Review

Peter Schlattmann*

Tutorial: statistical methods for the meta-analysis of diagnostic test accuracy studies

https://doi.org/10.1515/cclm-2022-1256
Received December 10, 2022; accepted December 21, 2022;
published online January 19, 2023

Abstract: This tutorial shows how to perform a meta-analysis of diagnostic test accuracy studies (DTA) based on a 2×2 table available for each included primary study. First, univariate methods for meta-analysis of sensitivity and specificity are presented. Then the use of univariate logistic regression models with and without random effects for e.g. sensitivity is described. Diagnostic odds ratios (DOR) are then introduced to combine sensitivity and specificity into one single measure and to assess publication bias. Finally, bivariate random effects models using the exact binomial likelihood to describe within-study variability and a normal distribution to describe between-study variability are presented as the method of choice. Based on this model summary receiver operating characteristic (sROC) curves are constructed using a regression model logit-true positive rate (TPR) over logit-false positive rate (FPR). Also it is demonstrated how to perform the necessary calculations with the freely available software R. As an example a meta-analysis of DTA studies using Procalcitonin as a diagnostic marker for sepsis is presented.

Keywords: area under the curve (AUC); diagnostic test accuracy (DTA); generalized linear mixed model (GLMM); meta-analysis; Procalcitonin; sensitivity; specificity; summary operator curve (sROC).

Introduction

The publication of meta-analyses [1–3] and especially meta-analyses of diagnostic test accuracy (DTA) studies [4–8] has a long tradition in Clinical Chemistry and Laboratory Medicine (CCLM). Such meta-analyses play an important role in health technology assessment [9]. Besides subject matters also methodological issues are of importance and thus are published in CCLM [10, 11].

There are numerous methods available for meta-analyses of DTA studies [12]. Basic requirement is the availability of a 2×2 table for each included primary study. First, we start with univariate methods for meta-analysis of sensitivity and specificity. That is, fixed and random effects univariate meta-analyses using logistic regression without and with random effects are presented. Next, diagnostic odds ratios (DOR) are introduced in order to combine sensitivity and specificity into one measure and to assess publication bias. Then, we present bivariate random effects meta-analyses with maximum likelihood (using the exact binomial likelihood to describe within-study variability) and a normal distribution to describe between-study variability. Finally, summary receiver operating characteristic (sROC) curves are constructed using regression models logit-true positive rate (TPR) over logit-false positive rate (FPR). Based on sROC curves the overall diagnostic performance can be evaluated using the area under the curve (AUC). The necessary calculations can be done with the freely available software R [13] and are described in detail in this review.

Motivating example

Worldwide, sepsis and its sequelae still remain a frequent cause of acute illness and death in patients with community and nosocomial acquired infections [14]. Sepsis may be seen as systemic inflammatory response due to infection. However, a gold standard for the proof of infection is missing. Depending on prior antibiotic therapy, bacteremia is found only in approximately 30% of patients with sepsis. Furthermore, early clinical signs of sepsis, like fever, tachycardia, and leucocytosis, are unspecific and overlap with signs also seen in a multitude of systemic inflammatory response syndromes (SIRS) in the absence of infection, especially in surgical patients. Other signs, such as arterial hypotension, thrombocytopenia, or elevated lactate levels indicate, too late, the progression to organ dysfunction. Thus, delay in diagnosis and treatment of sepsis causes increased mortality.

***Corresponding author: Peter Schlattmann**, Jena University Hospital, Institute of Medical Statistics, Computer and Data Sciences, Jena, Germany, E-mail: peter.schlattmann@med.uni-jena.de. https://orcid.org/0000-0001-7420-7707

In sepsis numerous humoral and cellular systems are activated, followed by a release of a multitude of mediators and other molecules that mediate the host response to infection. Several potential diagnostic indicators measured in the bloodstream have been evaluated for their clinical ability to assess the diagnosis and severity of sepsis. One of these, the 116 amino acid polypeptide procalcitonin (PCT) is frequently used when it comes to identify bacterial infections.

In this tutorial we will use Procalcitonin as an example for a meta-analysis of DTA studies using data from [15]. This is a meta-analysis of Procalcitonin (PCT) for diagnosis of sepsis in critically ill patients. Data sources were Medline, Embase, ISI Web of Knowledge, the Cochrane Library, Scopus, BioMed Central, and Science Direct, from inception to Feb 21, 2012, and reference lists of identified primary studies. Articles written in English, German, or French that investigated Procalcitonin for differentiation of septic patients – those with sepsis, severe sepsis, or septic shock – from those with a systemic inflammatory response syndrome of non-infectious origin were included. Excluded studies were studies of healthy people, patients without probable infection, and children younger than 28 days. Two independent investigators extracted patient and study characteristics, discrepancies were resolved by consensus. The search returned 3,487 reports, of which 31 fulfilled the inclusion criteria, accounting for 3,244 patients. Table 1 shows PCT data for diagnosis of sepsis which were extracted from the 31 studies.

Table 1: Meta-analysis of procalcitonin (PCT) for diagnosis of sepsis study data.

Name	Year	TP	FP	TN	FN	Cut-off, g/L
Ahmadinejad	2009	63	11	38	8	0.5
Al-Nawas	1996	73	45	170	49	0.5
Arkader	2006	12	0	14	2	2
Bell	2003	47	2	19	15	15.75
Castelli	2004	21	2	13	13	1.2
Clec'h	2006	29	2	38	7	1
Clec'h	2006	28	9	27	3	9.7
Dorizzi	2006	42	6	26	9	1
Du	2003	16	8	23	4	1.6
Gaini	2006	56	9	10	18	1
Gibot	2004	39	9	20	8	0.6
Groselj-Grenc	2009	20	3	9	4	0.28
Harbath	2001	58	4	14	2	1.1
Hsu	2011	31	0	11	24	2.2
Ivancevic	2008	34	5	17	7	1.1
Jimeno	2004	17	5	58	24	0.5
Kofoed	2007	77	23	32	19	0.25
Latour-Perez	2010	53	5	37	19	0.5
Meynaar	2011	31	9	35	1	2
Naeini	2006	22	1	24	3	0.5
Oshita	2010	76	11	45	36	0.5
Pavcnik-Arnol	2007	17	2	17	13	5.79
Ruiz-Alvarez	2009	65	9	16	13	0.32
Sakr	2008	82	92	116	37	2
Selberg	2000	19	5	6	3	3.3
Simon	2008	17	10	29	8	2.5
Suprin	2000	49	6	14	26	2
Tsalik	2011	168	33	56	79	0.1
Tsangaris	2009	19	2	21	8	1
Tugrul	2002	55	2	8	20	1.31
Wanner	2000	34	20	68	11	1.5

Univariate meta-analyses of sensitivity and specificity

Forest plots for sensitivity and specificity

One way to perform a diagnostic meta-analysis is to analyze sensitivity and specificity separately as those are key parameters when evaluating the performance of a binary diagnostic test [16]. This requires knowledge of a reference or gold standard which denotes the disease status D. The potential outcomes of a 2×2 table showing the disease status D in the columns and test results T in the rows are shown in Table 2. For a detailed description and examples see e.g. Schlattmann [17].

Table 2: Potential outcomes of a diagnostic test with known reference standard.

	Disease present D^+	Disease absent D^-	Total
Test positive T^+	True positive (TP)	False positive (FP)	TP + FP
Test negative T^-	False negative (FN)	True negative (TN)	FN + TN
Total	n_1	n_2	n

In a diagnostic meta-analysis we have for each individual study (i=1,…,k) study specific sensitivities (true positive rate, TPR) with $\widehat{\mathrm{Se}_i} = \frac{\mathrm{TP}_i}{\mathrm{TP}_i+\mathrm{FN}_i}$. Here TP_i are the true positives and FN_i are the false negatives according to a gold standard. Looking at the study from Ahmadinejad (2009) we find a sensitivity equal to $\widehat{\mathrm{Se}} = \frac{63}{63+8} = 0.887$. Likewise, for each study specificity (true negative rate, TNR) is given by $\widehat{\mathrm{Sp}}_i = \frac{\mathrm{TN}_i}{\mathrm{TN}_i+\mathrm{FP}_i}$. Here TN_i are the true negatives and FP_i are the false positives. Again, for the study by Ahmadinejad specificity is given by $\mathrm{spec} = \frac{38}{38+11} = 0.776$.

