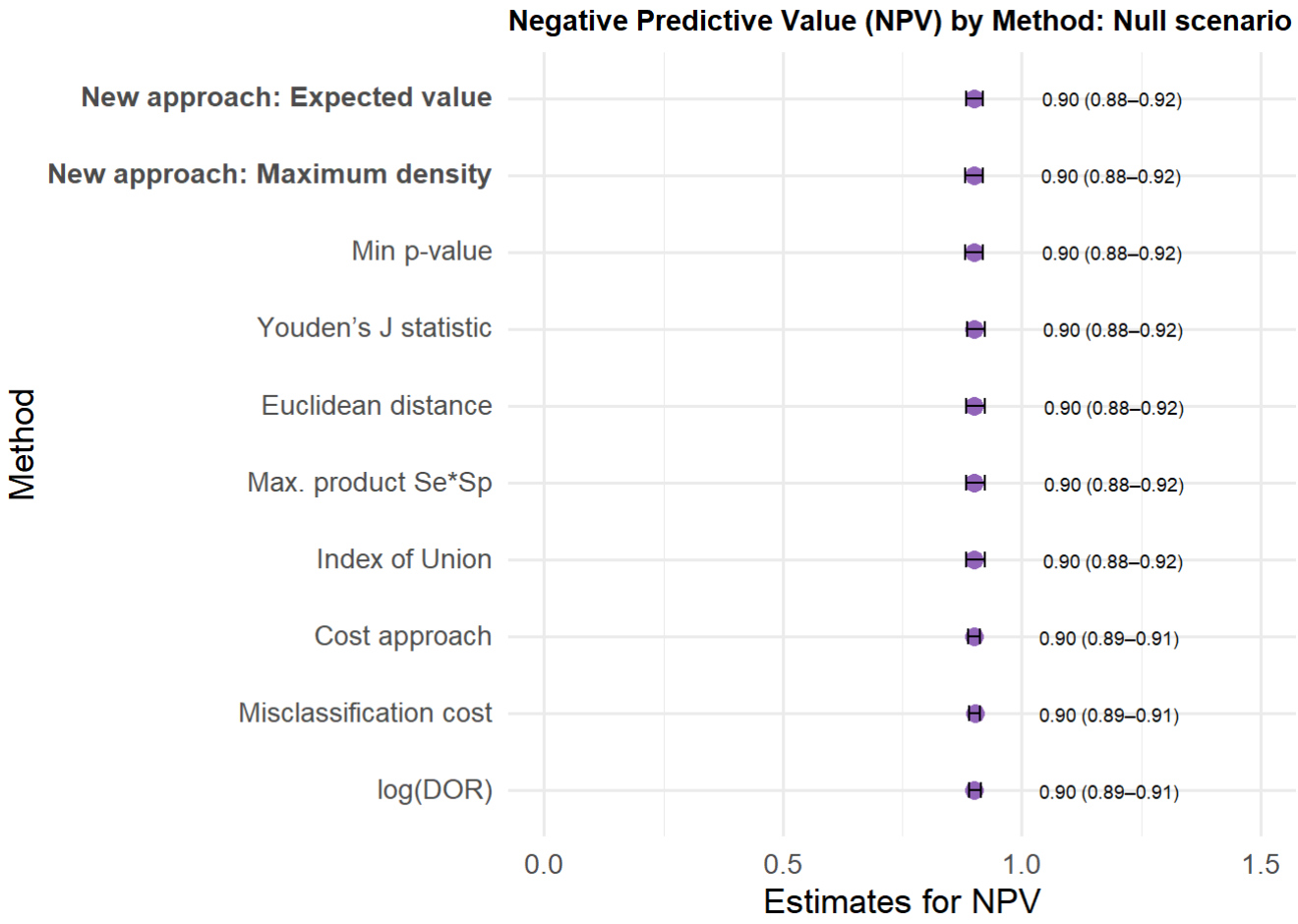
1.2.4 Positive Predictive Value (PPV)

Figure 4: Comparison of PPV for each method



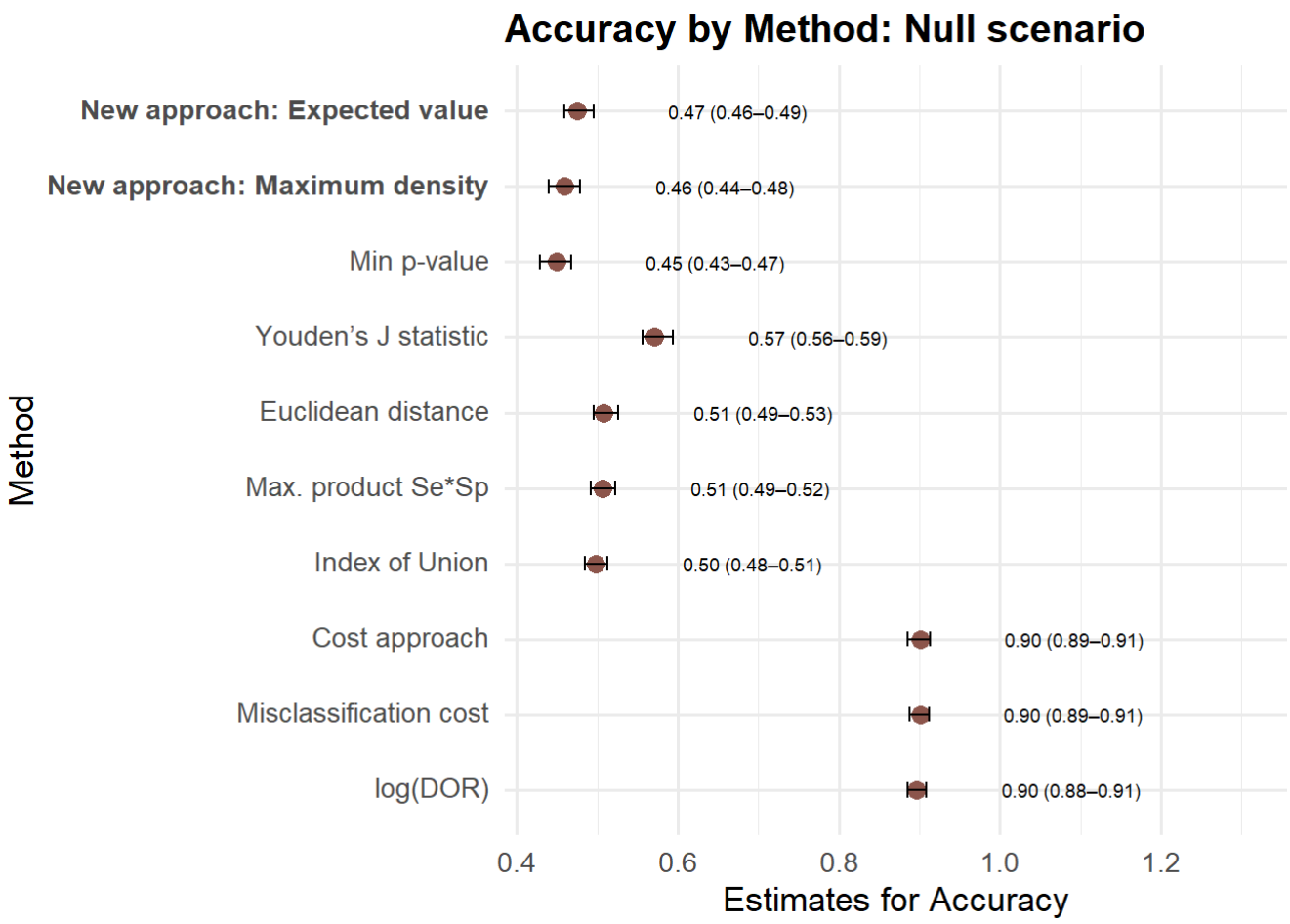
1.2.5 Negative Predictive Value (NPV)

