Statistical packages for diagnostic meta-analysis and their application

Philipp Doebler, Paul Bürkner and Gerta Rücker

Abstract The bivariate model has become a de facto standard in diagnostic meta-analysis. Complex iterative algorithms are needed to fit the model and thus a meta-analysis of diagnostic accuracy data is much aided by appropriate software packages. Also, graphical methods ease exploration, interpretation and communication in the context of a diagnostic meta-analysis. This chapter reviews existing software and discusses the relative merits of general packages and specialized packages for DTA meta-analysis. The use of software for diagnostic meta-analysis and especially fitting the bivariate model is illustrated with a sample workflow in the open-source statistical framework R. Some ways to extend the bivariate model and software for the case of multiple cut-off values per primary study are discussed.

Key words: diagnostic test accuracy, meta-analysis, software, bivariate model, meta-regression, SROC-curve, descriptive statistics

1 Introduction

The vast majority of meta-analytic approaches are intended for the analysis of univariate effect size measures. As a consequence, most software packages focus on meta-analytic techniques for univariate data, e.g. tools like RevMan (The Nordic Cochrane Centre, 2014) or R-packages like `meta` (Schwarzer, 2007). However, the

Philipp Doebler
Department of Statistics – TU Dortmund University. e-mail: doebler@statistik.tu-dortmund.de

Gerta Rücker
Institute for Medical Biometry and Statistics, Faculty of Medicine, Medical Center – University of Freiburg. e-mail: ruecker@imbi.uni-freiburg.de

Paul Bürkner
Institute of Psychology – Faculty of Psychology and Sport Sciences – WWU Münster. e-mail: paul.buerkner@uni-muenster.de

accuracy of a diagnostic test is typically evaluated in a positive and a negative arm relative to a gold standard (**thisvolumechapterXYZ**) and thus produces two endpoints, the sensitivity and false positive rate (**thisvolumechapterZapf**) calculated from a 2×2-table. While such a table can be boiled down to a univariate effect size and meta-analysed (e.g. Glas, Lijmer, Prins, Bonsel, & Bossuyt, 2003), it is recommended to employ the Reitsma et al. (2005) bivariate model that simultaneously analyse the reported pairs of sensitivity and false positive rate (Leeflang, Deeks, Gatsonis, & Bossuyt, 2008; Macaskill, Gatsonis, Deeks, Harbord, & Takwoingi, 2010). Alternatively, the HSROC-model of Rutter and Gatsonis (2001) can be used, which is equivalent to the bivariate model in the absence of covariates (Harbord, Deeks, Egger, Whiting, & Sterne, 2007). Regardless whether the bivariate model of the HSROC-model is employed, complex iterative algorithms are needed, so in contrast to, say, a DerSimonian and Laird (1986) meta-analysis, a spreadsheet program is not an option.

As a consequence, the meta-analysis of diagnostic test accuracy (DTA) requires specialized packages. Another reason to employ such packages are powerful graphical techniques like summary receiver operating characteristic (SROC) curves (**thisvolumechapterZapf**), which add substantially to a diagnostic meta-analysis in the presence of different (implicit or explicit) cut-off values. Note that packages for bivariate and multivariate meta-analyses such as the R-packages `metafor` (Viechtbauer, 2010) or `mvmeta` (Gasparrini, Armstrong, & Kenward, 2012) do not include functionality to produce SROC curves.

The aim of this chapter is to inform the diagnostic meta-analyst of software options and to present a workflow in R (R Core Team, 2017) with some detail including computer code. Others have recently contributed similar work to aid DTA meta-analysts in the analysis stage: A book chapter by Schwarzer, Carpenter, and Rücker (2015) discusses in detail how to perform an analysis in R and several recent tutorials and/or reviews exist in the medical literature Liu et al. (2013), Kim, Lee, Choi, Huh, and Park (2015), Lee, Kim, Choi, Huh, and Park (2015). The contribution by Macaskill et al. (2010), is worth mentioning for authors of Cochrane Reviews.

The remainder of this chapter is structured as follows: After an overview of existing packages and short discussion of their strength and weaknesses, techniques are presented for descriptive statistics in Section 2, the fitting of the bivariate model in Section 3, including the calculation and plotting for SROC-curves. A brief discussion wraps up the chapter and hints at computer code for advanced methods.

1.1 Overview of software packages

Table 1 contains an overview of software packages for diagnostic meta-analysis. The selection is based on packages that allow to fit the bivariate model and includes specialized packages for meta-analysis of DTA as well as general packages with capabilities for multivariate meta-analysis. The table omits discontinued packages and

those only suited for outdated approaches like the Moses-Littenberg SROC curve (e.g. RevMan[1], MetaDiSc). In addition to the packages found in Table 1 there are other packages that allow to fit generalized linear models (and hence the bivariate model as a special case), but we omit them as we are not aware that they have been referenced for this purpose in the literature. In addition, there are (R-)packages for special variants of the bivariate models not listed here (Schiller & Dendukuri, 2015; Nikoloulopoulos, 2016; Verde, 2017).

We caution the reader, that all general packages will require more effort in implementing the bivariate model and producing output specific for the DTA context such as SROC-curves. There is hence a trade-off between extensibility of the software and convenient use. Clearly, the diagnostic meta-analyst will have to balance these two factors, as extensibility comes with more time and effort in implementation or even requires programming skills. In the Discussion, we reference packages and computer code for some specialized analysis methods.

2 Sample workflow in R

Since all three authors of this chapter are biased towards R, we present a fairly detailed worked example with R code in the following section[2] For this purpose we mainly use the R package `mada` (Doebler, 2015), a specialized package for DTA meta-analysis.[3]

We begin by advising on the first steps of an analysis of DTA meta-analysis data. We show ways to import data and then demonstrate descriptive techniques that might be useful prior to an analysis with the bivariate model. We use selected variables from a dataset originally reported in Patrick et al. (1994) on the diagnostic accuracy of interviewer or self-administered questionnaires to detect smoking relative to biochemical goldstandards.

