Received: 6 January 2021 | Revised: 18 April 2022 | Accepted: 22 April 2022

DOI: 10.1002/jrsm.1579

RESEARCH ARTICLE

Research Synthesis Methods WILEY

A semiparametric approach for meta-analysis of diagnostic accuracy studies with multiple cut-offs

Cornelia Frömke[1] | **Mathia Kirstein**[1] | **Antonia Zapf**[2]

[1]Department of Information and Communication, Faculty for Media, Information and Design, University of Applied Sciences and Arts Hannover, Hannover, Germany

[2]Department of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Correspondence
Cornelia Frömke, Department of Information and Communication, Faculty for Media, Information and Design, University of Applied Sciences and Arts Hannover, Hannover, Germany.
Email: cornelia.froemke@hs-hannover.de

Abstract

The accuracy of a diagnostic test is often expressed using a pair of measures: sensitivity (proportion of test positives among all individuals with target condition) and specificity (proportion of test negatives among all individuals without target condition). If the outcome of a diagnostic test is binary, results from different studies can easily be summarized in a meta-analysis. However, if the diagnostic test is based on a discrete or continuous measure (e.g., a biomarker), several cut-offs within one study as well as among different studies are published. Instead of taking all information of the cut-offs into account in the meta-analysis, a single cut-off per study is often selected arbitrarily for the analysis, even though there are statistical methods for the incorporation of several cut-offs. For these methods, distributional assumptions have to be met and/or the models may not converge when specific data structures occur. We propose a semiparametric approach to overcome both problems. Our simulation study shows that the diagnostic accuracy is under-estimated, although this underestimation in sensitivity and specificity is relatively small. The comparative approach of Steinhauser et al. is better in terms of coverage probability, but may lead to convergence problems. In addition to the simulation results, we illustrate the application of the semiparametric approach using a published meta-analysis for a diagnostic test differentiating between bacterial and viral meningitis in children.

KEYWORDS

diagnostic accuracy studies, meta-analysis, multiple cut-offs, semiparametric

Highlights

What is already known?

- In recent years, various approaches for the meta-analysis of diagnostic accuracy studies with multiple cut-off values have been proposed.
- However, these approaches require certain assumptions regarding the underlying distribution as well as the correlation structure and may have convergence problems.

What is new?

- A semiparametric approach is proposed for the meta-analysis of diagnostic accuracy studies with multiple cut-off values.
- The statistical properties of the semiparametric approach are investigated and compared with an alternative approach

Potential impact for RSM readers?

- The semiparametric is conservative and underestimates sensitivity and specificity slightly.
- The coverage probability is below the nominal level.
- In contrast to iterative approaches, the approach yields always estimators.

1 | INTRODUCTION

The aim of diagnostic accuracy studies is to survey the accuracy of a new diagnostic test differentiating among individuals suffering from the disease of interest (or general denoted as "with target condition") and individuals without the disease (general "without target condition"). Therefore, two primary endpoints are of interest: (I) sensitivity, being the proportion of test positives among all individuals with target condition and (II) specificity, being the proportion of test negatives among all individuals without target condition. If the outcome of the diagnostic test is dichotomous, the estimation of sensitivity and specificity results in a 2×2 table and in a calculation of percentages.

According to the center for evidence–based medicine (CEBM) levels of evidence,[1] it is not sufficient to test sensitivity and specificity in a single study. It is necessary to investigate the diagnostic test in several studies and to combine the results of these individual studies with a meta-analysis in a systematic review. In contrast to therapeutic studies which commonly focus on one primary endpoint, diagnostic accuracy studies are interested in the co-primary endpoints sensitivity and specificity.[2] In an individual study at a specific cut-off, sensitivity and specificity are independent because they are estimated in different populations. However, beyond the studies and cut-offs within a single study, sensitivity and specificity are correlated, and a meta-analysis should model this correlation if possible.[3] Cochrane recommends two methods for meta-analyses: the bivariate logistic random-effects model and the hierarchical summary receiver operating characteristic (HSROC) model.[4,5] In addition, various alternative approaches with their respective strengths and weaknesses can be found in the literature.[6–8]

However, the requirements for a diagnostic meta-analysis may become even more complex. The outcome of the diagnostic test may be a clinical score or a biomarker measured on a discrete or continuous scale. In this case, the choice of a specific cut-off value discriminating between "test positive" or "test negative" determines sensitivity and specificity. In such a case, results for sensitivity and specificity may be presented for multiple test values, that is, cut-offs can vary both within one study and between individual studies. Here, the bivariate logistic regression model as well as the HSROC model are of limited use: The bivariate model is only recommended for binary tests or when different studies report similar cut-offs. The HSROC model is appropriate for continuous diagnostic tests when the cut-offs of the individual studies vary, but it does not consider multiple cut-offs within individual studies or even full ROC curves, respectively.[4,9] In the last years, several authors proposed approaches for the meta-analysis of full ROC curves. Hoyer et al. (2018) point out limitations of such methods, such as the requirement of equal cut-offs among the studies or the impossibility to handle extreme values of sensitivity and specificity if the approach does not incorporate a continuity correction.[10] They propose a bivariate time-to-event model for interval-censored data for the analysis of full ROC curves as an alternative.[10] However, this approach may suffer from convergence problems. Further approaches have been published recently.[11,12] In the context of a data challenge, the semiparametric approach presented here has already been compared with three current approaches for a concrete examplary meta-analysis.[13] The characteristics of all four approaches were also compared regarding requirements, procedures, results, and specific advantages.