Figure 5: Comparison of NPV for each method



1.2.6 Accuracy

Figure 6: Comparison of Accuracy for each method



Statistical report TFM Lucia Blanc

AUTHORS
Research team - Lucia Blanc
UOC
Product - Statistical report
Biostatistics Research Unit HUGTiP-IGTP - Lucia Blanc

PUBLISHED
May 23, 2025

1 Results for Small scenario

1.1 Comparison table

Table 1: Comparison of Cut-off Methods: Small scenario

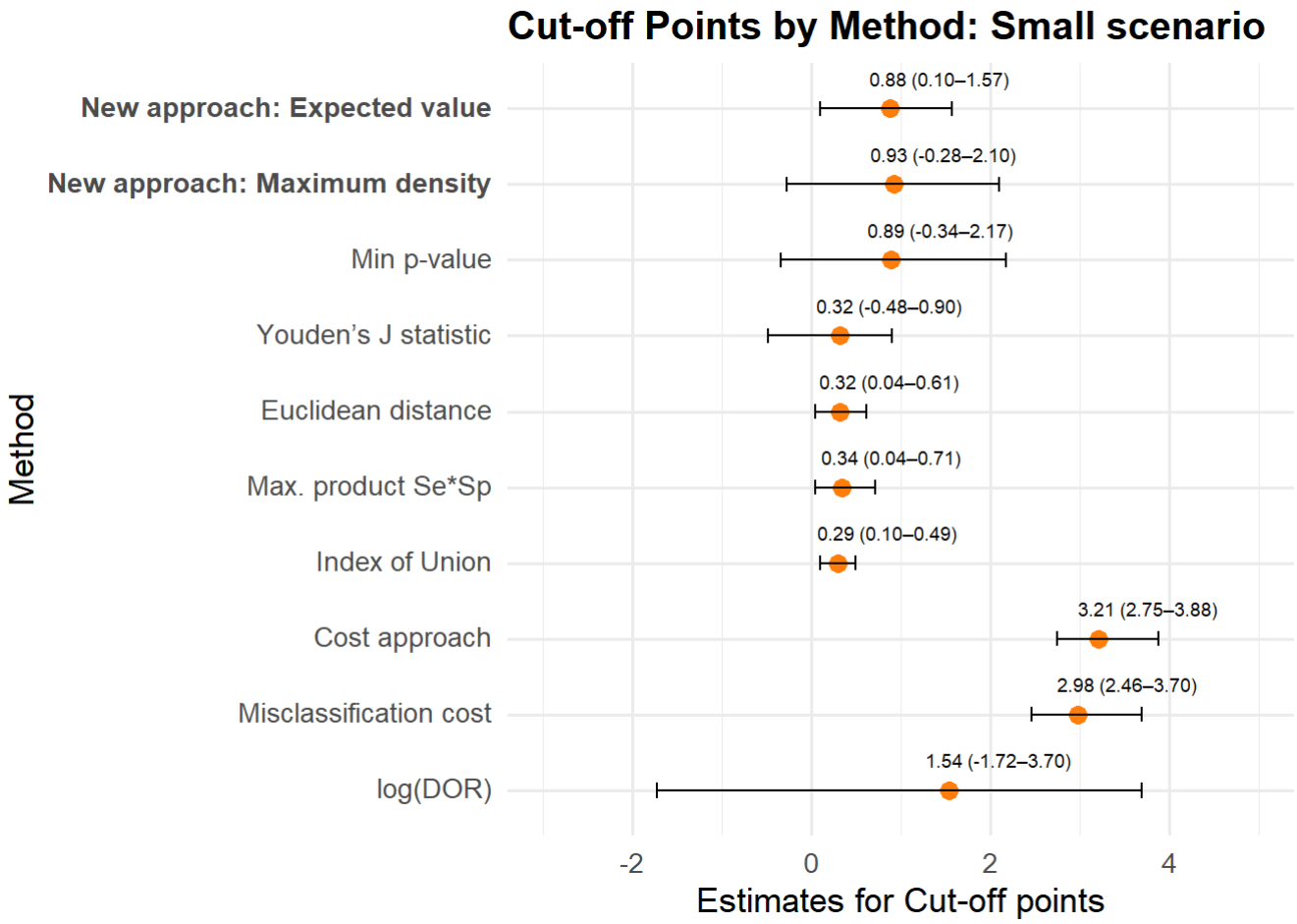
Comparison of Cut-off Methods: Small scenario

Method	Cut-off point [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	PPV [95% CI]	NPV [95% CI]	Accuracy [95% CI]
New approach: Expected value	0.88 [0.10; 1.57]	0.37 [0.28; 0.47]	0.81 [0.80; 0.84]	0.18 [0.13; 0.23]	0.92 [0.90; 0.94]	0.77 [0.74; 0.79]
New approach: Maximum density	0.93 [-0.28; 2.10]	0.35 [0.27; 0.46]	0.82 [0.80; 0.85]	0.18 [0.14; 0.23]	0.92 [0.90; 0.94]	0.78 [0.75; 0.80]
Min p-value	0.89 [-0.34; 2.17]	0.37 [0.28; 0.47]	0.81 [0.80; 0.84]	0.18 [0.13; 0.24]	0.92 [0.90; 0.94]	0.77 [0.75; 0.80]
Youden's J statistic	0.32 [-0.48; 0.90]	0.58 [0.48; 0.66]	0.62 [0.59; 0.64]	0.14 [0.11; 0.17]	0.93 [0.91; 0.94]	0.62 [0.60; 0.64]
Euclidean distance	0.32 [0.04; 0.61]	0.58 [0.48; 0.66]	0.62 [0.60; 0.64]	0.14 [0.11; 0.17]	0.93 [0.91; 0.94]	0.62 [0.60; 0.64]
Max. product Se*Sp	0.34 [0.04; 0.71]	0.57 [0.48; 0.65]	0.63 [0.61; 0.65]	0.14 [0.11; 0.17]	0.93 [0.91; 0.95]	0.62 [0.61; 0.64]
Index of Union	0.29 [0.10; 0.49]	0.59 [0.50; 0.68]	0.61 [0.58; 0.63]	0.14 [0.11; 0.17]	0.93 [0.91; 0.95]	0.61 [0.59; 0.63]
Cost approach	3.21 [2.75; 3.88]	0.01 [0.00; 0.02]	1.00 [1.00; 1.00]	0.45 [0.00; 1.00]	0.90 [0.88; 0.92]	0.90 [0.88; 0.92]
Misclassification cost	2.98 [2.46; 3.70]	0.01 [0.00; 0.02]	1.00 [1.00; 1.00]	0.37 [0.00; 1.00]	0.90 [0.88; 0.92]	0.90 [0.88; 0.92]
log(DOR)	1.54 [-1.72; 3.70]	0.15 [0.09; 0.21]	0.94 [0.92; 0.95]	0.22 [0.13; 0.30]	0.91 [0.89; 0.93]	0.86 [0.83; 0.88]