2.1 Importing data into R

After coding data (**thisvolumechapterXYZ**), the analyst obtains a raw data file. Importing data into R is often made easier by employing graphical user inter-

[1] We mention in passing that RevMan can plot SROC curves if suitable output is supplied from other packages.

[2] Example code for other packages can typically be found in the references in Table 1 or in the technical documentation.

[3] The package can be installed by typing `install.packages("mada")` at an R-prompt. After this, the package only needs to loaded once in an R session with `library(mada)`. The most current (development) version of `mada` is found at http://r-forge.r-project.org/projects/mada/. Some additional functionality of `mada` is explained in the package vignette that is automatically installed with the package and can be accessed by typing `vignette("mada")` at an R-prompt.

Table 1 Current software packages for diagnostic meta-analysis that include the bivariate model.

Statistical Framework	Package/Macro	Features and Notes	Reference(s)
		Open-Source Software	
BUGS-language	`WinBUGS`, `OpenBUGS`	general statistical package, Bayesian, extensible, programming needed	Lunn, Thomas, Best, and Spiegelhalter (2000), Lunn, Spiegelhalter, Thomas, and Best (2009)
	`jags`, `rjags`	general statistical package, Bayesian, extensible, programming needed	Plummer (2016)
R	`brms`	general mixed model package, Bayesian, extensible, implementation needed	Bürkner (2017)
	`lme4`	general mixed model package, extensible, implementation needed	Partlett and Takwoingi (2016)
	`mada`	specialized package for DTA meta-analysis, LMM-approximation to bivariate model, graphical methods	Doebler (2015)
	`meta4diag`	specialized package for DTA meta-analysis, Bayesian, graphical methods	Guo and Riebler (2016)
	`metafor`	general univariate and multivariate meta-analysis package, implementation needed	Viechtbauer (2010)
	`Metatron`	specialized package for DTA meta-analysis, multinomial processing tree models for imperfect gold standards	Huang (2014), Botella, Huang, and Suero (2013)
	`mvmeta`	general multivariate meta-analysis package, implementation needed	Gasparrini, Armstrong, and Kenward (2012)
		Proprietary Software	
MLwiN	-	general mixed model package	Charlton, Rasbash, Browne, Healy, and Cameron (2017)
SAS	`Proc NLMIXED`	general mixed model functions, implementation needed	Chu and Cole (2006), Arends et al. (2008)
	`Proc GLIMMIX`	general mixed model functions, implementation needed	Menke (2010)
	`METADAS`	specialized package for DTA meta-analysis	Takwoingi and Deeks (2011)
Stata	`glamm`	general mixed model package, implementation needed	Rabe-Hesketh, Skrondal, and Pickles (2004)
	`metandi`	specialized package for DTA meta-analysis, graphical methods	Harbord and Whiting (2010)
	`meqrlogit` (`xtmelogit`)	general binary mixed model package, implementation needed	Takwoingi (2016)
	`midas`	specialized package for DTA meta-analysis, graphical methods	Dwamena (2007)

faces (GUIs) like RStudio. Depending on the source of the data, the preinstalled R-package `foreign` can be helpful (say to read SPSS files) or the R-package `readxl` (for Microsoft Excel files; Wickham & Bryan, 2017). Typically a `data.frame` is obtained, i.e., an R-object containing data of different types (especially numerical and categorical data).

Some rows of the Patrick et al. (1994) smoking data are shown in Table 2. From some of the original primary studies more than one 2×2-table could be reasonably coded, since authors reported results for multiple samples, multiple screening tests or multiple gold standards. This corresponds to more than one row for some studies. Also, the data set does not contain the sensitivities and false-positive rates originally reported in some studies, but the (reconstructed) frequencies from the underlying 2×2-table. The chapter by **thisvolumechapterXYZ** discusses how to obtain them during coding. In the following, we will assume that at least the four columns `TP`, `FN`, `FP` and `TN` are present in the data, corresponding to the frequencies of true positives, false negatives, false positives and true negatives respectively.

Table 2 Selected rows of the Patrick et al. (1994) smoking data.

row	author	study_id	type	TP	FN	FP	TN	population
1	Bauman & Dent, 1982	1	SAQ	21	15	28	324	S
2	Bauman & Dent, 1982	1	SAQ	90	10	120	969	S
3	Bauman & Dent, 1982	1	SAQ	104	8	26	232	G
4	Bauman & Dent, 1982	1	SAQ	332	18	92	673	G
5	Bauman et al, 1982	2	SAQ	3	0	2	77	S
6	Bauman & Koch, 1983	3	SAQ	437	23	78	901	G
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
46	Vogt et al, 1977	23	IAQ	83	2	11	43	G
47	Vogt et al, 1977	23	IAQ	71	3	7	42	G
48	Vogt et al, 1977	23	IAQ	76	3	18	42	G
49	Wagenknecht et al, 1990	24	IAQ	1357	185	68	3322	G
50	Wald et al, 1981	25	IAQ	1649	17	423	6632	G
51	Williams et al, 1979	26	SAQ	19	2	1	96	S

Note. SAQ=self-administered questionnaire; IAQ=interviewer-administered questionnaire; S=student; G=general.