Here, we will present the semiparametric method for a meta-analysis of diagnostic accuracy studies considering the correlation structure of the two criteria as well as the multiple cut-offs within and among the individual studies. In a simulation study and for an examplary meta-analysis, we will compare the results with those of an alternative approach.

2 | METHODS

In this section, we introduce the semiparametric approach for meta-analysis of diagnostic accuracy studies with multiple cut-offs. This approach is an extension of the nonparametric analysis of diagnostic studies with repeated measures as proposed by Werner and Brunner (2007).[14] For details regarding the methods, we refer to the overview article by Brunner and Zapf (2014).[15]

Thus, in a study $s = 1,...,N$ we observe w_{si} diagnostic test results X_{sik} from individuals $k = 1,...,w_{si}$ without target condition $(i = 0)$ and or with target condition $(i = 1)$. Summarizing the number of individuals with or without target condition of all studies results in $w_{\cdot 0} = \sum_{s=1}^{N} w_{s0}$ individuals without target condition and $w_{\cdot 1} = \sum_{s=1}^{N} w_{s1}$ individuals with target condition and a total sample size of $w_{\cdot\cdot} = w_{\cdot 0} + w_{\cdot 1}$.

The estimation of the area under the curve (AUC), sensitivity, specificity and their corresponding confidence intervals require global as well as local ranked data. It should be noted that the estimators do not take the correlation structure among the individuals within one study into account. The global ranks include all observations from the N studies and both conditions and are denoted by

$$R_{sik} = \text{rank}^{(X_{sik})}.$$

Local ranks correspond to the ranked observations from all individuals with condition i from the N studies and we denote them by *asymptotic equivalence theorem* $R_{sik}^{(i)}$.

The AUC is equal to the relative effect

$$p = P(X_{s0k} < X_{s1k}) + \frac{1}{2}P(X_{s0k} = X_{s1k}),$$

which can be estimated by

$$\widehat{p} = \frac{1}{2} + \frac{1}{w_{\cdot\cdot}}\left(\overline{R}_{\cdot 1\cdot} - \overline{R}_{\cdot 0\cdot}\right). \tag{1}$$

Here, $\overline{R}_{\cdot i\cdot}$ denotes the mean of the global ranks for condition i:

$$\overline{R}_{\cdot i\cdot} = \frac{1}{w_{\cdot i}} \sum_{s=1}^{N} \sum_{k=1}^{w_{is}} R_{sik}.$$

Note that the meta-analysis estimator of the AUC is equal to the common estimator for single diagnostic accuracy studies with multiple observations per individual.[14,15]

Konietschke and Brunner (2009) showed that by means of the asymptotic equivalence theorem $\sqrt{N}(\widehat{p} - p) \tilde{N}(0, \sigma^2)$ holds.[16] Based on this theorem, an asymptotic range exceeding confidence interval for the AUC is given as

$$\widehat{p} \pm z_{1-\alpha/2} \cdot \sqrt{\frac{\widehat{\sigma}^2}{N}},$$

with the variance

$$\widehat{\sigma}^2 = \frac{N^2}{N-1}\left(\frac{1}{w_{\cdot 0} \cdot w_{\cdot 1}}\right)^2 \sum_{s=1}^{N} \left(D_{s1\cdot} - D_{s0\cdot} - \left(\frac{w_{s1}}{w_{\cdot 1}} D_{\cdot 1} - \frac{w_{s0}}{w_{\cdot 0}} D_{\cdot 0}\right)\right)^2,$$

where D_{sik} is the difference of the global and the local rank $R_{sik} - R_{sik}^{(i)}$, $D_{si\cdot} = \sum_{k=1}^{w_{si}} D_{sik}$ and $D_{\cdot i} = \sum_{s=1}^{N} D_{si\cdot}$.[14] This variance estimator $\widehat{\sigma}^2$ takes the correlation structure among the individuals within one study into account. Since it is known that for small sample sizes the confidence interval with the t-approximation has better statistical properties, we use the $(1-\alpha/2)$-quantile of the t-distribution with $df = N - 1$ degrees of freedom, such that

$$\widehat{p} \pm t_{df,1-\alpha/2} \cdot \sqrt{\frac{\widehat{\sigma}^2}{N}}. \tag{2}$$

We can define sensitivity and specificity in the same way as the AUC using pseudo samples at a specific cut-off.[17] For example, to estimate the sensitivity, we replace the observations of the individuals without target condition X_{s0k} by a one-point distribution at the chosen cut-off γ:

$$\Gamma(x) = \begin{cases} 0, x < \gamma \\ \frac{1}{2}, x = \gamma \\ 1, x > \gamma \end{cases}. \tag{3}$$

Afterwards, we rank the observations and estimate the sensitivity with Equation (1) and the corresponding confidence interval Equation (2). Likewise, replacing the observations of the individuals with target condition by a

one-point distribution results in the estimator of the specificity.