1.2 Comparison plots

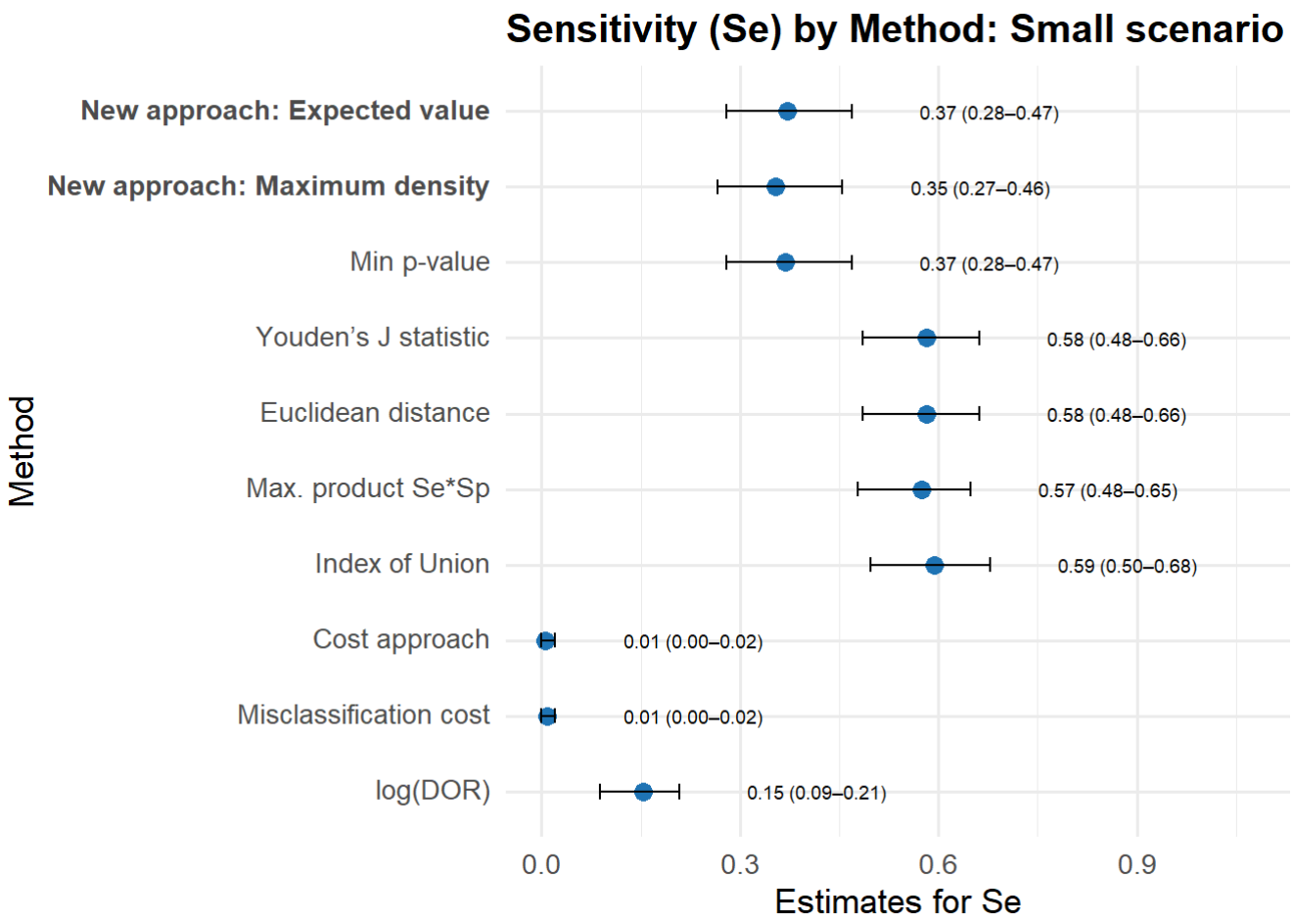
1.2.1 Cut-off points

Figure 1: Comparison of Cut-off point for each method



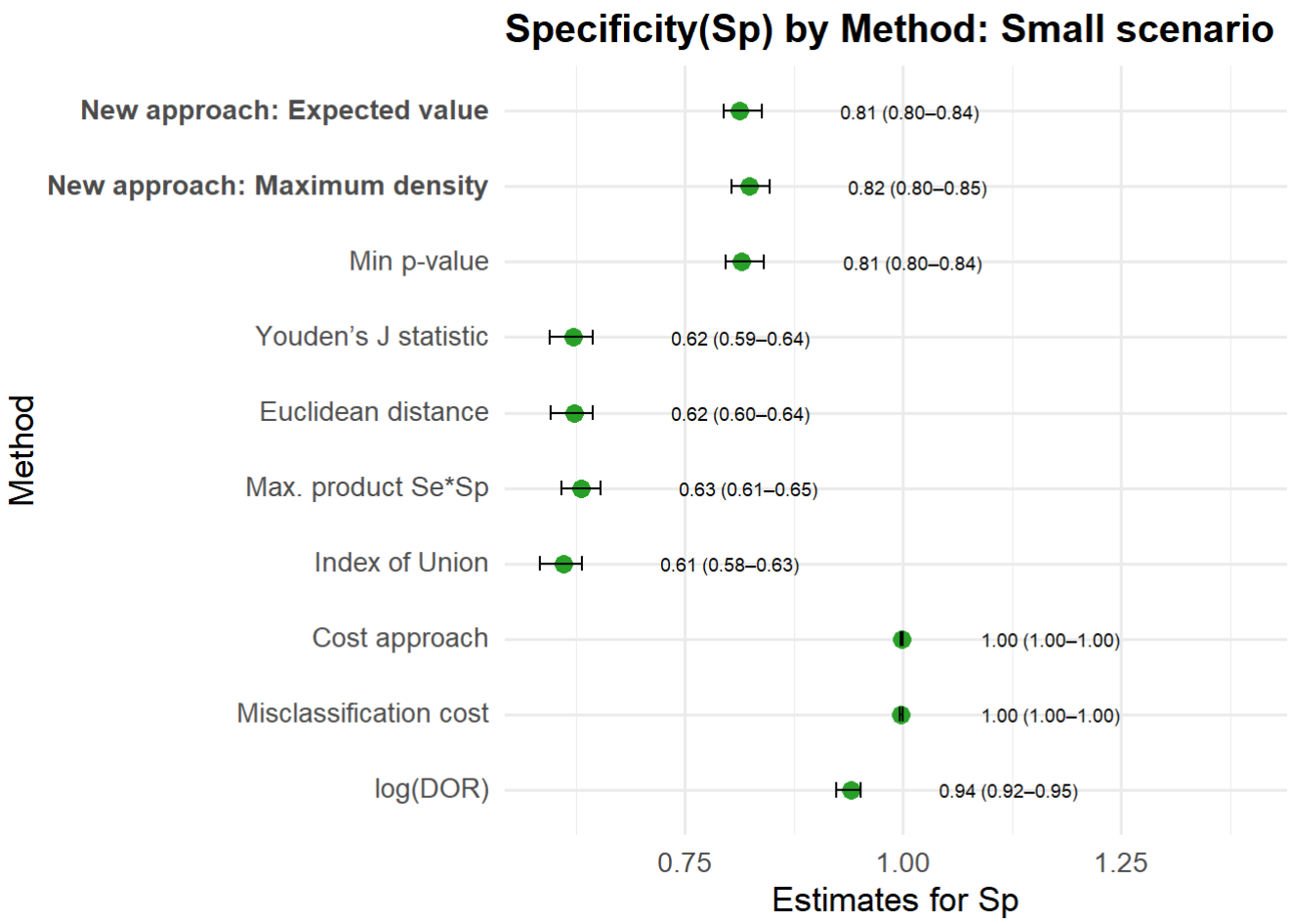
1.2.2 Sensitivity