2.2 Calculating summary statistics for each study

thisvolumechapterZapf discusses a range of useful summary statistics of diagnostic accuracy. The `madad` function can be conveniently used to calculate these. Note that by default a continuity correction of 0.5 is added to all cells, in case there is a zero cell in any 2×2-table. Confidence intervals are Wilson score intervals.

```
 KeywordToklibrary(mada) # load the mada package for this session
# In this example, an example data.frame named smoking
# with several variables used.
data(smoking) # make data available
# Many of the following commands assume that
# the data.frame contains variables for the
# frequencies named TP,FN,FP and TN. If not,
# the syntax has to be modified (see the manual).
descr <- madad(smoking) # includes continuity correction!
print(descr, digits = 2) # print lengthy results

## Descriptive summary of smoking with 51 primary studies.
## Confidence level for all calculations set to 95 %
## Using a continuity correction of 0.5 if applicable
##
## Diagnostic accuracies
##       sens 2.5% 97.5% spec 2.5% 97.5%
##  [1,] 0.58 0.42  0.72 0.92 0.89  0.94
##  [2,] 0.90 0.82  0.94 0.89 0.87  0.91
##  [3,] 0.92 0.86  0.96 0.90 0.85  0.93
## ...
## Test for equality of sensitivities:
## X-squared = 1569.401, df = 50, p-value = <2e-16
## Test for equality of specificities:
## X-squared = 1320.466, df = 50, p-value = <2e-16
##
## Diagnostic OR and likelihood ratios
##           DOR   2.5%    97.5%  posLR  2.5%   97.5% negLR 2.5% 97.5%
##  [1,]   15.79   7.41    33.67   7.20  4.61   11.24  0.46 0.31  0.67
##  [2,]   69.35  35.61   135.04   8.11  6.76    9.71  0.12 0.07  0.21
##  [3,]  107.86  48.16   241.58   9.04  6.28   13.01  0.08 0.04  0.16
## ...
## Correlation of sensitivities and false positive rates:
##    rho  2.5 % 97.5 %
##   0.27   0.00   0.51
```

Note that in addition to the sensitivity and specificity, χ^2-tests of equality are calculated, which typically confirms the presence of substantial heterogeneity in DTA meta-analysis data. We omit the discussion of the diagnostic odds ratios and the positive and negative likelihood ratios also resulting from a `madad` call and refer to Chapter XYZ of this volume for details on these statistics. Sometimes it is convenient to use output of R functions in subsequent caluclations:

```
# if you need to work with (part of) the output,
# check the structure:
str(descr)
## List of 17
##  $ sens              :List of 2
##   ..$ sens   : num [1:51] 0.581 0.896 0.925 0.947 0.875 ...
##   ..$ sens.ci: num [1:51, 1:2] 0.422 0.821 0.861 0.919 0.396 ...
##   .. ..- attr(*, "dimnames")=List of 2
##   .. .. ..$ : NULL
```

```
##   .. .. ..$ : chr [1:2] "2.5%" "97.5%"
## [...]

# From the structure, we see the list-structure
# and can use it to extract parts of the output:
descr$sens$sens # extract vector of sensitivities

##  [1] 0.58108108 0.89603960 0.92477876 0.94729345 0.87500000
## [...]

# redo calculations without continuity correction:
descr0 <- madad(smoking, correction = 0) # output omitted
```

The last line shows how to omit the continuity correction if desired (e.g., to reproduce original results).

2.3 Graphical techniques

Patterns can often be much more easily recognized from graphical representations of data than from tables. Pairs of sensitivity and false-positive rate should be plotted at some point of the analysis. In addition to the point estimates, their uncertainty is of interest. Especially outliers with large standard errors might otherwise influence the perception of the data.

Next we show how to produce a paired forest plot as well as a 'cross hairs' plot (Phillips, Stewart, & Sutton, 2010) and a plot with confidence ellipses. To prevent large and/or cluttered plots, we use an (essentially arbitrary) subset of the smoking data here for didactic purposes[4].

```
# First reduce to a subset of with independent 2x2-tables:
smoking1 <- subset(smoking, smoking$result_id == 1)
# Reduce further to a random (and essentially arbitray) subset of
# ten studies to prevent a cluttered plot:
set.seed(12345) # fix random number seed for reproducibilty
smoking1 <- subset(smoking1[sample(1:nrow(smoking1), 10), ])
smoking1 <- smoking1[order(smoking1$author), ] # reorder

# make forest plots
descr1 <- madad(smoking1) # data for forest plots
mynames <- smoking1$author # vector of names for forest plot
forest(descr1, "sens", snames = mynames, main = "Sensitivity")
forest(descr1, "spec", snames = mynames, main = "Specificity")

# make crosshair plot:
crosshair(smoking1, pch = ifelse(smoking1$type == "IAQ", 1, 2),
          col = ifelse(smoking1$population == "G", 1, 2),
          cex = 1.5)
```

[4] Note that the `subset` function is a convenient way in R to form subsets.

```
legend("bottomright", c("IAQ", "SAQ"), pch = 1:2, cex = 1.5)
legend("topright", c("general population", "student population"),
       pch = 15, col = 1:2, cex = 1.5)

# make ROC-ellipse plot
ROCellipse(smoking1, pch = ifelse(smoking1$type == "IAQ", 1, 2),
           col = ifelse(smoking1$population == "G", 1, 2),
           cex = 1.5)
legend("bottomright", c("IAQ", "SAQ"), pch = 1:2, cex = 1.5)
legend("topright", c("general population", "student population"),
       pch = 15, col = 1:2, cex = 1.5)
```

Figure 1 contains the corresponding output. Note that the variables type and population have been used to set color and shape of symbols, so that trends resulting from these covariates might be recognized.

Here, one learns about outliers (easily observed in the sensitivity forest plot or in the crosshair and ellipse plots) and heterogeneity (non-intersecting confidence ellipses). Also, the use of color reveals that diagnostic accuracy is not only determined by the underlying population, while the symbols indicate that IAQs (at least in the arbitrary subset) are more accurate, since they cluster in the top left corner.

2.4 Fitting the bivariate model

The bivariate model of Reitsma et al. (2005) has been introduced in Chapters XYZ and XYZ of this volume. As the de facto standard in DTA meta-analysis, fitting it deserves special attention in this chapter. Recall that there are (in the absence of covariates) five parameters of this model: The logit-transformed mean sensitivity and false positive rate, their between study variances (again on logit-scale) and the between study covariance (or equivalently the between study correlation). The interpretation of these parameters is covered in the chapters by **thisvolumechapterbuerkner** and **thisvolumechapterZapf** All packages mentioned in Table 1 can estimate these parameters, with a variety of algorithms. Roughly, the algorithms can be subdivided into frequentist algorithms, which are typically based on the maximum-likelihood principle, and Bayesian approaches, which lead to Markov-Chain-Monte-Carlo techniques (MCMC). In this section, we provide examples of software using both types of algorithms without going into their technical foundations.