This nonparametric approach requires individual patient data. Unfortunately, authors of diagnostic accuracy studies often only present aggregated information, that is, sensitivity and specificity for selected cut-offs. Therefore, individual fictitious data needs to be generated based on distributional assumptions of the test results as well as the following information: We know the number of individuals with or without target condition less or greater than the study-specific cut-offs γ_{sm} with $m = 1,...,t_s$ for each study s,. The cut-offs of one individual study are ordered, such that $\gamma_{s1} < ... < \gamma_{st_s}$. The simplest fictitious data generation here would consist in one-point distributions for the respective areas. However, the obtained results were not satisfactory (see online appendix S1). But unlike the generation of pseudo data in Equation (3) to estimate the AUC, for the fictitious individual data to perform the meta-analysis, we are not limited to the one-point distribution to generate pseudo data and we use the uniform distribution in the following. This data generation also requires a minimum of model specifications. Between the cut-offs γ_{sm} and $\gamma_{s(m+1)}$ the fictitious data are generated following a uniform distribution. Below the first cut-off value, the data are drawn uniformly distributed from the interval from the smallest cut-off value used in the whole meta-analysis and the smallest study-specific cut-off value, and correspondingly, above the last cut-off value, the data are drawn uniformly distributed from the interval from the largest study-specific cut-off value to the largest cut-off value used in the whole meta-analysis.

3 | MOTIVATING EXAMPLE

To illustrate the semiparametric method, we use data from a meta-analysis summarizing study results to differentiate between bacterial and viral meningitis in children.[18] As bacterial meningitis in children is life-threatening, a rapid diagnosis is of utmost importance to increase the survival rate. The correct therapy depends on the type of disease-causing agent, which can be bacterial or viral. The gold standard to differentiate between a viral and a bacterial infection are invasive methods such as lumbar puncture and cerebrospinal fluid analysis. Several studies investigated procalcitonin (PCT) as a new biomarker replacing the common markers of bacterial infection. Basically, the serum level of the peptide PCT increases in the course of a bacterial infection, but not in the course of viral infection. Hence, the accuracy of the diagnosis of bacterial infection depends on the chosen cut-off in serum PCT level: levels larger than the cut-off indicate a bacterial infection (with target condition) whereas lower levels indicate a viral one (without target condition).[18] Table 1 lists the results from the individual studies for all available cut-offs as reported by Henry et al. (2016).[18] We found discrepancies in the study of Dubos et al. (2006)[19] and extracted the study results from the original paper. Henry et al. (2016)[18] included only the results for one cut-off per study in the meta-analysis; the corresponding rows are in boldface in Table 1.

4 | SIMULATION STUDY

In this section, we present results from a simulation study investigating the performance of the semiparametric approach. First of all, performance was measured by the AUC to assess the overall diagnostic accuracy. However, as the single statistic can describe different ROC curves, we investigated sensitivity and specificity for a set of cut-off values as well.

Following the recommendations of Burton et al. (2006),[20] we report the percentage bias, the mean squared error (MSE) as well as the empirical coverage probability for the AUC, sensitivity, and specificity to evaluate the statistical properties.

4.1 | Settings and data generation

We have chosen a baseline scenario motivated by the study of Henry et al. (2016)[18] and varied the individual parameters. In the exemplary meta-analysis, there were eight studies and the individual studies included between 30 and 200 patients with an average prevalence of 35% and between one and seven reported cut-off values.

The simulations for the baseline scenario consist of the following three steps.

- In a first step, we generated data and cut-offs for each individual study. Corresponding to the exemplary meta-analysis, we generated data for eight studies. In the individual studies, prevalence is binomially distributed with $p = 0.35$ and the number of patients is drawn uniformly distributed from the interval (30; 200). The data were then drawn from a multivariate normal distribution $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ as covariance matrix with correlation equal to 0.6 and variance equal to 1 and with expectation 0 for the individuals without and $\Phi^{-1}(AUC)\sqrt{2}$ for the individuals with target condition for the pre-specified AUC = 0.85 (using the R package mvtnorm[21]). Given this AUC of 0.85, seven different cut-off values are set per study such that either the sensitivity is 70%, 80%, or 95%, or the

TABLE 1 Results from the individual studies as reported from Henry et al. (2016)[18] and Dubos et al. (2006)[19]