For a graphical presentation for each study sensitivity and specificity are calculated together with a 95% confidence interval and displayed in a so called forest plot. There are several ways to construct a confidence interval for a

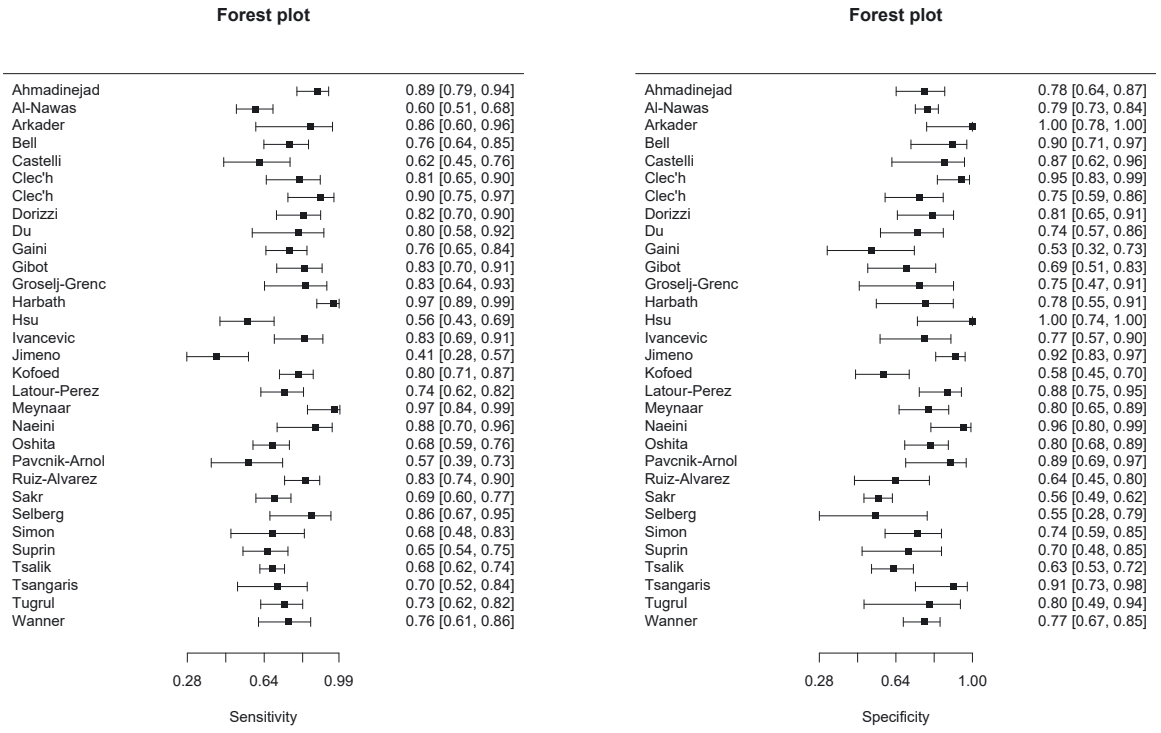

Figure 1: Univariate forest plots for sensitivity and specificity in Procalcitonin studies [15].

binomial proportion with different statistical properties [18, 19]. Figure 1 shows a forest plot of sensitivity of PCT on the left hand side and of specificity on the right hand side. In this plot set we see considerable heterogeneity for sensitivity ranging from 0.415 to 0.969. Likewise for specificity we find heterogenous results which range from 0.526 to 1.000.

Fixed and random effects models

Statistically speaking sensitivity and specificity are proportions and can be treated as such in a meta-analysis [20].

Standard fixed effects models for meta-analyses can be applied [21–23]. One approach is using log transformed odds of sensitivity and specificity (logit transform). Odds are defined as odds $= \frac{p}{1-p}$ where p is a probability. If we have a fair coin the odds for head are $\frac{0.5}{1-0.5} = 1$ meaning that head and tail are equally likely. A logit transform of sensitivity is given as $\text{logit}\,(\text{Se}) = \log\left(\frac{\text{Se}}{1-\text{Se}}\right)$, where log denotes the natural logarithm.

Summary estimates of logit-transformed sensitivity and logit-transformed specificity, respectively are obtained as a weighted average of the respective logit-transformed proportions of the individual studies. Weights are given by the inverse of the respective study specific variances. This has the disadvantage that in the case of zero entries undefined log odds occur. Thus, in the past years there has been a lively discussion how to avoid undefined log odds [24–26] by adding e.g. 0.5 to each cell of the study specific 2×2 table in case of zero cells.

To avoid this, we apply logistic regression models potentially with random effects aka generalized linear mixed models. That is, we assume that sensitivity and specificity respectively follow a binomial distribution. Thus, for each study:

$$\begin{aligned} \text{TP}_i &\sim \text{Binomial}\,(n_{1i}, \text{Se}_i) \\ \text{TN}_i &\sim \text{Binomial}\,(n_{2i}, \text{Sp}_i) \end{aligned}$$

A common effect logistic regression model for sensitivity has the form

$$\log\left(\frac{\text{Se}_i}{1-\text{Se}_i}\right) = \beta_0 \tag{1}$$

This is a generalized linear model with binomial errors, linear predictor β_0 and logistic link function. The left hand side shows the natural logarithm of the odds of sensitivity. The unknown parameter β_0 can be estimated using maximum likelihood using numerous statistical software packages such as R [13]. Also, this is a so called common effect model, since it assumes the overall sensitivity in each study is identical and given by

Table 3: PCT and sepsis – two univariate meta-analyses for sensitivity and specificity using common effect logistic regression models.

Parameter	Logit-transformed		Back-transformed		Heterogeneity variance
	Coefficient	Standard error	Estimate	95% CI	$\hat{\tau}^2$
Sensitivity	1.022	0.053	0.735	(0.715, 0.755)	–
Specificity	1.080	0.061	0.747	(0.723, 0.769)	–

$$\widehat{\mathrm{Se}} = \frac{\exp\left(\widehat{\beta}_0\right)}{1 + \exp\left(\widehat{\beta}_0\right)} \tag{2}$$

For the PCT data an application of two univariate common effects models for sensitivity and specificity yields the results presented in Table 3. Overall, assuming a common effect we find a sensitivity equal to 0.735 with 95% CI (0.715, 0.755) and a specificity equal to 0.747 with 95% CI (0.723, 0.769).

A common effect model assumes that the underlying true sensitivity is the same in each study. The overall variation and, therefore, the confidence intervals will reflect only random variation within each study but not any potential heterogeneity between the studies. Of course, the same applies for specificity.

Whether pooling of the data in this way is appropriate should be decided after investigating the heterogeneity of the study results. If the results vary substantially, no fixed effects pooled estimator should be presented [27]. As a result only estimators e.g. for selected subgroups should be calculated. The previous remark notwithstanding, a fixed effects meta-analysis is always valuable, since it tests the null-hypothesis that diagnostic accuracy was identical in all trials [28]. If the null-hypothesis is rejected then the alternative may be asserted that at least one study differs.

One way to address heterogeneity is the calculation of Cochran's Q-statistic and the I^2 measure. This describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance). For sensitivity we find $\widehat{I}^2$=69.8% and for specificity $\widehat{I}^2$=67.4%. Likewise, the test for heterogeneity turns out to be statistically significant (sensitivity Q=99.4, df=30, p<0.001, specificity Q=92.06, df=30, p<0.001).

Thus, the investigation of heterogeneity between studies is a main task in each meta-analysis [29]. Here a common effect model is not appropriate. Alternatively, a random effects model which incorporates variation between studies should be considered.

A random effects logistic regression model has then the form

$$\log\left(\frac{\mathrm{Se}_i}{1-\mathrm{Se}_i}\right) = \beta_0 + b_i, \quad b_i \sim N(0, \tau^2), i = 1, \ldots, k \tag{3}$$

Again, this is a generalized linear model with binomial errors, linear predictor β_0 and logistic link function. Additionally, we assume variability between studies given by the study specific departure b_i from the overall intercept β_0. For the b_i a normal distribution with expectation zero and heterogeneity variance τ^2 is assumed. The latter indicates variability between studies, i.e. heterogeneity. Both unknown parameters again can be estimated using maximum likelihood. Table 4 shows the result for the PCT data.

Overall, assuming a random effects model we find in Table 4 a sensitivity equal to 0.768 with 95% CI (0.720, 0.810) with heterogeneity variance $\widehat{\tau}^2$=0.36. For specificity, we find a value equal to 0.793 with 95% CI (0.743, 0.836) and heterogeneity variance $\widehat{\tau}^2$=0.379. Thus, we find substantial heterogeneity between studies.