Figure 2: Comparison of Sensitivity for each method



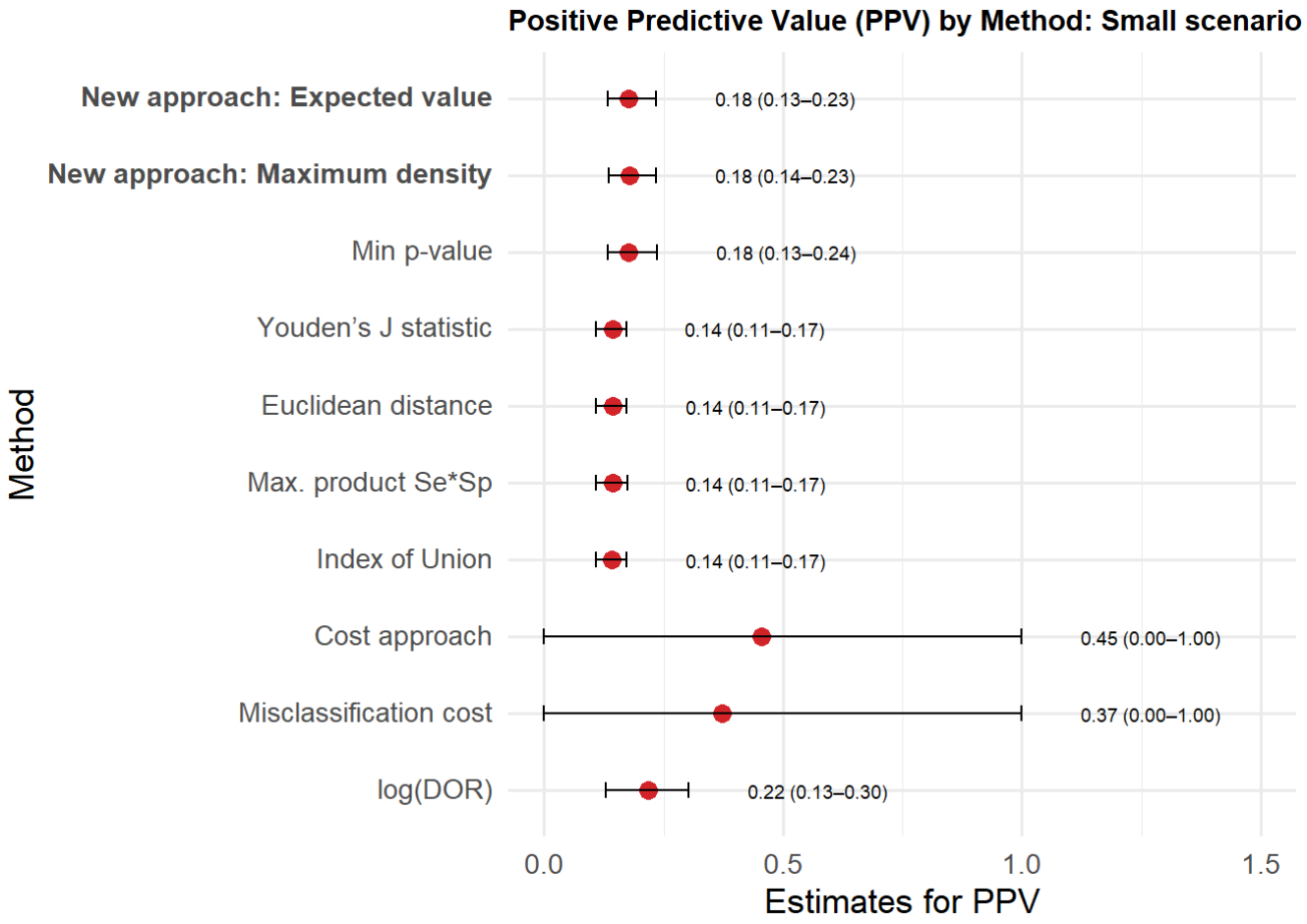
1.2.3 Specificity

Figure 3: Comparison of Specificity for each method



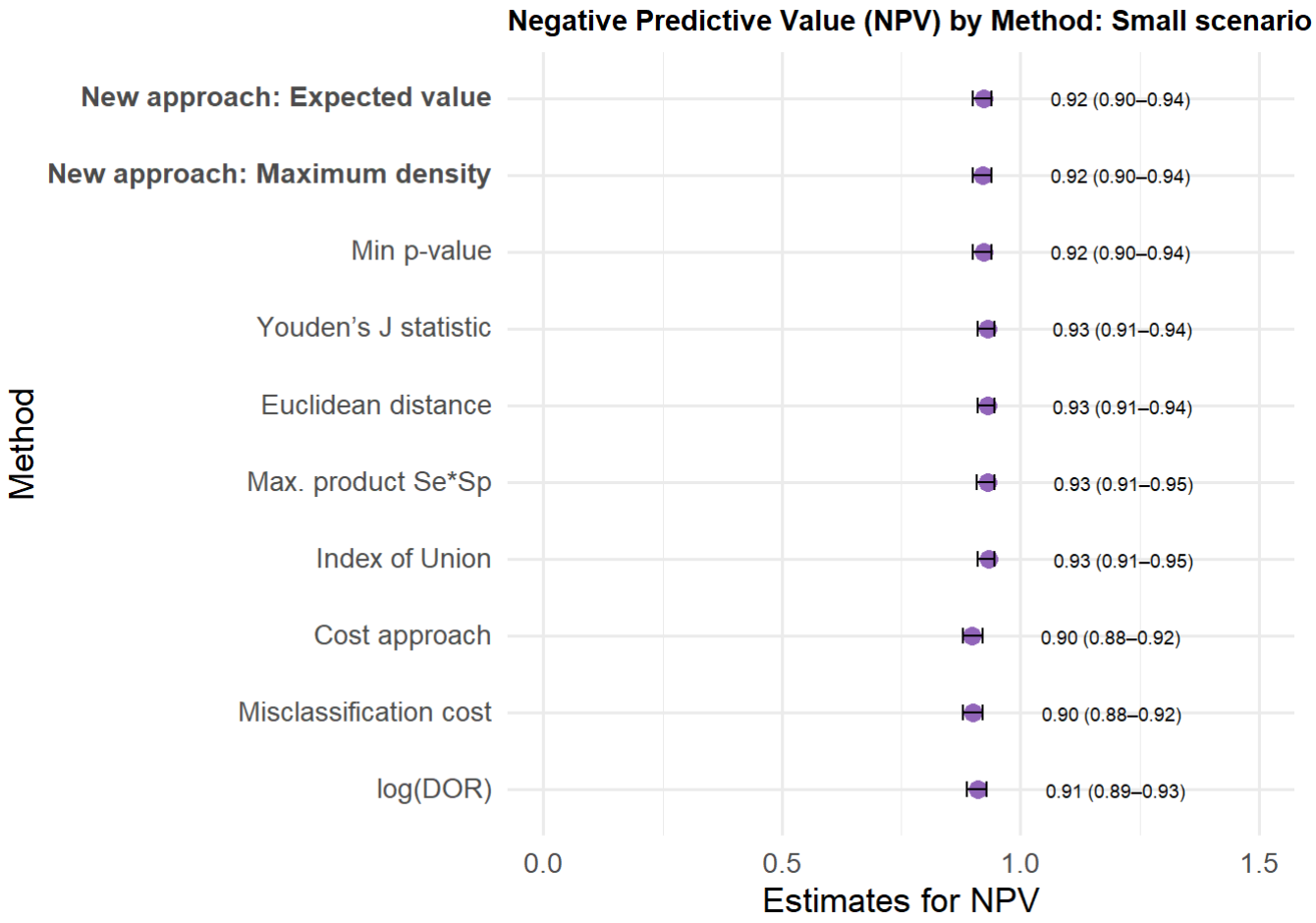
1.2.4 Positive Predictive Value (PPV)