Study number	Study identifier	Design	Cases (*n*) with bacterial infection	Cases (*n*) with viral infection	Cut-off (ng/ml)	TP (*n*)	FN (*n*)	TN (*n*)	FP (*n*)
1	Alkohli et al (2011)	Prospective	20	20	**2**	**20**	**0**	**13**	**7**
1	Alkohli et al (2011)	Prospective	20	20	10	18	2	17	3
2	Dubos et al (2006)	Retrospective	21	146	0.2	21	0	45	101
2	Dubos et al (2006)	Retrospective	21	146	**0.5**	**19**	**2**	**130**	**16**
3	Dubos et al (2008)	Retrospective	96	102	**0.5**	**95**	**1**	**85**	**17**
4	Gendrel et al (1998)	Prospective	23	51	**0.2**	**23**	**0**	**51**	**0**
4	Gendrel et al (1998)	Prospective	23	51	0.5	22	1	51	0
5	Ibrahim et al (2011)	Prospective	18	20	**0.5**	**17**	**1**	**19**	**1**
6	Liu et al (2006)	Prospective	18	23	0.445	18	0	20	3
6	Liu et al (2006)	Prospective	18	23	**0.5**	**18**	**0**	**21**	**2**
6	Liu et al (2006)	Prospective	18	23	1.05	18	0	21	2
6	Liu et al (2006)	Prospective	18	23	2	17	1	21	2
6	Liu et al (2006)	Prospective	18	23	4.06	16	2	22	1
6	Liu et al (2006)	Prospective	18	23	4.83	14	4	22	1
6	Liu et al (2006)	Prospective	18	23	8.395	14	4	23	0
7	Mayah et al (2013)	Prospective	26	32	**3.3**	**23**	**3**	**28**	**4**
8	Onal et al (2008)	Prospective	16	13	**0.5**	**15**	**1**	**13**	**0**

Note: Numbers in bold are the results used in Henry et al. (2016).[19]
Abbreviations: FN, false negatives; FP, false positives; TN, true negatives; TP, true positives.

specificity is 70%, 80%, or 95%, or the Youden value is maximal. From this pool of seven cut-off values, a random selection is then used, where the number of cut-off values is uniformly distributed from the interval (1; 7).
- In a second step, we calculated the corresponding entries of the fourfold tables (as these entries for the individual cut-offs are commonly the only information present for a meta-analysis).
- In the third step we finally applied our semiparametric approach including generation of uniformly distributed fictitious data as described above and estimation of the AUC, sensitivity, and specificity of the meta-analysis with corresponding confidence intervals.

Starting from the baseline scenario, the parameters were varied individually, resulting in 31 further scenarios (see Table 2).

- Table 2-To put the properties of our approach in perspective, we added the approach of Steinhauser et al. (2016) as a method of comparison, using the R package diagmeta (version 0.5–0).[11,22] We chose the CIDS model and logistic distribution as recommended and set log. cutoff to FALSE. For the confidence interval for the AUC, we chose the option that this is calculated based on the confidence region for the specificity given the sensitivity.

We simulated 1000 meta-analyses for each scenario. The simulation program was written in R 3.6.1[23] and the semiparametric approach was computed with the package diagnostic 0.4.4[24] (see supplementary material). The simulation study is based on the thesis of Kirstein (2019).[25]

4.2 | Simulation results

All simulation results are given in tables in the online appendix S1. For the baseline scenario, we see that the semiparametric approach leads to a larger relative bias in terms of underestimation for AUC and sensitivity than the approach of Steinhauser et al. (−4% and −2% versus −1% and −0.2%, respectively). In contrast, for specificity, the approach of Steinhauser leads to a larger bias in terms of overestimation (+2%), whereas the semiparametric approach continues to underestimate (−1%). The MSE is comparable for both approaches, the coverage for the semiparametric approach is clearly below the targeted 95% with 56%–78%. The approach of Steinhauser et al. leads to a slight too low coverage probability for the

AUC with 90, while for sensitivity and specificity the coverage probability is the targeted 95%.

These results are mirrored across all scenarios. Note that one scenario (sensitivity = 95%, specificity = 43%) results in such a strong bias for the semiparametric approach (−42%) that it would make the graphical representation difficult to read - therefore, we have left this scenario in the table in the appendix S1, but excluded it in the figures.

In Figure 1, one can see that the semiparametric approach underestimates the diagnostic accuracy almost always, whereas the approach of Steinhauser et al. overestimates specificity almost always. While for the AUC the bias of the semiparametric approach is relatively large with predominantly −4% to −5%, it amounts to about −2% for sensitivity and only between −1% and −2% for specificity. The same behavior of increasing point estimates is also seen with the Steinhauser et al. approach, on a smaller scale: for the AUC an underestimation of about 1%, for sensitivity virtually no bias, and for specificity an overestimation.

The mean squared error is correspondingly larger for the semiparametric approach than for the Steinhauser et al. approach, with the difference being largest for AUC and relatively small for sensitivity (see Figure 2).

Accordingly, the coverage probability for the Steinhauser approach is somewhat low for the AUC but very good for sensitivity and specificity, whereas it is clearly too low for the semiparametric approach (median coverage for AUC, sensitivity, and specificity: semiparametric 52%, 78%, 79%; Steinhauser et al. 90%, 94%, 95%, see Figure 3).

Looking at the results of the semiparametric approach when varying the parameters, it is noticeable that there are no substantial changes in the bias and MSE for the AUC, but the coverage probability is particularly low with large sample sizes, which is due to the narrower confidence intervals. Conversely, the highest coverage probability is obtained with small sample sizes, few studies and low prevalences due to the resulting wide confidence intervals.