The bivariate model assumes that independent 2×2-tables are available. Since there are multiple rows for some of the studies in the smoking dataset we only analyze the very first 2×2-table from each study subsequently, but hint how to overcome this restriction at the end of this section. Also, we reduce the dataset further to include only the self-administered questionnaire (SAQ) data.

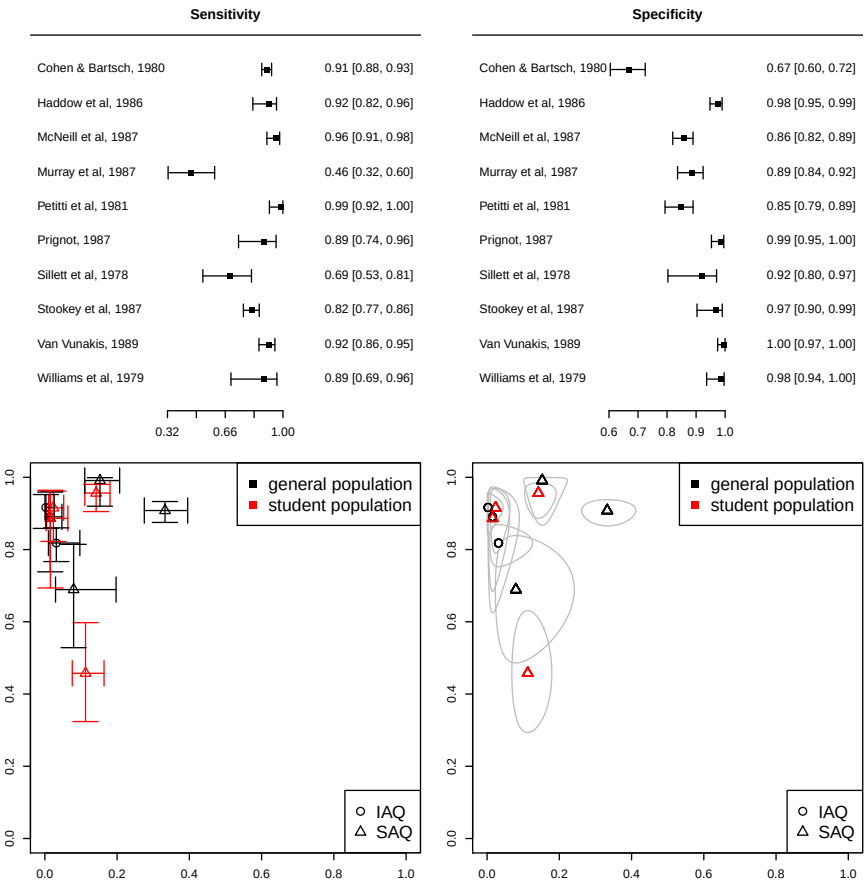

Fig. 1 Paired forest plots (top), 'cross hairs' plot (bottom left) and confidence ellipses plot (bottom right) for a subset of the smoking data.

2.5 Fitting the bivariate model without covariates

We now show two ways to fit the bivariate model in R. The `reitsma` function from the R-package `mada` implements a linear mixed model approximation to the bivariate model, which parallels the implementation with SAS Proc MIXED by Reitsma et al. (2005) with restricted maximum likelihood estimation (REML). Chu and Cole (2006) caution, that this approximation is slightly biased can be improved upon by fitting a generalized linear mixed model, a point made more precise in the recent simulation study of Richter, Schlattmann, and Dewey (to appear). In R, the `fit.bivar` function from the R-package `Metatron`, which implements the generalized linear mixed model version of the bivariate model, gives similar results as SAS Proc NLMIXED. Both R-functions discussed here need data from 2×2-tables as in Table 2. After fitting the model, a summary is produced, which we annotate:

```
# smoking2 is to contain only data for the SAQs from independent 2x2-tables:
smoking2 <- subset(smoking, smoking$result_id == 1 & smoking$type == "SAQ")
library(mada) # LMM-approximation to the bivariate model
smoking2 <- smoking[smoking$type == "SAQ",]
# if the dataset contains columns names TP,FN,FP and TN, use
fit1 <- reitsma(smoking2)
summary(fit1) # detailed output
## Call:  reitsma.default(data = smoking2)
##
## Bivariate diagnostic random-effects meta-analysis
## Estimation method: REML
```

First, we learn what the input was (which is more useful if covariates are added to the model) and that REML-estimation was performed (by default, some other estimators are available). We then see estimates of the fixed effects of the model, which are the logit-transformed sensitivity and false positive rate:

```
## Fixed-effects coefficients
##              Estimate Std. Error   z Pr(>|z|) 95%ci.lb 95%ci.ub
##
## Fixed-effects coefficients
##                   Estimate Std. Error      z Pr(>|z|) 95%ci.lb 95%ci.ub
## tsens.(Intercept)     1.68       0.47   3.56     0.00     0.76     2.61
## tfpr.(Intercept)     -2.46       0.24 -10.27     0.00    -2.93    -1.99
## sensitivity           0.84          -      -        -     0.68     0.93
## false pos. rate       0.08          -      -        -     0.05     0.12
##
## tsens.(Intercept) ***
## tfpr.(Intercept)  ***
## sensitivity
## false pos. rate
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Confidence intervals and asymptotic Wald-tests indicate here that the logit transformed accuracy parameters are significantly different from 0, which means the pooled sensitivity and false positive rate are signficantly different from .50. These backtransformed estimates are available in two extra lines. Note that for reasons of formatting, the significance codes for the fixed effects occupy four additional lines in this output, where the three stars for sensitivity and false positive rate indicate $p < 0.001$ and the backtransformed parameters do not have any stars, as no inference is performed. The output then contains the standard deviations of the random effects and an estimate of the correlation (.50 here), followed by the log-likelihood and fit measures:

```
## Variance components: between-studies Std. Dev and correlation matrix
##       Std. Dev tsens  tfpr
## tsens     1.80  1.00    .
## tfpr      0.90  0.50 1.00
##
```