Figure 4: Comparison of PPV for each method



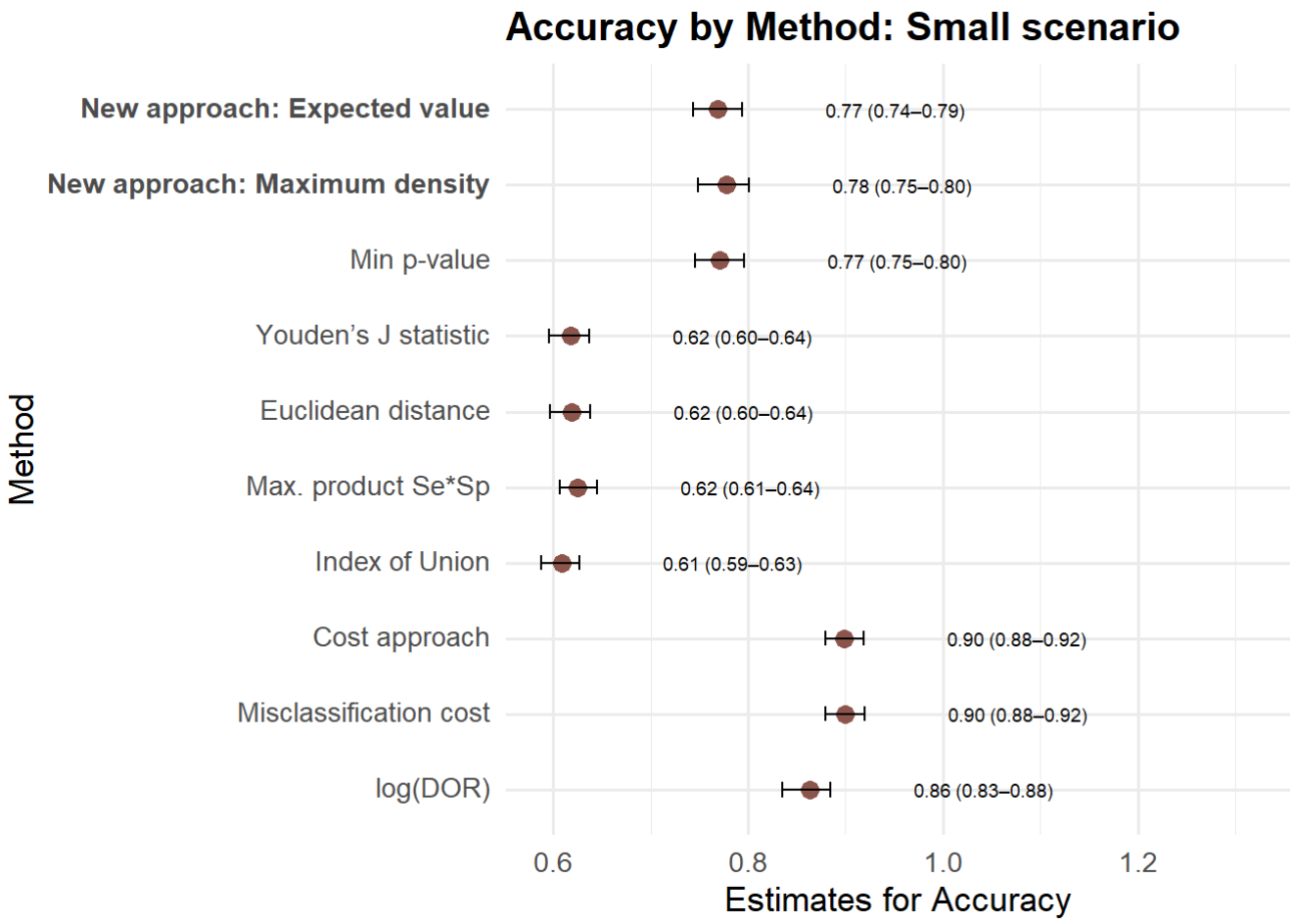
1.2.5 Negative Predictive Value (NPV)

Figure 5: Comparison of NPV for each method



1.2.6 Accuracy

Figure 6: Comparison of Accuracy for each method



Statistical report TFM Lucia Blanc

AUTHORS
Research team - Lucia Blanc
UOC
Product - Statistical report
Biostatistics Research Unit HUGTiP-IGTP - Lucia Blanc