Regarding the bias, it is noticeable that for sensitivity the bias is largest for a high AUC (and relatively low sensitivity) and small sample size, and smallest for a high sensitivity and large sample size. Correspondingly, for specificity, the bias is largest at a high sensitivity (and thus low specificity) and smallest at a low prevalence and thus larger number of individuals without a target condition.

TABLE 2 Variations of the individual parameters in the simulation study

Parameter	Baseline scenario	Variations
Number of studies	8	4, 15
Minimum and maximum number of individuals per study	(30; 200)	(30; 500), (30; 700), (100; 200), (100; 500), (100; 700), (30; 700), (30; 30), (100; 100), (200; 200); (500; 500)
Correlation between studies	0.6	0.3, 0.4, 0.5, 0.7, 0.8, 0.9
Pre-specified AUC	0.85	0.9, 0.95
Pre-specified sensitivity	0.8	0.85, 0.9, 0.95
Prevalence	0.35	0.15, 0.2, 0.25, 0.3, 0.4, 0.45, 0.5

Abbreviation: AUC, area under the curve.

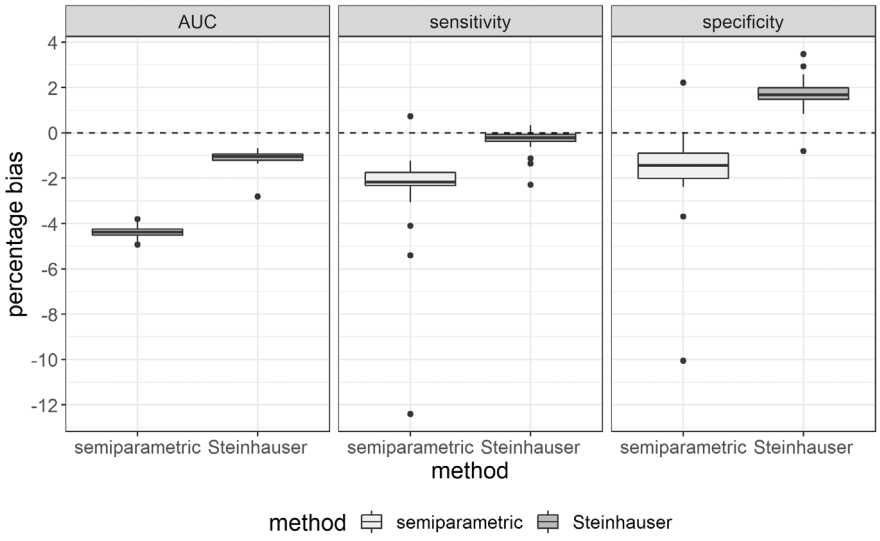

FIGURE 1 Percentage bias for the semiparametric and the Steinhauser approach in 30 tested scenarios. Panels are divided by AUC, sensitivity, and specificity. AUC, area under the curve

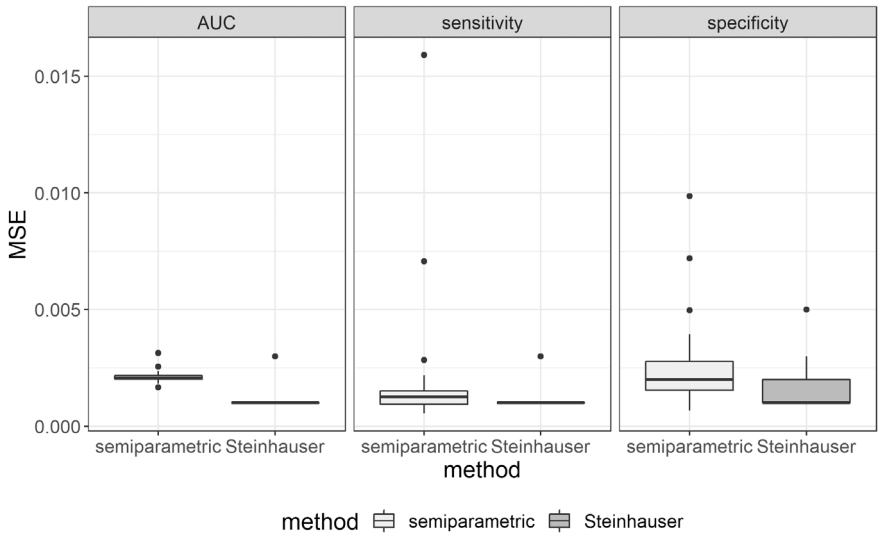

FIGURE 2 Mean squared error (MSE) for the semiparametric and the Steinhauser approach in 30 tested scenarios. Panels are divided by AUC, sensitivity, and specificity. AUC, area under the curve