```
##  logLik   AIC    BIC
##  32.82 -55.64 -48.31
```

Note that the log-likelihood includes terms for the Jacobian of the logit-transformation, which might differ from implementations with SAS Proc MIXED. For further details see Doebler, Holling, and Böhning (2012). The remainder of the output contains an estimate of the area under the SROC-curve (AUC) with a value of .949, which is close to optimal (though the partial AUC of .869 is a bit more modest). More details on SROC-curves follow in the next section. The SROC-curve is calculated based on the parametrization of the HSROC-model and the parameters of this model are also given:

```
## AUC:  0.949
## Partial AUC (restricted to observed FPRs and normalized):  0.869
##
## HSROC parameters
##       Theta      Lambda        beta sigma2theta sigma2alpha
##       -1.14        4.67       -0.69        1.21        1.62
```

In sum, all parameters of the bivariate model are found in the output of `summary(fit1)`: The pooled logit-transformed sensitivity and false positive rate are found in the `Estimate` column. For convenience, also the back-transformed values are given (.84 and .08 here). The between study standard deviations of the random effects follow (1.80 and 0.90 here) together with their correlation (.50 here), from which the covariance can be computed if necessary.

2.6 SROC-curves

In the majority of areas of application of DTA meta-analysis, explicit or implicit cut-off values are used to dichotomize the result of the screening test. Authors of primary studies choose cut-off values to compromise between false-positive rate and sensitivity. On the level of the primary studies, the curve representing the different trade-offs of false-positive rate vs. sensitivity, is known as the receiver operating characteristic (ROC) curve.

Variation in the cut-off leads to different pairs of false-positive rate and sensitivity, even if the primary studies were otherwise equal and hence to (apparent) heterogeneity on the meta-analytic level. As a consequence summary ROC (SROC) curves are of special interest in DTA meta-analyses. From the parameters of the bivariate model SROC-curves can be computed. As a default, we recommend the SROC-curve suggested by Rutter and Gatsonis (2001) for the HSROC model. A straightforward way to plot this SROC-curve in R is a simple call of `plot`: If `fit1` is an object produced by the `reitsma` function, then `plot(fit1, predict = TRUE)` produces the SROC-curve and a prediction region.

Figure 2 displays the output of the call to `plot`: We see the pair of pooled accuracies together with a 95%-confidence region, the analogon of a 95%-confidence

interval in the bivariate case. Also, a 95%-prediction region is displayed (dashed line). This can be interpreted as follows: If a new study was to be performed, its pair of sensitivity and false-positive rate would end up in the prediction region with a probability of 95%.

Next to the SROC-curve discussed here, other SROC-models exist. We will come back to software pckages for older and more current models in the Discussion.

2.7 The bivariate model with covariates

If the DTA meta-analyst's interest is in the influence of a covariate on the diagnostic accuracy, an extension of the bivariate model is needed. We use the general term covariate here to include the categorical case and the continuous case. Examples of categorical covariates include screening test type and population (sub-)type, while mean age and publication year are typically treated as continuous covariates.

Chapter XYZ of this volume by **thisvolumechapterZapf** should be consulted for a detailed specification of the technical details of the extension by covariates. In a nutshell, a regression of the logit-transformed sensitivity and/or false-positive rate on the covariate(s) is added to the model. Estimating such a model results in separate (fixed) regression coefficients for sensitivity and false-positive rate. Significance tests for the regression coefficients are a useful by-product. Some expertise with regression modelling is helpful when conducting a DTA meta-analysis with covariates and in fact, the bivariate model with covariates is an example of a multivariate meta-regression.

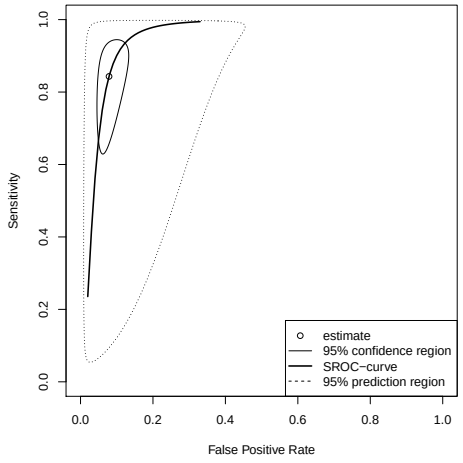

Fig. 2 Graphical representation of a bivariate model fit.

Note that the multivariate meta-regression discussed here, similar to its univariate counterpart, assumes that the between study covariance of the random effect is the same for all combinations of the covariates. For the case of a single categorical covariate, i.e. subgroups, this implies that the between study covariance is identical in all subgroups. This assumption can be relaxed with subgroup-specific between study covariance, but we do not discuss this case here.

All packages in Table 1 are capable of fitting the bivariate model with covariates. A relatively compact syntax can be used in R, so we demonstrate this for the smoking data. We use the categorical moderator questionnaire type with levels interviewer-administered and self-administered (IAQ and SAQ), so that differences in diagnostic accuracy for these two types of screening measures can be studied.

```
# smoking3 is a subset of the smoking data with independent 2x2-tables:
smoking3 <- subset(smoking, smoking$result_id == 1)
fit_type1 <- reitsma(smoking3, formula = cbind(tsens,tfpr) ~ type)
```

Again, fitting does not produce output right away but a detailed summary is produced by calling `summary`:

```
summary(fit_type1)

## Call:  reitsma.default(data = smoking3,
##                        formula = cbind(tsens, tfpr) ~ type)
##
## Bivariate diagnostic random-effects meta-analysis
## Estimation method: REML
```

Again we learn what the input was and that REML estimation was performed (by default). The output continues by estimates of the logit-transformed sensitivity for the IAQ studies (which are represented by the model's intercept term) and the regression coefficients for SAQ are then interpreted as log odds ratios. The DTA meta-analyst could consider to backtransform the log odds ratios for readers more familiar with odds ratios, say as in the regressions tabulated by Karrasch et al. (2017).