Overall this approach seems to provide useful results in terms of sensitivity and specificity as e.g. investigated by Simel and Bossuyt [30]. However, we do not have any information on the correlation between sensitivity and specificity and the magnitude of the overall diagnostic performance.

Diagnostic odds ratio (DOR)

So far, we have considered sensitivity and specificity as a pair for each study. There have been many attempts to merge the results of a diagnostic study into one single measure. One proposal is the diagnostic odd ratio (DOR) [11, 31]

Table 4: PCT and sepsis – two univariate meta-analyses for sensitivity and specificity using random effects logistic regression models.

Parameter	Logit scale		Back-transformed		Heterogeneity variance
	Mean	Standard error	Estimate	95% CI	$\hat{\tau}^2$
Sensitivity	1.198	0.128	0.768	(0.720, 0.810)	0.360
Specificity	1.343	0.144	0.793	(0.743, 0.836)	0.379

$$\mathrm{DOR} = \frac{\frac{\widehat{\mathrm{Se}}}{1-\widehat{\mathrm{Se}}}}{\frac{1-\widehat{\mathrm{Sp}}}{\widehat{\mathrm{Sp}}}} = \frac{\frac{\mathrm{TP}}{\mathrm{FN}}}{\frac{\mathrm{FP}}{\mathrm{TN}}} = \frac{\mathrm{TP}\times\mathrm{TN}}{\mathrm{FN}\times\mathrm{FP}} \tag{4}$$

This is the ratio of the odds of a positive test result for a person with the disease divided by the odds for a positive test result for a healthy person. The value of a DOR ranges from 0 to infinity, where higher values indicate better discriminatory test performance. The synthesis of diagnostic odds ratios is straightforward and follows standard meta-analysis methods. Summary estimates of diagnostic odds ratios are obtained as a weighted average of the respective log transformed DORs of the individual studies. The weights are given by the inverse of the respective study specific variances.

First, investigating heterogeneity between studies we find substantial heterogeneity (Q=89.00, df=30, p<0.001, $\widehat{I}^2$=66.3%. Thus, we apply a random effects mode with an overall DOR=11.698, 95% CI (8.301, 16.486). Thus, we see discriminatory potential of Procalcitonin for the diagnosis of sepsis.

Apart from challenges in interpreting diagnostic odds ratios, a disadvantage is that it is impossible to weight the true positive and false positive rates separately. Likewise, it is impossible to distinguish between tests with high sensitivity and low specificity and tests with low sensitivity and high specificity. Furthermore no direct investigation of the correlation between sensitivity and specificity is possible. Thus, bivariate models are preferable and introduced in Section 3.2.

Publication bias

Publication bias is a major form of bias in any meta-analysis. That is, if the studies that are included in a review have results that systematically differ from relevant studies that are missed, then the findings will be compromised by publication bias. Thus, researchers are advised to perform a thorough literature search and to investigate publication bias. Following Deeks et al. [32] we present the effective sample size funnel plot together with the associated regression test of asymmetry. The effective sample size plot (Figure 2) takes the DOR on the x-axis of the plot and $1/\sqrt{\mathrm{ESS}}$ on the y-axis. ESS stands for effective sample size and is proportional to $\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$. This test is based on the regression of log(DOR) against $1/\sqrt{\mathrm{ESS}}$, weighting by ESS.

Unfortunately, for our example there is publication bias present (Test result: t=4.11, df=29, p-value=0.0003). More details are shown in Section 4.3. As result in a first step a repeated literature search would take place.

Bivariate diagnostic meta analysis

Plots of sensitivity and specificity in the summary receiver operator curve (sROC) space

Procalcitonin is a continuous diagnostic marker. Until now we have assumed that we are dealing with a binary diagnostic test. A frequent cut-off value equals 0.5 g/L. Values larger or equal than 0.5 g/L indicate a positive test and smaller values indicate a negative test result and thus we have transformed the continuous marker Procalcitonin into a binary test. Obviously, other cut-off values could be used. For example we could apply a cut off value ≥2.0 g/L. As a result increasing the cut-off value from 0.5 g/L to 2.0 g/L will lead to a decreased sensitivity and an increased specificity. This idea is depicted in Figure 3.

Descriptive statistics of the Procalcitonin data applied in Schlattmann [17] find a median PCT value equal to 0.2 g/L with a minimum equal to 0.01 g/L and a maximum of 200 g/L. Obviously, we could use any value between minimum and maximum as a cut off value and calculate the corresponding sensitivity and specificity.

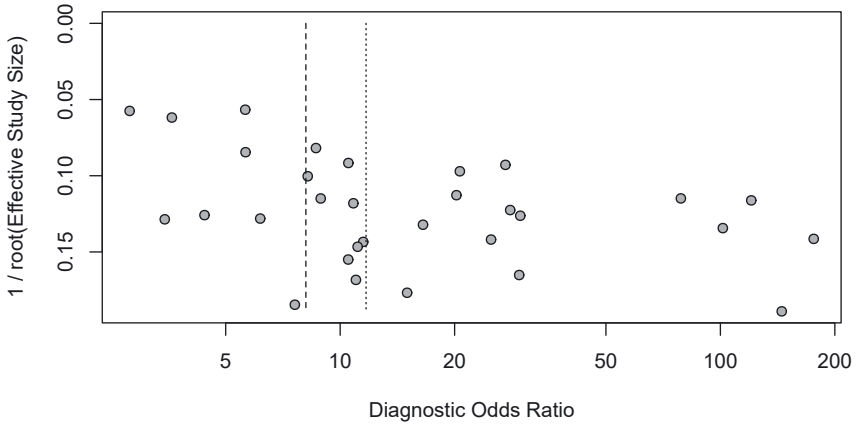

Figure 2: Diagnostic odds ratio against $1/\sqrt{\mathrm{ESS}}$, where ESS stands for effective sample size.

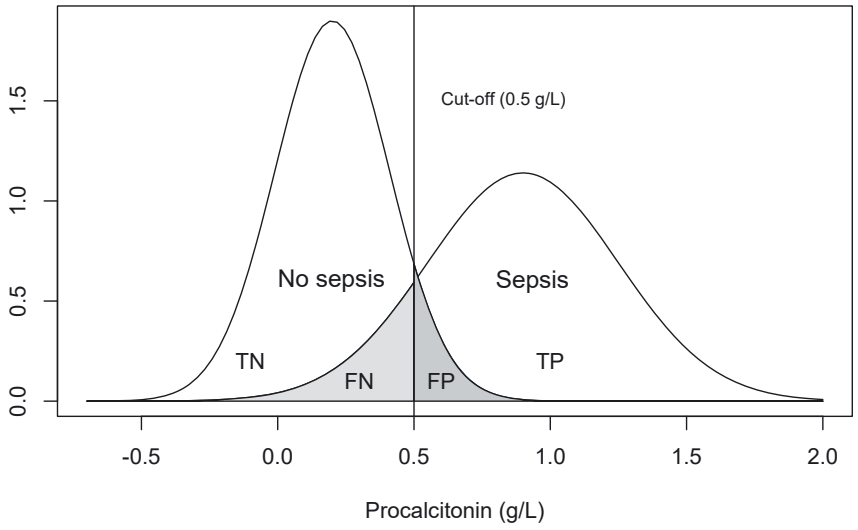

Figure 3: Illustration of the cut-off value problem for Procalcitonin. Variation of the cut-off value c leads to an increased specificity and decreased sensitivity if c is moved to the right, and vice versa if c is moved to the left.

This is done when we create a receiver operator curve (ROC) [33] which is obtained by calculating the sensitivity and specificity of every observed data value and plotting sensitivity against 1-specificity. A test that perfectly discriminates between the two groups would yield a "curve" that coincided with the left and top sides of the plot since we would not have any false negative (FN) or false positive (FP) values. A useless test would give a straight line from the bottom left corner to the top right. This implies that a true positive and a false positive test result are equally likely.

The performance of the test can be assessed by using the area under the receiver operating characteristic curve (AUC). This area may be interpreted as the probability that a random person with the disease has a higher value of the measurement than a random person without the disease. A perfect test would have an AUC=1 and a useless test has an AUC=0.5. This is shown in Figure 4.