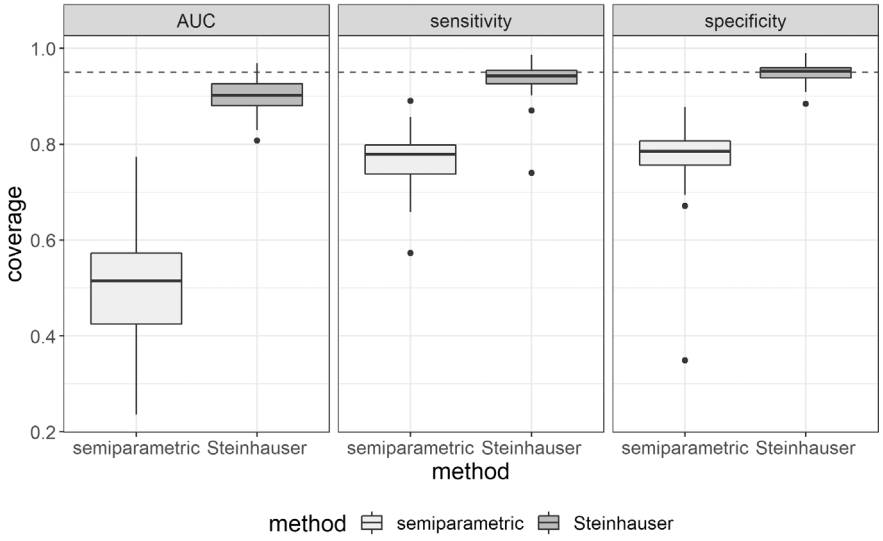

FIGURE 3 Coverage probability for the semiparametric and the Steinhauser approach in 30 tested scenarios. Panels are divided by AUC, sensitivity, and specificity. AUC, area under the curve

While no large effects can be observed for sensitivity and specificity in the MSE, the correlation between the studies is a notable additional modifier in the coverage probability, with a low-correlation leading to a higher coverage probability.

The approach of Steinhauser et al. does not converge in four scenarios in up to 1% of the simulation runs: in case of low prevalence of 15%, extreme sample sizes (w_{is} fixed at 30 or between 30 and 700) and few studies ($N = 4$), whereas the semiparametric approach always yields results.

5 | NUMERICAL EXAMPLE

To estimate pooled sensitivities and specificities, Henry et al. (2016) applied a random-effects model. However, no details are provided about the model specification and method used. Further, they computed the summary ROC (SROC) curve and estimated the AUC. The authors presented an AUC of 0.97 for the eight studies. Unfortunately, the authors did not consider the varying cut-offs among and within the individual studies. Instead, they selected one cut-off for each study resulting in estimators based on cut-offs of 0.2 ng/ml (1 study), 0.5 ng/ml (5 studies), 2 ng/ml (1 study), and 3.3 ng/ml (1 study) and thus discarded study results from 9 out of 17 cut-offs (52.9%). For a subgroup of five studies with the common cut-off of 0.5 ng/ml, the pooled sensitivity was 0.97 (95% confidence interval [CI] = 0.93–0.99) and the pooled specificity was 0.88 (95% CI = 0.84–0.92).

With the semiparametric approach as well as with the approach of Steinhauser et al., we can take into account all 17 cut-offs. Compared to the reported AUC of 0.97, Steinhauser's approach yielded a similar result with AUC = 0.964, whereas the semiparametric approach

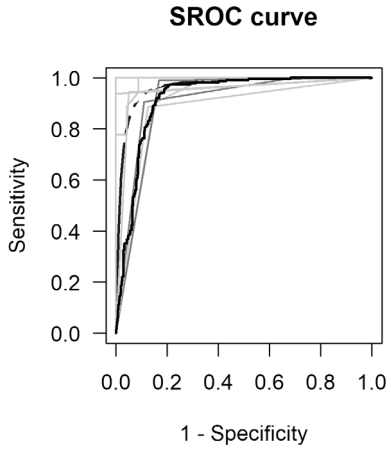

FIGURE 4 ROC curves from the meningitis study. The black curve refers to the meta-analysis results (Steinhauser = dashed, semiparametric = solid) and the gray curves refer to the results from the individual studies (light gray sample sizes below 100, dark gray between 100 and 200). ROC, receiver operating characteristic

TABLE 3 Sensitivity and specificity with corresponding 95%-confidence intervals for the meta-analysis of meningitis diagnostic accuracy studies

Approach	Sensitivity (95% CI)	Specificity (95% CI)
Henry et al.	0.97 (0.93; 0.99)	0.88 (0.84; 0.92)
Semiparametric	0.97 (0.943; 0.998)	0.80 (0.604; 0.993)
Steinhauser et al.	0.95 (0.838; 0.986)	0.87 (0.741; 0.938)

estimated the AUC lower with 0.918 (95% CI = 0.869–0.966). The approach of Steinhauser et al. gives two confidence intervals for the AUC, based on the confidence region of the sensitivity given the specificity and vice versa. In this example, the estimators and 95% confidence intervals are 0.965 (0.378–0.998) for sensitivity given specificity and 0.965 (0.217–0.996) for specificity given sensitivity. Thus, the confidence intervals are very broad.