```
## Fixed-effects coefficients
##                Estimate Std. Err.   z Pr(>|z|) 95%ci.lb 95%ci.ub
## tsens.(Intercept) 2.81   0.49   5.74    0.00    1.85    3.78
## tsens.typeSAQ    -1.17   0.63  -1.84    0.07   -2.41    0.08
## tfpr.(Intercept) -3.34   0.31 -10.73    0.00   -3.95   -2.73
## tfpr.typeSAQ      0.88   0.39   2.27    0.02    0.12    1.65
##
## tsens.(Intercept) ***
## tsens.typeSAQ       .
## tfpr.(Intercept)  ***
## tfpr.typeSAQ        *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The covariance matrix of the random effects an fit measures then follow, similar to the case without covariates. Note that not parameters of an SROC-curve are reported, as there is, at least in general, not a unique curve in models with covariates[5]:

```
## Variance components: between-studies Std. Dev and correlation
## matrix
##       Std. Dev tsens  tfpr
## tsens    1.508 1.000     .
## tfpr     0.875 0.551 1.000
##
##   logLik      AIC      BIC
##   70.721 -127.441 -113.783
```

2.8 Fitting strategies for advanced models

The bivariate model is a special case of a generalized linear mixed model (GLMM; e.g. Demidenko, 2013; Brown & Prescott, 2015), a type of regression model that includes fixed and random effects. From the GLMM perspective, a range of extensions of the bivariate model are possible, including multiple cutoff values per study (**thisvolumechapterZapf**) or trivariate extensions including the observed prevalence (Chu, Nie, Cole, & Poole, 2009; Hoyer & Kuss, 2015; Nikoloulopoulos, to appear). As a discussion of all extensions is beyond the scope of this chapter, only an example implementation of the bivariate model as a GLMM is provided.

The multi-purpose R-package `brms` is used in our sample implementation, as it is a (relatively) convenient open source alternative to commercial packages for GLMMs (Bürkner, 2017) and also because its Bayesian approach to parameter estimation allows to include prior information, which is in contrast to frequentist packages like `lme4` (Bates, Mächler, Bolker, & Walker, 2015). The following code presents an example of an analysis of the smoking data in `brms` with and without covariates[6].

First, the data is rearranged, so that each study fills two rows: one for the positive arm of the study (i.e. `condition = "yes"`) and the other for the negative arm. Many GLMM packages expect data arranged in this fashion:

```
nstudy <- nrow(smoking3)
# convert data to long format:
smoking3_long <- with(smoking3,
  data.frame(P = c(TP, FP), N = c(FN, TN),
             condition = rep(c("yes", "no"), each = nstudy),
             type = rep(type, 2),
```

[5] Note that for special cases like a binary covariate, ploting SROC-curves for the parameters corresponding to each of both levels of the covariates is meaningful. For an example see Meyer, Frings, Rücker, and Hellwig (2017).

[6] Note that `brms`'s syntax is very similar to `lme4`'s so that the sample code below can be adapted. For similar `lme4` code, also consult Partlett and Takwoingi (2016)

```
              study = rep(1:nstudy, 2))
)
smoking3_long$total <- with(smoking3_long, P + N)
```

Next we load the package and the bivariate model is fitted with and without the `type` covariate.

```
library(brms) # load brms package
# fit a GLMM corresponding to the bivariate model
fit <- brm(P | trials(total) ~ 0 + condition +
            (0 + condition | study),
           data = smoking3_long, family = binomial())
summary(fit) # obtain model parameters
# produce plot of marginal effects:
marginal_effects(fit, conditions = list(total = 1))

# add study type as a covariate (IAQ vs. SAQ)
fit_type <- brm(P | trials(total) ~ 0 + condition +
                  condition:type +(0 + condition | study),
                data = smoking3_long, family = binomial())
summary(fit_type) # check influence of covariate
# plot marginal effects for type:
marginal_effects(fit_type, "condition:type",
                  conditions = list(total = 1))
```

Parameter estimates similar to `mada`'s result (not shown), and the discrepancies are a consequence of `brms`' Bayesian approach and, more importantly, the fact that `mada` uses a linear approximation to a GLMM. The `marginal_effects` function produces a graphical display of the estimated pooled sensitivities and false positive rates (we omit the output for space constraints). Further details are found in the documentation of `brms` and in Bürkner (2017).

3 Discussion

We have discussed software options for DTA meta-analysis and, given the space constraint of a single chapter, could not cover everything in detail. Hints at software packages for some additional aspects are provided as part of the discussion.

3.1 SROC-models

3.1.1 Moses-Littenberg SROC-approach

The Moses-Littenberg SROC-curve (Moses, Shapiro, & Littenberg, 1993) might be convenient for exploratory purposes, though it might lead to a curve with negative slope if thresholds are similar in all studies. We do not reiterate the theory behind

these curves here as Chapter XYZ of this volume covers it (**thisvolumechapterbuerkner**). The Moses-Littenberg SROC-curve can be produced with the help of `RevMan` or `mada`. Hand calculation with any statistical package is feasible but typically inconvenient.

3.1.2 Software for current SROC-approaches

Recently, several models featuring educated guesses for the ROC-curves at the primary study level have been proposed in the literature. All models mentioned in this paragraph are complementary to the bivariate model, as they produce additional insight into the distribution and especially the heterogeneity of the underlying ROC-curves (**thisvolumechapterZapf**).

Holling, Böhning, and Böhning (2012b) propose an adjusted profile maximum likelihood estimator (APMLE) for the so-called Lehmann family of (S)ROC-curves. This estimator is available in `mada` in the `phm` function. The Lehmann family approach has inspired several other methods: Holling, Böhning, and Böhning (2012a) cluster the Lehmann family curves with semiparametric mixtures and the approach could be implemented in R using the R-package `CAMAN`[7] (Schlattmann, Höhne, & Verba, 2016), though no convenient off-the-shelf implementation is available. In a similar fashion, the covariate adjusted mixtures employing t_α-(S)ROC curves instead of Lehmann curves proposed by Doebler and Holling (2015) could be implemented, again with some programming on the side of the user. The variant of Charoensawat, Böhning, Böhning, and Holling (2014) can be used with any package for univariate meta-analysis, say with `meta` or `metafor` in R. Another line of SROC-models starts with the weighted Youden-index models of Rücker and Schumacher (2010), implemented in `mada` in the `rsSROC` function. An extension of this approach by Steinhauser, Schumacher, and Rücker (2016) is discussed in the subsequent section of models for multiple thresholds.