PUBLISHED
May 23, 2025

1 Results for Moderate scenario

1.1 Comparison table

Table 1: Comparison of Cut-off Methods: Moderate scenario

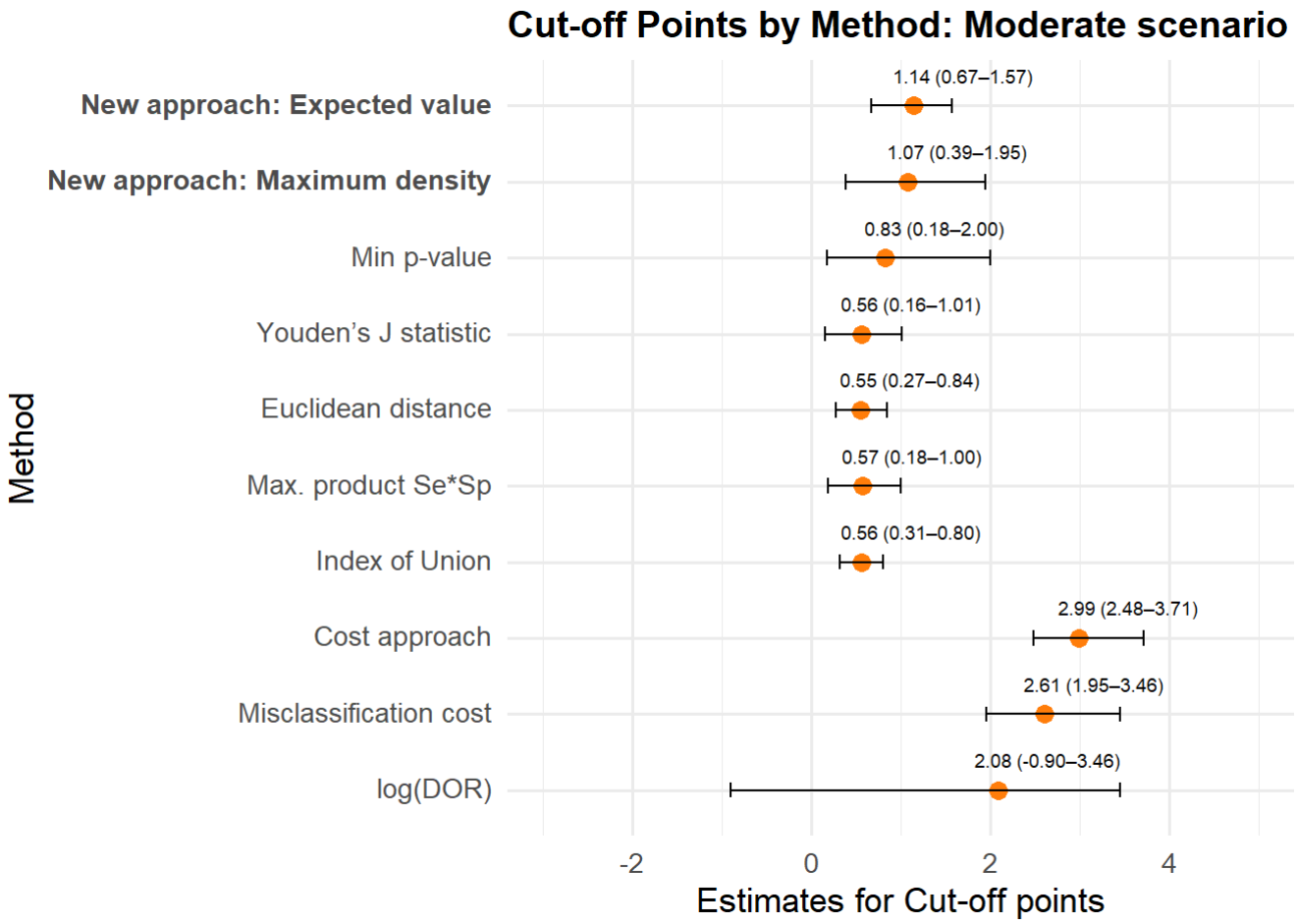
Comparison of Cut-off Methods: Moderate scenario

Method	Cut-off point [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	PPV [95% CI]	NPV [95% CI]	Accuracy [95% CI]
New approach: Expected value	1.14 [0.67; 1.57]	0.43 [0.30; 0.58]	0.87 [0.84; 0.91]	0.28 [0.19; 0.40]	0.93 [0.91; 0.96]	0.83 [0.80; 0.86]
New approach: Maximum density	1.07 [0.39; 1.95]	0.46 [0.33; 0.62]	0.86 [0.83; 0.89]	0.27 [0.18; 0.38]	0.94 [0.91; 0.96]	0.82 [0.78; 0.85]
Min p-value	0.83 [0.18; 2.00]	0.57 [0.44; 0.70]	0.80 [0.77; 0.83]	0.24 [0.19; 0.32]	0.94 [0.91; 0.96]	0.77 [0.74; 0.81]
Youden's J statistic	0.56 [0.16; 1.01]	0.68 [0.55; 0.82]	0.71 [0.68; 0.75]	0.21 [0.17; 0.28]	0.95 [0.92; 0.97]	0.71 [0.68; 0.74]
Euclidean distance	0.55 [0.27; 0.84]	0.69 [0.57; 0.82]	0.71 [0.67; 0.74]	0.21 [0.17; 0.28]	0.95 [0.92; 0.97]	0.71 [0.68; 0.74]
Max. product Se*Sp	0.57 [0.18; 1.00]	0.68 [0.55; 0.82]	0.72 [0.68; 0.75]	0.21 [0.17; 0.28]	0.95 [0.92; 0.97]	0.71 [0.68; 0.74]
Index of Union	0.56 [0.31; 0.80]	0.68 [0.55; 0.82]	0.71 [0.68; 0.75]	0.21 [0.17; 0.28]	0.95 [0.92; 0.97]	0.71 [0.68; 0.74]
Cost approach	2.99 [2.48; 3.71]	0.03 [0.00; 0.07]	1.00 [0.99; 1.00]	0.70 [0.00; 1.00]	0.90 [0.87; 0.92]	0.90 [0.87; 0.92]
Misclassification cost	2.61 [1.95; 3.46]	0.06 [0.00; 0.12]	1.00 [0.99; 1.00]	0.59 [0.00; 1.00]	0.90 [0.87; 0.92]	0.90 [0.87; 0.92]
log(DOR)	2.08 [-0.90; 3.46]	0.14 [0.04; 0.24]	0.98 [0.97; 0.99]	0.45 [0.20; 0.67]	0.91 [0.88; 0.93]	0.90 [0.87; 0.92]