The differences in the AUC are also clearly reflected in the SROC curves in Figure 4. The SROC curve of the approach of Steinhauser et al. lies above the SROC curve of the semiparametric approach (both with thick lines). In addition, the ROC curves of the individual studies are plotted (the studies with a sample size below 100 in light gray and the two studies by Dubos with a sample size of close to 200 in dark gray). It seems that the results of the semiparametric approach are more strongly influenced by the large studies than those of Steinhauser et al. (2016).

In Table 3 we present in addition the estimated sensitivity and specificity with their corresponding 95% confidence intervals for the mentioned cut-off of 0.5 ng/ml. The differences fit the different SROC curves.

6 | DISCUSSION

In this article, we presented a semiparametric approach for meta-analyses of diagnostic accuracy studies with multiple cut-offs. In the literature, the pragmatic approaches of using only one cut-off per study (as in the example study) or performing individual meta-analyses for the individual cut-offs is often applied in this situation. However, these approaches lead to a loss of information and are not reasonable or not possible at all if the cut-offs are very different. In contrast, the semiparametric approach can take all information into account and provide a summary ROC curve with corresponding area under the curve and pairs of sensitivity and specificity for individual cut-offs. We investigated the statistical properties in a simulation study and compared it with the approach of Steinhauser et al. (2016). In terms of relative bias, the semiparametric approach is inferior to that of Steinhauser et al. for AUC and sensitivity, but superior for specificity. However, the underestimation of the semiparametric approach in sensitivity and specificity is relatively small, ranging from 1 to approximately 2%. Regarding coverage probability, the semiparametric approach is inferior to Steinhauser et al. While the semiparametric approach always yields results, this is not the case for the approach of Steinhauser et al. in extreme scenarios.

The approach of Steinhauser et al. yields two confidence intervals for the AUC, one based on the confidence region for sensitivity given the specificity and vice versa. However, in the example study the confidence intervals were very broad. Furthermore, the results of the example study may indicate that the semiparametric approach is more reflective of the large studies.

The strength of the paper is that a real-world meta-analysis was chosen as the starting point, and from here, different parameters were varied in the simulation study. Thus, the approach is evaluated under realistic conditions. Furthermore, the comparison with the approach of Steinhauser et al. allows an assessment of the results. A weakness of the work is that not all parameter variations were combined and also some settings, such as skewed distributions and between-study heterogeneity, were not taken into account. Furthermore, only one alternative approach was used for comparison. Reasons for this are the existing implementation of the approach of Steinhauser et al. in R and that an even more extensive simulation study would go beyond the scope of this article.

The available approaches which can consider multiple and varying cut-offs have different prerequisites, enable the output of different results, and have specific advantages (cf. Table 4 in Zapf et al., 2021). The main advantage of the semiparametric approach is that it is not

an iterative procedure and therefore always produces results. Since the semiparametric approach underestimates diagnostic accuracy and is thus a conservative procedure, an option would also be to use this approach as a fallback option if the primary approach does not converge. It should, however, be considered, that fictitious data has to be generated. In this article we present results based on uniform distributed data and provide results based on single point distributed data in the appendix S1. The differences are substantial and we would recommend the uniform distribution for generating fictitious data. The statistical properties are better and the sROC curve is very well interpretable. The use of the single point distribution is not recommended from our point of view, even if it is easier from a programming point of view. Whether other distributions show even better properties would have to be checked. However, then again more assumptions would be needed and the goal was to remain as far as possible assumption-free. From our point of view, the decision how to generate the fictitious data has to be made in any case before starting the data analysis in order not to compromise the validity of the study.

So far, there is no systematic comparison of the different approaches, but this work is a first step.

Another approach that can be used comes from Riley and colleagues who consider the situation of different cut-off values as a meta-analysis with missing data and proposes an imputation of these missing cut-off values.[26–28] In doing so, they assume that the logit sensitivity and specificity follow a multivariate-normal distribution.

Currently, a large simulation study is being conducted to systematically compare the four approaches used in the data challenge. If possible, the imputation approach of Riley and colleagues should also be included. Among other things, skewed data will be generated, between-study heterogeneity will be represented, and all parameter variations will be combined. This will also be an opportunity to investigate whether the dependence on study size that has been suggested here can be confirmed. If so, and if this is not deemed reasonable, a possible adaptation of the semiparametric approach would be the unweighted variance estimation of the accuracy measures (see Brunner and Zapf, 2014).

AUTHOR CONTRIBUTIONS

AZ and CF developed the semiparametric approach. MK drafted the simulation program and AZ and CF designed and performed the simulations. CF and AZ drafted the manuscript and all authors revised it.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The R-simulation program and data and R-program for the example meta-analysis is available in the supplementary material.