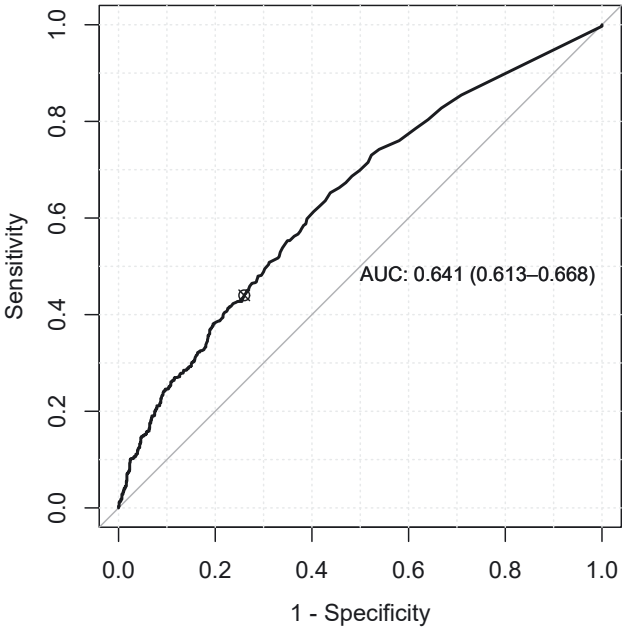

Figure 4: Procalcitonin as biomarker for sepsis: Receiver operator curve and area under the curve. The point shows sensitivity and specificity for a cut-off value of 0.5 g/L.

In diagnostic meta-analyses often only a single cut-off value for a specific study is provided. Hence not the study specific ROC curve is available but only the corresponding TP; FN, FP and TN as shown in Table 1, where e.g. Ahmadinejad applies a cut-off value of 0.5 g/L.

In order to display variation between studies due to different cut-off values plots in ROC space may be constructed. Here, a simple scatterplot of sensitivity vs. 1-specificity of each study is useful. Additional information showing also the variability within a study is shown in a cross-hair plot [34] which shows 1-specificity (false positive rate) vs. sensitivity together with the respective study specific 95% confidence intervals.

In Figure 5 the scatterplot shows variation in cut-off points as well in accuracy. Looking at the cross-hair plot on the right side we see also high variability of sensitivities and false positive rates indicating considerable heterogeneity.

Univariate meta-analyses provide single estimates of sensitivity and specificity. Here, we might be interested in a joint pair together with a confidence region. Also we saw that heterogeneity is common in DTA studies. One reason is variation in cut-off points used in the individual studies. Another reason might be due to differences in the respective patient populations. Thus, we might be interested in a prediction region which shows where future studies might fall. Finally, the construction of a summary ROC curve across studies (sROC) might be of interest. These aims can be reached using an appropriate model, that is a bivariate statistical model.

Bivariate generalized linear mixed modes

The logistic models used so far have the disadvantage to ignore the bivariate structure of the data. Thus, frequently a bivariate linear random effects model is used for a DTA

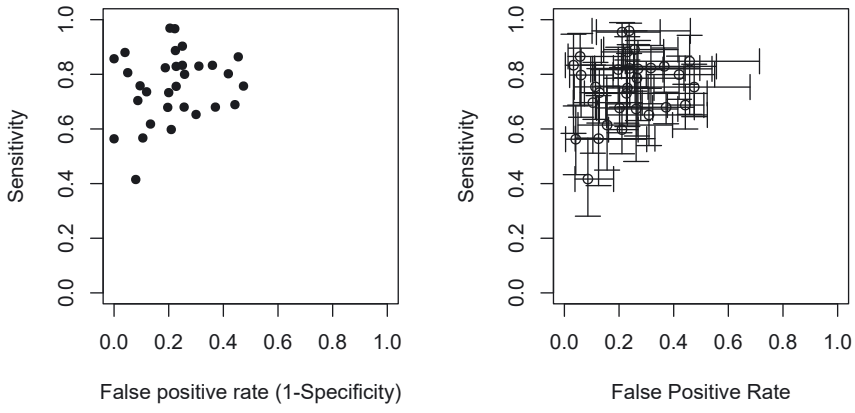

Figure 5: Scatterplot and cross-hair plot in ROC-space for the Procalcitonin data.

meta-analysis which was introduced by Reitsma et al. [35]. This model uses logit-transformed sensitivity and logit-transformed specificity simultaneously. Here it is assumed that the true logit-transformed sensitivities of the individual studies follow a normal distribution with a common mean value and between-study variability as in the univariate random effects model. Variation between studies can be attributed to unobserved heterogeneity due to e.g. heterogeneous study populations. Likewise, for the true logit-transformed specificities a normal distribution with a common mean value and between-study variability is assumed.

Now, this model introduces potential correlation between the true logit-transformed sensitivity and specificity within studies by assuming a bivariate normal distribution for the random effects. Besides variability between studies in the true underlying sensitivities and specificities, there is also variation due to sampling. Studies differ in size and thus in variation. Thus, on the second level of the model study specific variances of logit-transformed sensitivity and specificity are incorporated in order to take sampling variability into account.

As a result of this bivariate model approach summary estimates for sensitivity and specificity are obtained. In addition, based on the model's assumption of bivariate normality an sROC curve can then be constructed from the parameter estimates of the model. Performing a bivariate linear random effects model for meta-analysis of diagnostic accuracy can be done using the 'reitsma' function implemented in the freely available R-package 'mada' [36].

However, to synthesize data, an exact binomial rendition [37] of the linear bivariate mixed-effects regression model developed by van Houwelingen et al. [38] for meta-analysis of treatment trials, modified for synthesis of diagnostic test data builds an alternative. As in the linear mixed effects model the correlation between sensitivity and specificity is taken care of. Furthermore, in contrast to a logit transformation no ad hoc continuity correction to avoid zero cells in the 2×2 table is required. Thus, this model is preferable as shown in simulation studies [39] and empirical comparisons [40]. Hence, in the following we concentrate on this bivariate logistic regression model with random effects (bivariate GLMM).

Since we present our results in ROC space we make a slight shift of presentation. We now model the false positive rate, i.e. 1-specificity. As in the case of univariate models we assume a binomial distribution for sensitivity and 1-specificity respectively. Hence the binomial distribution depicts within study variability of the i=1,…,k studies:

$$\begin{aligned} \mathrm{TP}_i &\sim \text{Binomial}\,(n_{1i}, \mathrm{Se}_i) \\ \mathrm{FP}_i &\sim \text{Binomial}\,(n_{2i}, 1-\mathrm{Sp}_i) \end{aligned}$$

A bivariate random effects logistic regression model has then the form

$$\begin{aligned} \log\left(\frac{\mathrm{Se}_i}{1-\mathrm{Se}_i}\right) &= \beta_0 + \mu_i \\ \log\left(\frac{1-\mathrm{Sp}_i}{\mathrm{Sp}_i}\right) &= \beta_1 + \nu_i \end{aligned}$$

Between study variability is addressed using a bivariate normal distribution with

$$\begin{pmatrix} \mu_i \\ \nu_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right) \text{with} \quad \Sigma = \begin{pmatrix} \sigma_\mu^2 & \rho\sigma_\mu\sigma_\nu \\ \rho\sigma_\mu\sigma_\nu & \sigma_\nu^2 \end{pmatrix}. \qquad (5)$$

Here Σ denotes the covariance matrix of the bivariate random effects distribution, where σ_μ^2 denotes the between study variability of sensitivity on the logit scale. Likewise σ_ν^2 denotes the between study variability of 1-specificity on the logit scale, whereas ρ denotes the correlation between sensitivity and 1-specificity. Estimation can again be done using maximum likelihood with general statistical software such as R as shown in Section 4.4.

For our example we obtain the following results shown in Table 5.

Table 5: PCT and sepsis – bivariate meta-analyses for sensitivity and specificity based on bivariate random effects logistic regression model.

Parameter	Logit scale		Back-transformed		Heterogeneity
	Mean	Standard error	Estimate	95% CI	Σ
Sensitivity	1.189	0.128	0.767	(0.790, 0.809)	0.357
1-Specificity	−1.340	0.144	0.208	(0.742, 0.835)	0.384
Correlation					0.23
Specificity	1.340	0.144	0.792	(0.165, 0.258)	0.384

Based on the bivariate mixed effects logistic regression model we obtain an overall sensitivity equal to 0.767 and an overall specificity equal to 0.792. In terms of heterogeneity we find a variance between studies for sensitivity on the logit scale σ_μ^2 equal to 0.357 and likewise for 1-specificity a variance σ_ν^2 equal to 0.384. Importantly, we find a positive correlation, which implies a negative correlation=−0.23 between sensitivity and specificity. Only in this case the construction of a sROC curve is recommended [41].