3.2 Multiple thresholds

If 2×2-tables for more than one cutoff value are available from some of the primary studies (say from ROC-curves in the primary studies), one has to be careful not to treat them as independent estimates. Also, the diagnostic meta-analyst might want to obtain pairs of pooled sensitivity and false-positive rate for common cut-off values. In this situation, the diagnostic meta-analyst could consider to reduce the coded data in several ways: A reduction of the data could be to select a single 2×2-table per study, which clearly entails a loss of information, or to form subsets of the data for each threshold. Subsetting the data in this fashion is only advisable if enough studies end up in each subset, so it might not be possible in some DTA meta-analyses, also

[7] `CAMAN` is also the backbone for the implementation of the semiparamtric mixture approach of Schlattmann, Verba, Dewey, and Walther (2015), which extends the bivariate model.

see the discussion by Macaskill et al. (2010) and empirical work on the introduced bias by Levis et al. (2017). This problem has lead to special models for this situation.

We mention some of the existing models for multiple thresholds and what kind of code they supply for fitting the models. Dukic and Gatsonis (2003), generalizing the HSROC-model of Rutter and Gatsonis (2001), propose a Bayesian approach for which code in the BUGS language is available[8]. Implementing the approach in full will require some additional programming.

Hamza, Arends, van Houwelingen, and Stijnen (2009) extend the bivariate model of Reitsma et al. (2005) in a hierarchical fashion and obtain a multivariate random effects model. Code for SAS NLMIXED is supplied in the paper, but the approach is known to be prone to convergence problems and assumes that 2×2-tables for the same set of cutoff values can be coded for each study. Putter, Fiocco, and Stijnen (2010) instead argue in favor of an approach based on survival methods, for which R code is available as supporting information. The survival approach was not convincing enough in a simulation study of Simoneau et al. (to appear) compared to an approach with the bivariate model.

Riley et al. (2015) propose a model that handles missing cutoff values. Code in Stata is available as an additional file on the journal's website. Steinhauser et al. (2016) build on ideas of Rücker and Schumacher (2010) to present a model that handles multiple thresholds per study to estimate pooled sensitivity and false positive rate as well as an SROC-curve. R Code for this approach is part of the supplementary files for this paper. Hoyer, Hirt, and Kuss (to appear) propose an approach for meta-analysis of full ROC curves based on information from all thresholds by using bivariate time-to-event-models for interval-censored data with random effects. They supply SAS code for their approach. For some additional current approaches (Riley et al., 2014; Martínez-Camblor, 2017), we are not aware of readily available implementations.

3.3 The right tool for the job

The diagnostic meta-analyst is advised to select the appropriate software package at the planning stage of the meta-analysis, when it is decided which analyses are to be carried out. Preferably, the meta-analysis follows a protocol (**thisvolumechapterXYZ**; Deeks, Wisniewski, & Davenport, 2013), similar to that of a randomized clinical trial, and so software should be part of this protocol. From our experience, an early decision will help to organize the coding process and data preparation, as it is clear which form the data file must have to be amenable for the statistical analyses. Regardless of the chosen analysis methods, analysts can choose from a range of algorithms, several Bayesian and many frequentist, and software packages (open source and proprietary). Depending on programming skills, prior experience with mixed regression models and time budget, the meta-analyst is advised to compromise be-

[8] At the time of writing, code is found on V. Dukic's homepage: http://amath.colorado.edu/faculty/vdukic/software/ROC.html

tween flexibility and extensibility of the package on the one hand, and ease of use on the other hand.

Key message

- A number of statistical packages allow to fit the bivariate model.
- Specialized packages for DTA meta-analysis include convenient options for plotting SROC-curves.
- Currently, familiarity with more general packages is needed for special models, including those for multiple thresholds.

References

Arends, L., Hamza, T., Van Houwelingen, J., Heijenbrok-Kal, M., Hunink, M., & Stijnen, T. (2008). Bivariate Random Effects Meta-Analysis of ROC Curves. *Medical Decision Making*, *28*, 621–638.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Botella, J., Huang, H., & Suero, M. (2013). Multinomial tree models for assessing the status of the reference in studies of the accuracy of tools for binary classification. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00694

Brown, H. & Prescott, R. (2015). *Applied mixed models in medicine* (3rd). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118778210

Bürkner, P.-C. (2017). brms: an R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi:10.18637/jss.v080.i01

Charlton, C., Rasbash, J., Browne, W., Healy, M., & Cameron, B. (2017). Mlwin [computer program] version 3.00. Centre for Multilevel Modelling, University of Bristol.

Charoensawat, S., Böhning, W., Böhning, D., & Holling, H. (2014). Meta-analysis and meta-modelling for diagnostic problems. *BMC Medical Research Methodology*, *14*(1), 56. doi:10.1186/1471-2288-14-56

Chu, H. & Cole, S. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology*, *59*(12), 1331–1332.

Chu, H., Nie, L., Cole, S., & Poole, C. (2009). Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: alternative parameterizations and model selection. *Statistics in medicine*, *28*(18), 2384–2399.

Deeks, J., Wisniewski, S., & Davenport, C. (2013). Chapter 4: guide to the contents of a Cochrane Diagnostic Test Accuracy Protocol. In J. Deeks, P. Bossuyt, & C. Gatsonis (Eds.), *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0.0.* Available from: http://srdta.cochrane.org/.

Demidenko, E. (2013). *Mixed models: theory and applications*. Hoboken, NJ: John Wiley & Sons. doi:10.1002/0471728438

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*, 177–188.