1.2 Comparison plots

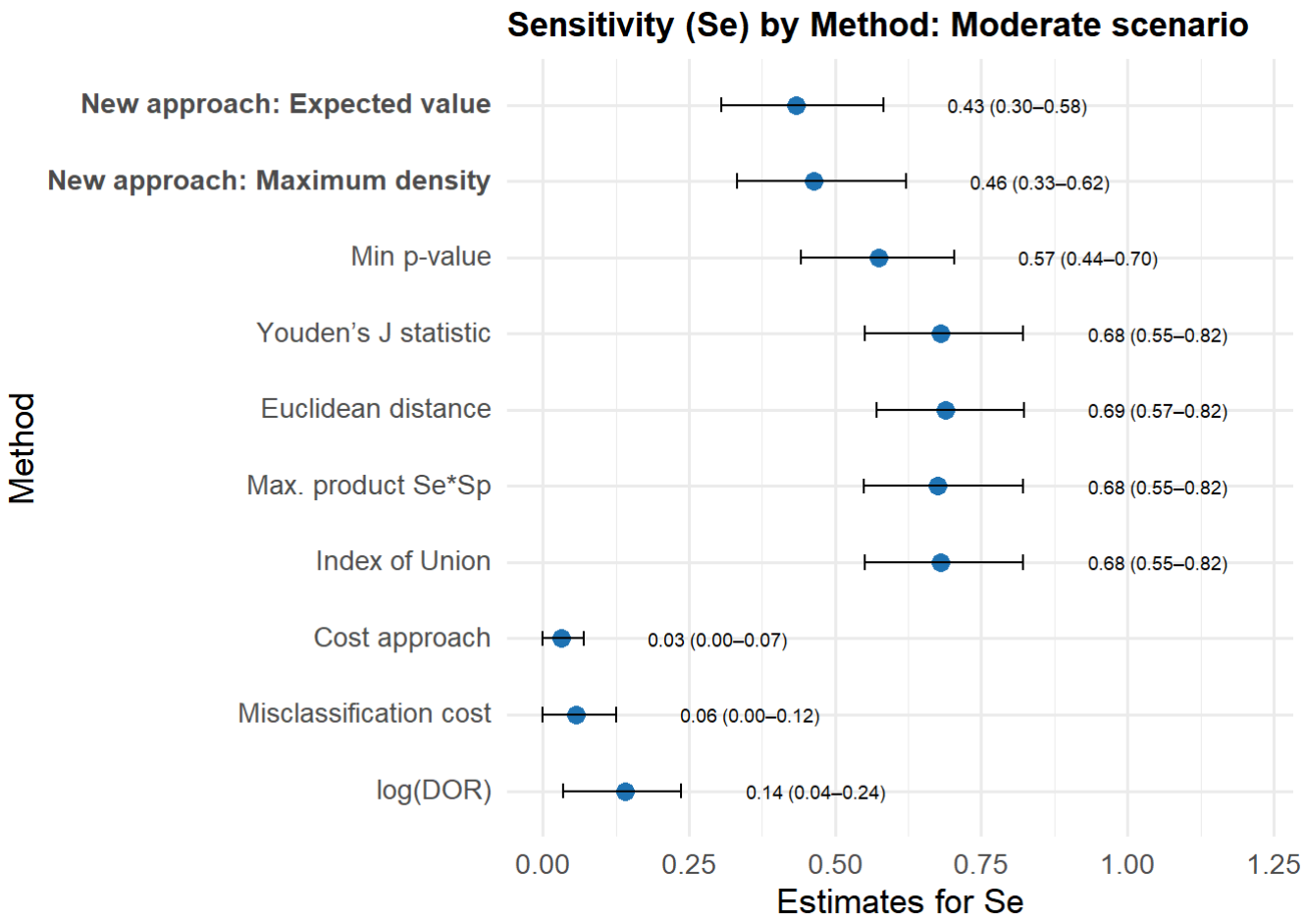
1.2.1 Cut-off points

Figure 1: Comparison of Cut-off point for each method



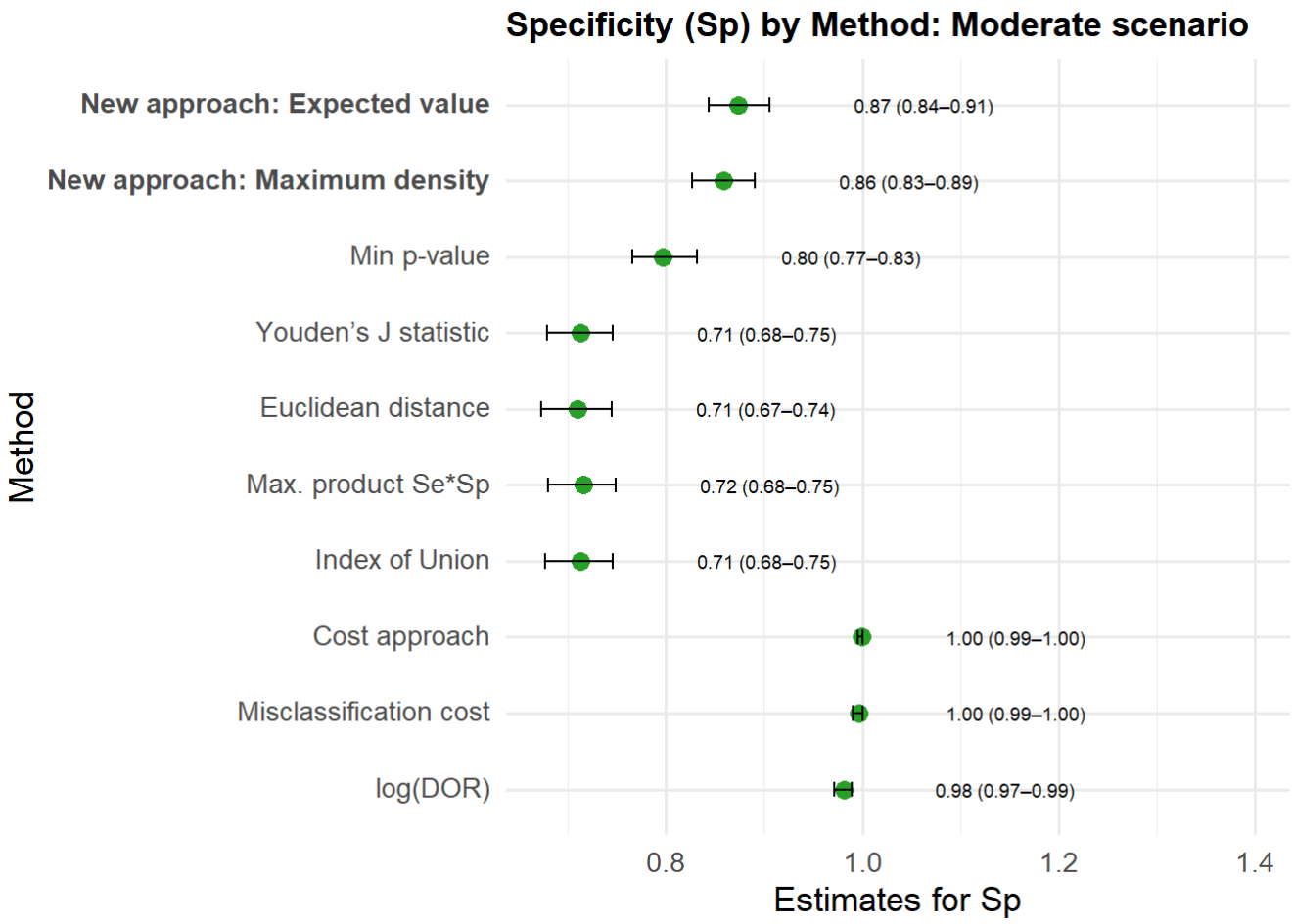
1.2.2 Sensitivity

Figure 2: Comparison of Sensitivity for each method



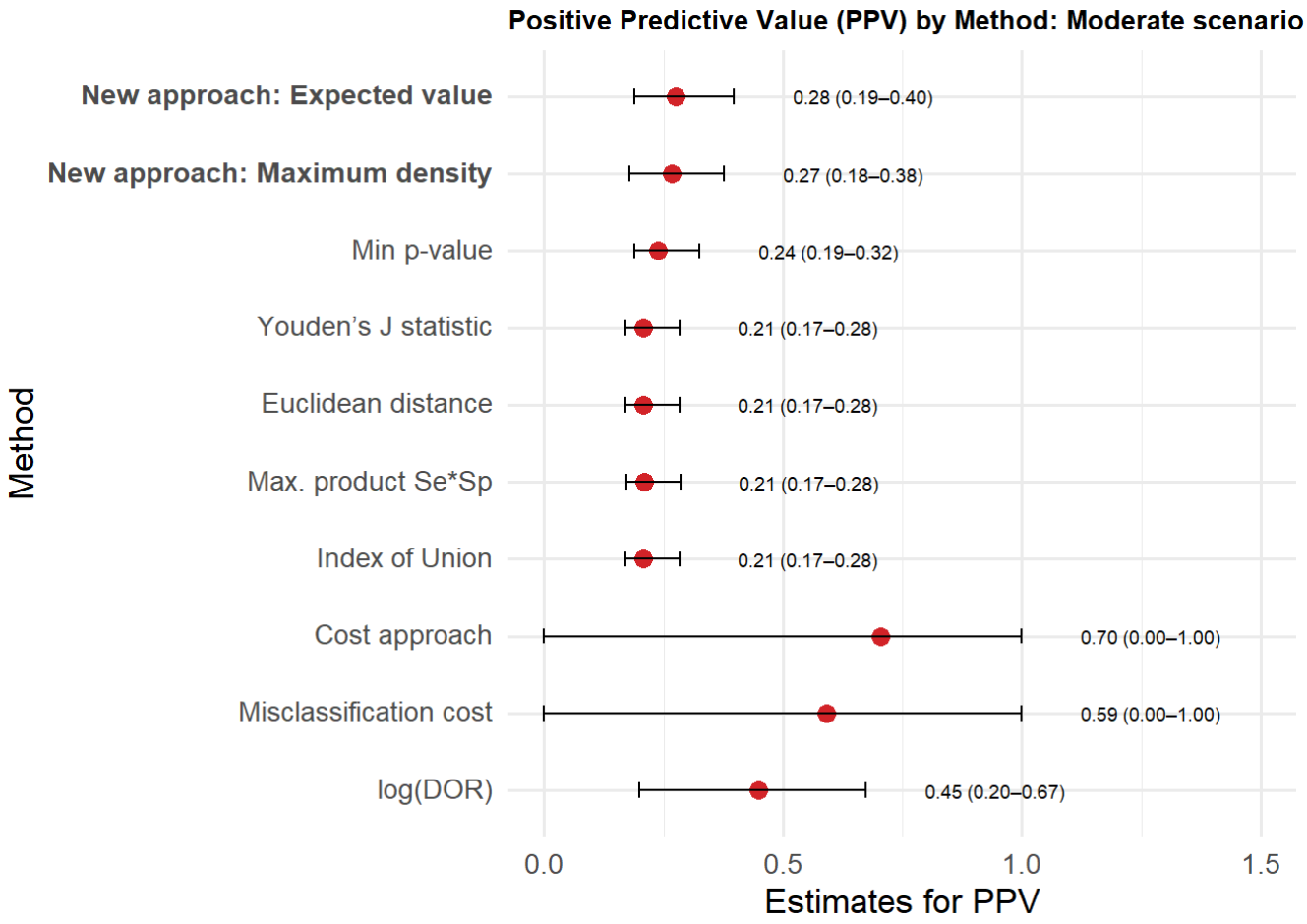
1.2.3 Specificity