Summary receiver operator curve (sROC curve)

According to item 21 of the PRISMA statement for DTA meta-analyses, [42], test accuracy, including variability should be reported. This includes summary results as well as confidence and prediction intervals respectively.

One way to address diagnostic test accuracy is to estimate the receiver operator curve based on the available data from the different studies. There are several methods available for sROC curve construction [43]. Here we apply the regression line of logit transformed sensitivity η based on logit transformed 1-specifcicity ξ. That is

$$\eta = \beta_0 + \frac{\rho\sigma_\mu\sigma_\nu}{\sigma_\nu^2}(\xi - \beta_1) \tag{6}$$

When transformed to the ROC space we obtain the sROC curve indicating the median sensitivity for a specific false positive rate. Figure 6 shows the sROC curve, the joint estimate of sensitivity and 1-specificity together with a 95% confidence and prediction region. This prediction region indicates the extent of statistical heterogeneity by depicting a region within which, assuming the model is correct, we have 95% confidence that the true sensitivity and specificity of a future study will take place. Obviously, for Procalcitonin we find substantial heterogeneity.

When evaluating the diagnostic performance of a biomarker the area under the curve is of interest. To restrict the computation of the AUC to the observed false positive rates leads to the partial area under the curve (pAUC). This summary index is considered to be more practically relevant than the area under the entire ROC curve (AUC) because it avoids extrapolation. For the data at hand we obtain a pAUC equal to 0.629 and for completeness an AUC=0.799 indicating helpful diagnostic performance.

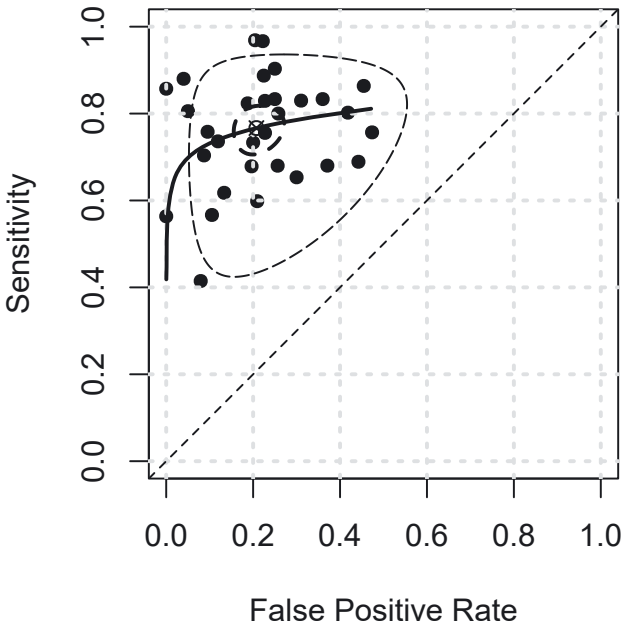

Figure 6: Procalcitonin as biomarker for sepsis: sROC curve (solid line). The point shows the joint estimate of sensitivity and 1-specificity together with a 95% confidence (dashed line) region and 95% prediction region (dotted line).

Using R

The freely statistical package R [13] may be used to perform the necessary calculations. The software can be obtained at https://cran.r-project.org. A useful integrated software environment is given by RStudio which is freely available for personal use: https://posit.co/. When using RStudio, R scripts can be used in order execute the relevant R commands. The following commands are found also as Supplementary Material in a file named.

`DTA_meta_analysis_tutorial.R`

Importing and manipulating data

Make sure you are working in the right directory. Please give the path to your directory where you save the file containing the data. For example:

```
setwd("M:/Gauss/schlatt/cclm/publi/meta-analysis")
```

The data from our example are read from an Excel .csv file and stored under the name 'PCT'. The command 'read.csv2' reads Excel files in .csv format. First comes the name of the file. Then 'header=T' implies that the first line contains the variable names.

```
PCT<-read.csv2("cclm_procalcitonin.csv",header=T)
```

The object 'PCT' contains the data and can be modified. For example the data column TP' contains the true positives as explained in Table 2.

The command 'attach' provides access to the individual elements of the data object 'PCT'.

In a first step we create a new variable called 'n1' which is a new column in our data set. To do this the syntax 'PCT$n1' is applied. Important, by using 'PCT$n1' a new column 'n1' is added to the dataframe 'PCT'. This variable contains the total number of diseased persons per study and is given as the sum of true positives TP and false negatives FN. In a similar way we create the variable 'n2', i.e. the total number of healthy individuals. The command 'head' shows the first six lines of the dataframe 'PCT'.

The symbol '#' indicates a comment which will not be executed by the program.

```
attach(PCT)
# calculate n1 (diseased persons) and create a new column named n1
# in the dataframe named PCT
PCT$n1<-TP+FN
# calculate n2 (healthy persons) and create a new column
PCT$n2<-FP+TN
# use attach again in order to make the newly created columns directly available
attach(PCT)
# calculate sensitivity and round to 3 digits
PCT$sens<-round(TP/n1,3)
# calculate specificity and round to 3 digits
PCT$spec<-round(TN/n2,3)
head(PCT)
```

	Study	Author	Year	TP	FP	TN	FN	Cut_off	n2	sens	spec
1	1	Ahmadinejad	2009	63	11	38	8	0.50	49	0.887	0.776
2	2	Al-Nawas	1996	73	45	170	49	0.50	215	0.598	0.791
3	3	Arkader	2006	12	0	14	2	2.00	14	0.857	1.000
4	4	Bell	2003	47	2	19	15	15.75	21	0.758	0.905
5	5	Castelli	2004	21	2	13	13	1.20	15	0.618	0.867
6	6	Clec'h	2006	29	2	38	7	1.00	40	0.806	0.950

Two univariate meta-analyses

Construction of forest plots sensitivity and specificity

Next we load the package 'mada' [36] and create a forest plot of sensitivity and specificity. First, we calculate basic measures of diagnostic accuracy and save it to the object 'PCT.d'. In case of zeros cells we do not make any corrections.

```
# load package 'mada'
library(mada)
# Calculate basic measures of diagnostic accuracy (sensitivity, specificity etc. for each study).
PCT.d<-madad(PCT,correction.control="none")
```

In the next step we construct a forest plot of sensitivity and specificity using the function 'forest' where we submit the object 'PCT.d' as an argument. Another argument is the type of plot. We start with sensitivity and thus we use type="sens". The plot for specificity is obtained in a similar way.

```
# forest plot of sensitivity and specificity side by side
old.par<-par()
plot.new()
par(fig=c(0, 0.5, 0, 1), new=TRUE)
forest(PCT.d, type="sens", xlab="Sensitivity", snames
=Author)
par(fig=c(0.5, 1, 0, 1), new=TRUE)
forest(PCT.d, type="spec", xlab="Specificity",snames
=Author)
par(old.par)
```

This code creates Figure 1. Since we want to show the plots side by side we store previous graphics environment parameters as 'old.par'. After finishing the plot we restore the previous graphics environment with 'par(old.par)'.

Meta-analysis for proportions

Next we apply the R package 'meta' [44]. This can be used to perform a meta-analysis treating sensitivity and specificity as proportions.

```
# Univariate meta-analysis with package meta
library(meta)
# Meta-analysis for sensitivity as a proportion
# Use function metaprop with true positives TP and total
number of diseased n1
m.sens<-metaprop(TP,n1,studlab=paste(Study,Year),
data=PCT)
# show result
summary(m.sens)
```

This gives the following truncated result:

```
                          proportion         95%-CI
Ahmadinejad 2009          0.8873             [0.7900; 0.9501]
Al-Nawas 1996             0.5984             [0.5058; 0.6861]
Arkader 2006              0.8571             [0.5719; 0.9822]
Bell 2003                 0.7581             [0.6326; 0.8578]
Castelli 2004             0.6176             [0.4356; 0.7783]
.....
Number of studies combined: k=31
Number of observations: o=1863
Number of events: e=1370

                          proportion         95%-CI
Common effect model       0.7354             [0.7149; 0.7549]
Random effects model      0.7683             [0.7201; 0.8103]

Quantifying heterogeneity:
tau^2=0.3637; tau=0.6031; I^2=69.8% [56.5%; 79.1%]; H=1.82
[1.52; 2.19]

Test of heterogeneity:
Q d.f.                    p-value            Test
99.40                     30<0.0001          Wald-type
127.46                    30<0.0001          Likelihood-Ratio

Details on meta-analytical method:
- Random intercept logistic regression model
- Maximum-likelihood estimator for tau^2
- Logit transformation
- Clopper-Pearson confidence interval for individual studies
```

The common effect model as shown is the model shown in Eq (1). The result is back-transformed as in Eq (2). Likewise the random effects refers to the model in Eq (3).