ORCID

Cornelia Frömke https://orcid.org/0000-0002-4028-0604
Antonia Zapf https://orcid.org/0000-0002-8467-0508

REFERENCES

1. CEBM Levels of evidence working group. "The Oxford levels of evidence 2". Oxford Centre for Evidence-Based Medicine https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009. 2022.
2. European Medicines Agency. Guideline on clinical evaluation of diagnostic agents. 2009. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003580.pdf.
3. Leeflang MG, Deeks D, Gatsonis C, Bossuyt PM. (on behalf of the Cochrane diagnostic test accuracy working group)Systematic reviews of diagnostic test accuracy. *Ann Intern Med*. 2008; 149:889-897.
4. Leeflang MMG. Systematic reviews and meta-analyses of diagnostic test accuracy. *Clin Microbiol Infect*. 2014;20(2):105-113.
5. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-990.
6. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol*. 2006;59(12):1331-1332.
7. Kuss O, Hoyer A, Solms A. Meta-analysis for diagnostic accuracy studies: a new statistical model using beta-binomial distributions and bivariate copulas. *Stat Med*. 2014;33(1):17-30.
8. Zapf A, Hoyer A, Kramer K, Kuss O. Nonparametric meta-analysis for diagnostic accuracy studies. *Stat Med*. 2015;34(29): 3831-3841.
9. Zapf A. In: Biondi-Zoccai G, ed. *Appraising Heterogeneity. Diagnostic Meta-Analysis*. Springer International Publishing; 2018: 125-160.
10. Hoyer A, Hirt S, Kuss O. Meta-analysis of full ROC curves using bivariate time-to-event models for interval-censored data. *Res Synth Methods*. 2018;9(1):62-72.
11. Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Med Res Methodol*. 2016;16(97):97.
12. Jones HE, Gatsonsis CA, Trikalinos TA, Welton NJ, Ades AE. Quantifying how diagnostic test accuracy depends on threshold in a meta-analysis. *Stat Med*. 2019;38(24):4789-4803.
13. Zapf A, Albert C, Frömke C, et al. Meta-analysis of diagnostic accuracy studies with multiple thresholds: comparison of different approaches. *Biom J*. 2021;63(4):699-711.
14. Werner C, Brunner E. Rank methods for the analysis of clustered data in diagnostic trials. *Comput Stat Data Analysis*. 2007; 51:5041-5054.
15. Brunner E, Zapf A. Nonparametric ROC analysis for diagnostic trials. In: Balakrishnan N, ed. *Statistics in Clinical Trials* –

Planning, Analysis, and Inferential Methods. Vol 2. John Wiley & Sons; 2014:471-483.
16. Konietschke F, Brunner E. Nonparametric analysis of clustered data in diagnostic trials: estimation problems in small sample sizes. *Comput Stat Data Analysis*. 2009;53(3):730-741.
17. Lange K, Brunner E. Sensitivity, specificity and ROC-curves in multiple reader diagnostic trials—a unified, nonparametric approach. *Stat Methodol*. 2012;9(4):490-500.
18. Henry BM, Roy J, Ramakrishnan PK, Vikse J, Tomaszewski KA, Walocha JA. Procalcitonin as a serum biomarker for differentiation of bacterial meningitis from viral meningitis in children: evidence from a meta-analysis. *Clin Pediatr (Phila)*. 2016;55(8):749-764.
19. Dubos F, Moulin F, Gajdos V, et al. Serum procalcitonin and other biologic markers to distinguish between bacterial and aseptic meningitis. *J Pediatr*. 2006;149(1):72-76.
20. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med*. 2006;25(24):4279-4292.
21. Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2021). Mvtnorm: Multivariate Normal and t Distributions. R Package Version 1.1-2, https://CRAN.R-project.org/package=mvtnorm.
22. Rücker G, Steinhauser S, Kolampally S, Schwarzer G (2022). Diagmeta: meta-analysis of diagnostic accuracy studies with several cutpoints. R Package Version 0.5–0, https://cran.r-project.org/web/packages/diagmeta/index.html.
23. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2020 https://www.R-project.org/
24. Rooney D. Covariate adjusted nonparametric estimation of sensitivity and specificity in clustered data [Master Thesis]. Heidelberg: Ruprecht-Karls-Universität Heidelberg; 2017.
25. Kirstein M. Simulation der überdeckungswahrscheinlichkeit von nichtparametrischen konfidenzintervallen zur metaanalyse von diagnosegütestudien mit multiplen trennwerten [Bachelor Thesis]. Hochschule Hannover –University of Applied Sciences and Arts; 2019.
26. Riley RD, Ahmed I, Ensor J, et al. Meta-analysis of test accuracy studies: an exploratory method for investigating the impact of missing thresholds. *Syst Rev*. 2015;4:12.
27. Riley RD, Takwoingi Y, Trikalinos T, et al. Meta-analysis of test accuracy studies with multiple and missing thresholds: a multivariate-normal model. *J Biometr Biostat*. 2014;5:3.
28. Ensor J, Deeks JJ, Martin EC, Riley RD. Meta-analysis of test accuracy studies using imputation for partial reporting of multiple thresholds. *Res Synth Methods*. 2018;9(1):100-115.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Frömke C, Kirstein M, Zapf A. A semiparametric approach for meta-analysis of diagnostic accuracy studies with multiple cut-offs. *Res Syn Meth*. 2022;13(5):612-621. doi:10.1002/jrsm.1579