Figure 3: Comparison of Specificity for each method



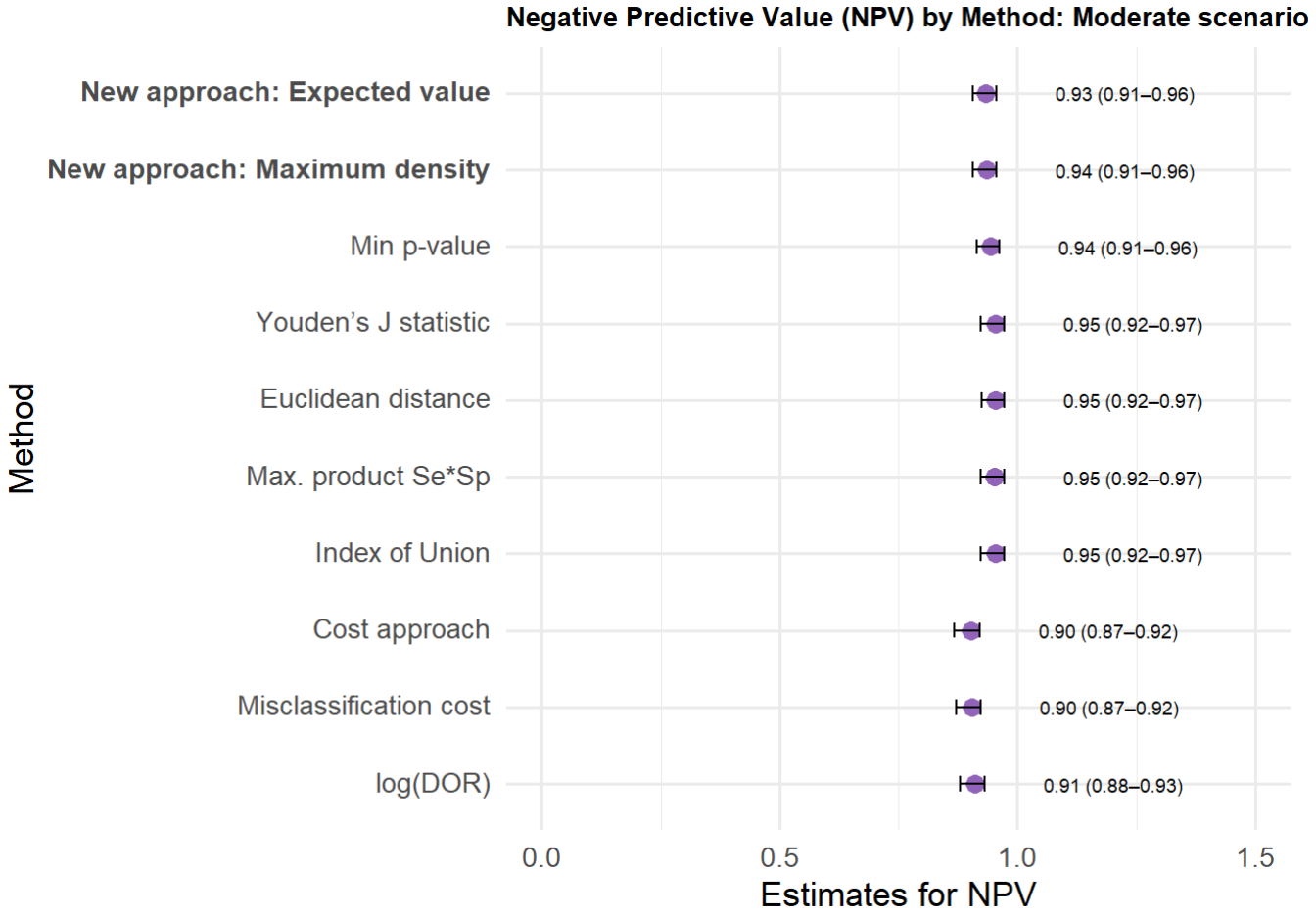
1.2.4 Positive Predictive Value (PPV)

Figure 4: Comparison of PPV for each method



1.2.5 Negative Predictive Value (NPV)

Figure 5: Comparison of NPV for each method



1.2.6 Accuracy

Figure 6: Comparison of Accuracy for each method