Logistic regression models

Alternatively, the functions 'glm' in order to calculate the parameters of the common effect model in Eq (1) and 'glmer' from the library 'lme4' of the random effects model in Eq (3) may be used. In a meta analysis we have for each study the number of true positives TP and the number of diseased n_1. This denominator needs to be taken into account. In R the combination of true positives TP and false negatives n_1-TP=FN builds the dependent variable. This is done using the command 'cbind(TP,FN)'. In order to perform a logistic regression model, we declare the dependent variable to follow a binomial distribution by using the command 'family=binomial()'.

```
# Common effect model (logistic regression)
# dependent variable is given by true positives and false
negatives
# Logistic regression intercept only model
sens.common<-glm(cbind(TP,FN) 1,family=binomial(),
data=PCT)
# show result
summary(sens.common)
glm(formula=cbind(TP, FN)  1, family=binomial(), data=PCT)

Deviance Residuals:
Min                 1Q        Median    3Q        Max
-4.3159             -1.2301   0.4208    1.5085    4.8411

Coefficients:
             Estimate  Std.      z       Pr(>|
                       Error     value   z|)
(Intercept)  1.02206   0.05252   19.46   <2e-
                                         16***
_
Signif.      0 '***'   0.001     0.01    0.05    0.1    1
codes:                 '**'      '*'     '.'     ' '
```

We can back-transform this result using the library 'emmeans' and the command 'lsmeans' where the stored result of model 'sens.common' is given as an argument. The second argument type="response" demands transformation of the result.to the original scale.

```
# Obtain estimates on the original scale together with a 95%
confidence interval
# library emmeans is required
library(emmeans)
lsmeans(sens.common, 1,type="response")
```

```
          lsmean    SE       df     asymp.LCL    asymp.UCL
overall   0.735     0.0102   Inf    0.715        0.755

Confidence level used: 0.95
Intervals are back-transformed from the logit scale
```

The random effects logistic regression model is obtained in a similar way using the function 'glmer' from the library 'lme4'. Additionally, we have to define random effects. This is done incorporating the additional term '(1| Study)' into the model which indicates the random effects following a normal distribution.

```
# Random effects logistic regression model for sensitivity
# library lme4' required
library(lme4)
sens.glmm<-glmer(cbind(TP,FN) 1+(1|
Study),family=binomial(),data=PCT)
#show result
summary(sens.glmm)
Generalized linear mixed model fit by maximum likelihood (Lap-
lace Approximation) ['glmerMod']

Random effects:
Groups              Name            Variance     Std.Dev.
Study               (Intercept)     0.36         0.6

Number of obs: 31, groups: Study, 31

Fixed
effects:
                Estimate   Std.       z        Pr(>|
                           Error      value    z|)
(Intercept)     1.1980     0.1284     9.327    <2e-16
                                                ***
_
```

The heterogeneity variance τ^2 is given as the variance of the random effects for the intercept and equals 0.36 as shown in Table 4. Again the overall sensitivity can be back transformed to the original scale using 'lsmeans'.

```
library(emmeans)'
# transform back to original scale
lsmeans(sens.glmm, 1,type="response")
```

```
1          lsmean    SE       df     asymp.LCL    asymp.UCL
overall    0.768     0.0229   Inf    0.72         0.81

Confidence level used: 0.95
Intervals are back-transformed from the logit scale
```

For specificity we proceed in a similar way (not shown).

Diagnostic odds ratio (DOR) and publication bias

For the calculation of the DOR and the assessment of publication bias we use again the library 'meta' with the function 'metabin'. Necessary arguments are the true positives TP, the number of diseased n_1, the false positives FP and the number of healthy subjects n_2.

Based on the stored results in 'm.dor' we show the results with 'summary(m.dor)' and construct the funnel plot shown in Figure 2.

Next we perform the regression test for publication bias:

Clearly, we find publication bias. In real life this needs to be investigated further, e.g. by a repeated literature search.

Bivariate meta-analysis

Plots in ROC space

We start with the R code necessary to create Figure 5. First, we use the command 'par(mfrow=c(1,2))' . This creates two plots in a row. Then we create a scatterplot using base R and next a cross hair plot which requires the libarary 'mada' [36].

```
# Diagnostic Odds Ratio (DOR)
# Use function 'metabin' from the package meta
# Arguments are true positives, number of diseased n1, fals positives FP and
# total number of healthy persons n2
m.dor<-metabin(TP,n1,FP,n2,studlab=paste(Author,Year),sm="DOR",data=PCT)
# show result
summary(m.dor)
Number of studies combined: k=31
Number of observations: o=3244
Number of events: e=1720
```

	DOR	95%-CI	z	p-value
Common effect model	8.1247	[6.8427; 9.6469]	23.91	<0.0001
Random effects model	11.6982	[8.3007; 16.4864]	14.05	<0.0001

```
Quantifying heterogeneity:
tau^2=0.5203 [0.2108; 1.4446]; tau=0.7213 [0.4592; 1.2019]
I^2=66.3% [50.9%; 76.9%]; H=1.72 [1.43; 2.08]
Test of heterogeneity:
```

Q	d.f.	p-value
89.00	30	<0.0001

```
# Publication bias
# Show funnel plot with DOR on the x-axis and 1/ESS^0.5 on the y-axis
funnel(m.dor)

# regression for publication bias.
metabias(m.dor)
Funnel plot test for diagnostic odds ratios
Test result: t=4.11, df=29, p-value=0.0003


Sample estimates:
```

bias	se.bias	intercept	se.intercept
18.4114	4.4803	0.5191	0.4709

```
Details:
- multiplicative residual heterogeneity variance
(tau^2=69.6852)
- predictor: inverse of the squared effective sample size
- weight: effective sample size
- reference: Deeks et al. (2005), J Clin Epid


# attach(PCT)
# Analyses in ROC space
# Show two plots a in a row
par(mfrow=c(1,2))
# scatter plot
```

(continued)

```
par(pty="s") # use square format
plot(1-spec,sens,xlim=c(0,1),ylim=c(0,1)
  ,xlab="False positive rate
(1-Specificity)",ylab="Sensitivity",pch=16)
# Crosshair plot
par(pty="s") # use square format
crosshair(PCT)
# restore to one plot per page
par(mfrow=c(1,1))
```

Bivariate logistic regression model with random effects

In order to use the bivariate logistic regression model with random effects we first need to transpose the data from 'wide' to 'long' format. Furthermore, we need new variables indicating disease status called 'healthy' and 'diseased'. Also we need new outcome variables "positive" for positive test results and "negative" vice versa. We can use the function 'reshape' where we create a new dataframe under the name 'long'. Next, the new variable 'healthy' as '1-diseased' is created and the data are sorted by study 'id'.

```
long<-reshape(PCT, direction="long", varying=list(c("TP" , "FP") , c("FN","TN" ) ) ,
  timevar="diseased" , times=c(1,0) , v.names=c("positive","negative") )
# create new variable "healthy"
long$healthy<-1-long$diseased
# sort by id
long<-long[order(long$id),]
```

Looking at the first six lines of the dataframe 'long' with the command 'head(long)' gives:

```
head(long)
```

Study	Author	Year	Cut_off	n1	n2	sens	spec	diseased	positive	negative	healthy
1	Ahmadinejad	2009	0.5	71	49	0.887	0.776	1	63	8	0
1	Ahmadinejad	2009	0.5	71	49	0.887	0.776	0	11	38	1
2	Al-Nawas	1996	0.5	122	215	0.598	0.791	1	73	49	0
2	Al-Nawas	1996	0.5	122	215	0.598	0.791	0	45	170	1
3	Arkader	2006	2.0	14	14	0.857	1.000	1	12	2	0
3	Arkader	2006	2.0	14	14	0.857	1.000	0	0	14	1

The data in long form contains the necessary information for the calculation of the bivariate random effects logistic model. Next, we use the function 'glmer' from the library 'lme4'. Now, our dependent variable is the combination of positive test results and negative test results in a matrix of the respective columns of the dataframe 'long' using.

```
cbind(positive,negative).
```

Now, apply the covariate "healthy" for healthy subjects coded '1' if true and '0' otherwise. This covariate quantifies the mean false positive rate on the logit scale. The covariate 'diseased' is coded '1' for diseased subjects and '0' otherwise and quantifies the mean true positive rate on the logit scale (i.e. sensitivity).
We do not want an intercept, thus our formula is '~0+healthy + diseased' for the fixed effects. For bivariate random effects we use '+(0+healthy + diseased| Study)'. Finally, we assume a binomial distribution thus family=-binomial). Hence the following code is applied and the result is stored as 'pct.glmm2'.

```
# Estimate parameters of the model
pct.glmm2<-glmer(cbind(positive,negative) 0+diseased+healthy+(0+diseased+healthy|Study),
          data=long, family=binomial )
# Show results
summary(pct.glmm2)
```

Using the command 'summary' with 'pct.glmm2' as an argument shows the results. Let's start with the covariance matrix of the random effects:

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
Family: binomial (logit )
Formula: cbind(positive, negative)  0 + diseased + healthy + (0 + diseased +
healthy | Study)
Random effects:
```

Groups	Name	Variance	Std.Dev. Corr
Study	diseased	0.3565	0.5971
healthy	0.3838	0.6195	0.23

We are interested in variability between studies, thus 'Groups name' refers to 'Study'. The variance equal to 0.3565 associated to 'diseased' is the variability between studies for the true positive rates σ_μ^2 in Eq (5). Likewise the variance equal to 0.3838 refers to the variability between studies σ_ν^2 for false positives rates. Finally, 'rho' denotes the correlation ρ in Eq (5). Next we look at the fixed effects:

```
Fixed effects:
```

	Estimate	Std. Error	z value	Pr(>|z|)
diseased	1.1892	0.1284	9.264	<2e-16 ***
healthy	−1.3395	0.1443	−9.280	<2e-16 ***

The estimate of the covariate 'diseased' is an estimate of β_0 equal to 1.1892. Likewise the estimate –1.3395 for the covariate healthy is an estimate of β_1 which denotes the false positive rate on the logit scale.

In order obtain the estimates on the original scale we use again the command 'lsmeans'

```
lsmeans(pct.glmm2, diseased,type="response")
diseased %in% healthy
```

diseased	healthy	lsmean	SE	df	asymp.LCL	asymp.UCL
1	0	0.767	0.0230	Inf	0.719	0.809
0	1	0.208	0.0237	Inf	0.165	0.258

The first line of the output with the covariates 'diseased' equal to 1 and healthy equal to '0' refers to sensitivity which is equal to 0.767. Likewise, the second line shows the false positive rate equal to 0.208. This completes the results shown in Table 5.

Summary receiver operator curve (sROC curve) with R

To the authors knowledge for the bivariate random effects logistic regression model no ready to use libraries or functions are available in R. Thus, for this article the function 'plot.sROC' was written. For a detailed description see the appendix.
The call of the function is simple.

```
# Create sROC curve with confidence and prediction ellipsoid
plot.sROC(PCT,pct.glmm2,conf=T,predict=T)
```

This function takes four arguments. The first one is the data set in wide format. It is mandatory, that e.g. true positives are named 'TP' in capital letters. The same applies for the other cells of the 2×2 table as denoted in Table 2. The next argument is the result of the bivariate random effects model. Submit here the current model. In our example the result of the model is stored under the name 'pct.glmm2'. Next 'conf=T' implies that a 95% confidence ellipsoid is desired and 'predict=T' implies the same for the 95% prediction region.

The function prints the area under the curve under the sROC curve as a result and creates a plot as shown in Figure 6.

Other software for meta-analysis of DTA studies

Admittedly, R is not very user friendly. The command line can be quite demanding to a beginner of R although the graphical user interface RStudio may help a bit. Thus, for an overview on alternatives which can be used for bivariate GLMM see Wang and Leefland [45]. For the commercially available software packages SAS and STATA the learning curve is similarly steep, but e.g. for SAS 'proc glimmix' a macro [46] and STATA the macro 'metadta' are available [47]. Alternatively, an interactive web based application called MetaDTA [48] could be used.

Acknowledgments: I would like to thank my friends from the Editorial Board of *CCLM* for the fun of editing *CCLM* together in the past years.
Research funding: None declared.
Author contributions: The author has accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: The author states no conflict of interest.
Informed consent: Not applicable.
Ethical approval: Not applicable.

Appendix: R code of the function plot.sROC

This function takes four arguments. The first one is the data set in wide format. It is mandatory, that e.g. true positives are named 'TP' in capital letters. The same applies for the other cells of the 2×2 table as denoted in Table 2. The next argument is the result of the bivariate random effects model. Submit here the current model. In our example the result of the model is stored under the name 'pct.glmm2'. Next 'conf=T' implies that a 95% confidence ellipsoid is desired and 'predict=T' implies the same for the 95% prediction interval.

If e.g. no prediction interval is wanted change this to 'predict=F'. The function prints the value of the extrapolated AUC and the partial pAUC based on the observed false positive rate as a result.

In order to use the function either mark the whole body of the function and use the 'run' button in RStudio. Alternatively, mark the whole body of the function and paste it into the R console.

The function is part of the Supplementary Material and the R script is named.

```
plot_sROC.R
# required packages
# package for generalized linear mixed models
library(lme4)
# post processing of model resuts
library(emmeans)
# needed for logit and expit transformation for plots in ROC space
library(rje)
plot.sROC<-function(data,model,conf=T,predict=T)
{
# calculate sensitivity
sens<-data$TP/(data$TP+data$FN)
# calculate false positive rate (1-specificity)
fpr<-data$FP/(data$TN+data$FP)
# find maximum of false positive rate on logit scale
max.fpr<-logit(max(fpr))
# find minimum of false positive rate on logit scale or set it closeto zero
min.fpr<-ifelse(min(fpr)<0.00025,logit(0.00025),logit(min(fpr)))
# extract regression coefficients
coef<-fixef(model)
# mean logit sensitivity (TPR)
eta<-coef[1]
# mean logit false positive rate
xi<-coef[2]
# variance covariance matrix of random effects
vc<-VarCorr(model)
# extract data and store as data.frame
# print(vc,comp=c("Std.Dev."))
temp<-as.data.frame(vc)
# covariance of random effects
cov<-temp$vcov[3]
# random effects variance of logit 1-specificity
varxi<-temp$vcov[2]
# random effects variance of logit sensitivity
vareta<-temp$vcov[1]
# save Variance-Covariance Matrix of random effects as matrix
Sigma<-matrix(c(vareta,cov,cov,varxi),nrow=2,byrow=T)
# sRoc curve eta on xi
# estimate slope of eta on xi regression line
beta<-cov/varxi
# estimate intercept of eta on xi regression line
alpha<-eta-cov/varxi*xi
# generate x axis: logit false positives from observed min to max
x<-seq(min.fpr,max.fpr,by=0.01)
# generate regression line in logit ROC-space
line<-alpha+beta*x
# total n of regression line
```

(continued)

```
nn<-length(line)
# transform to scale of TPR and FPR in ROC space
s<-expit(line)
# partial Area under the curve using trapezoidal rule for numerical integration
pAUC<-(s[1]/2 + sum(s[2:(nn-1)]) + s[nn]/2)/nn
# plot FPR and TPR together with sROC curve in ROC space
par(pty="s") # use a square plotting region
# plot FPR and TPR
plot(fpr,sens,pch=16,xlim=c(0,1),ylim=c(0,1), xlab="False Positive Rate",
ylab="Sensitivity")
# add grid
grid(lwd=2)
# show line where FPR equal to TPR (useless test)
abline(0,1,lty=2)
# plot summary estimates of FPR and TPR
points(expit(xi),expit(eta),pch=13)
# plot sROC curve in ROC space
lines(expit(x),s,lwd=2)
# confidence ellipsoid if desired (conf=T)
if(conf==T)
{
# extract variance covariance matrix of model coefficients (fixed effects)
rvar<-vcov(model)
# calculate correlation
r<-rvar[2,1][1,2]/(sqrt(rvar[1,1])*sqrt(rvar[2,2]))
# critical value Chi-Square distribution with two df
c<-sqrt(qchisq(0.95,2))
# generate values from zero to 2*pi (pi=3.1415.)
t<-seq(0,2*pi,0.001)
# y axis mean TPR+c* error*cos(t)
mueta<-eta+c*sqrt(rvar[1,1])*cos(t)
# axis mean FPR+c*standard error+cos(t+acos(r))
muxi<-xi+c*sqrt(rvar[2,2])*cos(t+acos(r))
# Transform to scale of sensitivity and specificity and plot SROC curve
lines(expit(muxi),expit(mueta),lwd=2,lty=2)
}
# prediction ellipsoid (if desired)
if(predict==T)
{
# create new matrix as sum of covariance matrix of coefficients
# and covariance matrix of random effects
rvar<-rvar+Sigma
# Same calculations as for the confidence ellipsoid
r<-rvar[2,1][1,2]/(sqrt(rvar[1,1])*sqrt(rvar[2,2]))
c<-sqrt(qchisq(0.95,2))
t<-seq(0,2*pi,0.001)
```

(continued)

```
mueta<-eta+c*sqrt(rvar[1,1])*cos(t)
muxi<-xi+c*sqrt(rvar[2,2])*cos(t+acos(r))
lines(expit(muxi),expit(mueta),lty=5)
}
# full AUC
x<-seq(logit(0.01),logit(0.99),by=0.01)
# generate regression line in logit ROC-space
line<-alpha+beta*x
# total n of regression line
nn<-length(line)
# transform to scale of TPR and FPR in ROC space
s<-expit(line)
# partial Area under the curve using trapezoidal rule for numerical integration
AUC<-(s[1]/2 + sum(s[2:(nn-1)]) + s[nn]/2)/nn
# print AUC and pAUC
cat("AUC=",AUC,"pAUC",pAUC,"\n")
}
```

References

1. Lippi G, Mattiuzzi C, Cervellin G. C-reactive protein and migraine. Facts or speculations? Clin Chem Lab Med 2014;52:1265–72.
2. Braga F, Pasqualetti S, Ferraro S, Panteghini M. Hyperuricemia as risk factor for coronary heart disease incidence and mortality in the general population: a systematic review and meta-analysis. Clin Chem Lab Med 2016;54:7–15.
3. Heilmann E, Gregoriano C, Wirz Y, Luyt CE, Wolff M, Chastre J, et al. Association of kidney function with effectiveness of procalcitonin-guided antibiotic treatment: a patient-level meta-analysis from randomized controlled trials. Clin Chem Lab Med 2021;59: 441–53.
4. Yang H, Gu Y, Chen C, Xu C, Xi Bao Y. Diagnostic value of pro-gastrin-releasing peptide for small cell lung cancer: a meta-analysis. Clin Chem Lab Med 2011;49:1039–46.
5. van Harten AC, Kester MI, Visser PJ, Blankenstein MA, Pijnenburg YAL, van der Flier WM, et al. Tau and p-tau as CSF biomarkers in dementia: a meta-analysis. Clin Chem Lab Med 2011;49:353–66.
6. Yu S, jie Yang H, qin Xie S, Bao YX. Diagnostic value of HE4 for ovarian cancer: a meta-analysis. Clin Chem Lab Med 2012;50:1439–46.
7. Agnello L, Vidali M, Giglio RV, Gambino CM, Ciaccio AM, Sasso BL, et al. Prostate health index (PHI) as a reliable biomarker for prostate cancer: a systematic review and meta-analysis. Clin Chem Lab Med 2022;60: 1261–77.
8. Lippi G, Henry BM, Adeli K. Diagnostic performance of the fully automated Roche Elecsys SARS-CoV-2 antigen electrochemiluminescence immunoassay: a pooled analysis. Clin Chem Lab Med 2022;60:655–61.
9. Ferraro S, Biganzoli EM, Castaldi S, Plebani M. Health Technology Assessment to assess value of biomarkers in the decision-making process. Clin Chem Lab Med 2022;60:647–54.
10. Oosterhuis WP, Niessen RWLM, Bossuyt PMM. The science of systematic reviewing studies of diagnostic tests. Clin Chem Lab Med 2000;38:577–88.
11. Cleophas TJ, Zwinderman AH. Meta-analyses of diagnostic studies. Clin Chem Lab Med 2009;47:1351–4.
12. Dahabreh IJ, Trikalinos TA, Lau J, Schmid C. An empirical assessment of bivariate methods for meta-analysis of test accuracy [internet]. Rockville, MD, USA: Agency for Healthcare Research and Quality; 2012.
13. R Core Team. R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/.
14. Fleischmann C, Scherag A, Adhikari NKJ, Hartog CS, Tsaganos T, Schlattmann P, et al. Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. Am J Respir Crit Care Med 2016;193:259–72.
15. Wacker C, Prkno A, Brunkhorst FM, Schlattmann P. Procalcitonin as a diagnostic marker for sepsis: a systematic review and meta-analysis. Lancet Infect Dis 2013;13:426–35.
16. Altman DG, Bland JM. Statistics notes: diagnostic tests 1: sensitivity and specificity. BMJ 1994;308:1552.
17. Schlattmann P. Statistics in diagnostic medicine. Clin Chem Lab Med 2022;31:801–7.
18. Vollset SE. Confidence intervals for a binomial proportion. Stat Med 1993;12:809–24.
19. Agresti A, Coull BA. Approximate is better than “exact” for interval estimation of binomial proportions. Am Statistician 1998; 52:119–26.
20. Schwarzer G, Carpenter J, Rücker G. Meta-analysis with R. Heidelberg, New York: Springer; 2014.
21. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. BMJ 1997;315:1533–7.
22. Sutton AJ, Higgins JP. Recent developments in meta-analysis. Stat Med 2008;27:625–50.
23. Schlattmann P. Medical applicatons of finite mixture models. Heidelberg, New York: Springer; 2009.
24. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med 2004;23:1351–75.

25. Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. Stat Med 2007;26:53–77.
26. Rucker G, Schwarzer G, Carpenter J, Olkin I. Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. Stat Med 2009;28:721–38.
27. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011;342:d549.
28. Senn S. Trying to be precise about vagueness. Stat Med 2007;26:1417–30.
29. Thompson S. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994;309:1351–5.
30. Simel DL, Bossuyt PMM. Differences between univariate and bivariate models for summarizing diagnostic accuracy may not be large. J Clin Epidemiol 2009;62:1292–300.
31. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003; 56:1129–35.
32. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58:882–93.
33. Altman DG, Bland JM. Statistics Notes: diagnostic tests 3: receiver operating characteristic plots. BMJ 1994;309:188.
34. Phillips B, Stewart LA, Sutton AJ. 'Cross hairs' plots for diagnostic meta-analysis. Res Synth Methods 2010;1:308–15.
35. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58:982–90.
36. Dobler P. mada: meta-analysis of diagnostic accuracy; 2022. R package version 0.5.11. Available from: https://CRAN.R-project.org/package=mada.
37. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. J Clin Epidemiol 2006;59:1331–2. author reply 1332–3.
38. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med 2002; 21:589–624.
39. Hamza TH, Reitsma JB, Stijnen T. Meta-analysis of diagnostic studies: a comparison of random intercept, normal-normal, and binomial-normal bivariate summary ROC approaches. Med Decis Making 2008; 28:639–49.
40. Rosenberger KJ, Chu H, Lin L. Empirical comparisons of meta-analysis methods for diagnostic studies: a meta-epidemiological study. BMJ Open 2022;12:e055336.
41. Chappell FM, Raab GM, Wardlaw JM. When are summary ROC curves appropriate for diagnostic meta-analyses? Stat Med 2009;28: 2653–68.
42. Salameh JP, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. BMJ 2020;370:m2632.
43. Arends LR, Hamza TH, van Houwelingen JC, Heijenbrok-Kal MH, Hunink MG, Stijnen T. Bivariate random effects meta-analysis of ROC curves. Med Decis Making 2008;28:621–38.
44. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Base Ment Health 2019;22: 153–60.
45. Wang J, Leeflang M. Recommended software/packages for meta-analysis of diagnostic accuracy. J Lab Precis Med 2019;4:22.
46. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. Methods Inf Med 2010; 49:62–4.
47. Nyaga VN, Arbyn M. Metadta: a Stata command for meta-analysis and meta-regression of diagnostic test accuracy data – a tutorial. Arch Publ Health 2022;80:95.
48. Freeman SC, Kerby CR, Patel A, Cooper NJ, Quinn T, Sutton AJ. Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA. BMC Med Res Methodol 2019;19:81.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/cclm-2022-1256).