Doebler, P. (2015). *mada: meta-analysis of diagnostic accuracy*. R package version 0.5.7. Retrieved from https://CRAN.R-project.org/package=mada

Doebler, P. & Holling, H. (2015). Meta-analysis of diagnostic accuracy and ROC curves with covariate adjusted semiparametric mixtures. *Psychometrika*, *80*(4), 1084–1104. doi:10.1007/s11336-014-9430-0

Doebler, P., Holling, H., & Böhning, D. (2012). A mixed model approach to meta-analysis of diagnostic studies with binary test outcome. *Psychological Methods*, *17*(3), 418–436. doi:10.1037/a0028091

Dukic, V. & Gatsonis, C. [C.]. (2003). Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. *Biometrics*, *59*(4), 936–946.

Dwamena, B. (2007). midas: Stata module for meta-analytical integration of diagnostic accuracy studies. http://econpapers.repec.org/software/bocbocode/s456880.htm.

Gasparrini, A., Armstrong, B., & Kenward, M. (2012). Multivariate meta-analysis for non-linear and other multi-parameter associations. *Statistics in Medicine*, *31*(29), 3821–3839.

Glas, A., Lijmer, J., Prins, M., Bonsel, G., & Bossuyt, P. (2003). The Diagnostic Odds Ratio: A Single Indicator of Test Performance. *Journal of Clinical Epidemiology*, *56*, 1129–1135.

Guo, J. & Riebler, A. (2016). *meta4diag: meta-analysis for diagnostic test studies*. R package version 2.0.5. Retrieved from https://CRAN.R-project.org/package=meta4diag

Hamza, T., Arends, L., van Houwelingen, H., & Stijnen, T. (2009). Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Medical Research Methodology*, *9*(1), 73.

Harbord, R., Deeks, J., Egger, M., Whiting, P., & Sterne, J. (2007). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*, *8*, 239–251.

Harbord, R. & Whiting, P. (2010). metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata Journal*, *9*, 211–229.

Holling, H., Böhning, W., & Böhning, D. (2012a). Likelihood-based clustering of meta-analytic SROC curves. *Psychometrika*, *77*(1), 106–126.

Holling, H., Böhning, W., & Böhning, D. (2012b). Meta-analysis of diagnostic studies based upon SROC-curves: a mixed model approach using the lehmann family. *Statistical Modelling*, *12*, 347–375.

Hoyer, A., Hirt, S., & Kuss, O. (to appear). Meta-analysis of full ROC curves using bivariate time-to-event models for interval-censored data. *Research Synthesis Methods*.

Hoyer, A. & Kuss, O. (2015). Meta-analysis of diagnostic tests accounting for disease prevalence: a new model using trivariate copulas. *Statistic in Medicine*, *34*(11), 1912–1924. doi:10.1002/sim.6463

Huang, H. (2014). *Metatron: meta-analysis for classification data and correction to imperfect reference*. R package version 0.1-1. Retrieved from https://CRAN.R-project.org/package=Metatron

Karrasch, S., Linde, K., Rücker, G., Sommer, H., Karsch-Völk, M., Kleijnen, J., ... Schneider, A. (2017). Accuracy of FE_{NO} for diagnosing asthma: a systematic review. *Thorax*, *72*(2), 109–116.

Kim, K., Lee, J., Choi, S., Huh, J., & Park, S. (2015). Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part I. general guidance and tips. *Korean Journal of Radiology*, *16*(6), 1175–1187.

Lee, J., Kim, K., Choi, S., Huh, J., & Park, S. (2015). Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part II. statistical methods of meta-analysis. *Korean Journal of Radiology*, *16*(6), 1188–1196.

Leeflang, M., Deeks, J., Gatsonis, C., & Bossuyt, P. (2008). Systematic Reviews of Diagnostic Test Accuracy. *Annals of Internal Medicine*, *149*, 889–897.

Levis, B., Benedetti, A., Levis, A., Ioannidis, J., Shrier, I., Cuijpers, P., ... Thombs, B. (2017). Selective cutoff reporting in studies of diagnostic test accuracy: a comparison of conventional and individual-patient-data meta-analyses of the Patient Health Questionnaire-9 depression screening tool. *American Journal of Epidemiology*, *185*(10), 954–964. doi:10.1093/aje/kww191

Liu, Z., Yao, Z., Li, C., Liu, X., Chen, H., & Gao, C. (2013). A step-by-step guide to the systematic review and meta-analysis of diagnostic and prognostic test accuracy evaluations. *British Journal of Cancer*, *108*(11), 2299–2303.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: evolution, critique and future directions. *Statistics in Medicine*, *28*(25), 3049–3067.

Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.

Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., & Takwoingi, Y. (2010). Chapter 10: analysing and presenting results. In J. Deeks, P. Bossuyt, & C. Gatsonis (Eds.), *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0*. Available from: http://srdta.cochrane.org/. The Cochrane Collaboration.

Martínez-Camblor, P. (2017). Fully non-parametric receiver operating characteristic curve estimation for random-effects meta-analysis. *Statistical Methods in Medical Research*, *26*(1), 5–20. doi:10.1177/0962280214537047

Menke, J. (2010). Bivariate random-effects meta-analysis of sensitivity and specificity with sas proc glimmix. *Methods of Information in Nedicine*, *49*(1), 54–64.

Meyer, P., Frings, L., Rücker, G., & Hellwig, S. (2017). 18F-FDG PET in Parkinsonism: differential diagnosis and cognitive impairment in Parkinson's disease. *Journal of Nuclear Medicine*. doi:10.2967/jnumed.116.186403

Moses, L., Shapiro, D., & Littenberg, B. (1993). Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Statistics in medicine*, *12*(14), 1293–1316.

Nikoloulopoulos, A. (2016). *CopulaREMADA: copula mixed effect models for bivariate and trivariate meta-analysis of diagnostic test accuracy studies*. R package version 1.0. Retrieved from https://CRAN.R-project.org/package=CopulaREMADA

Nikoloulopoulos, A. (to appear). A vine copula mixed effect model for trivariate meta-analysis of diagnostic test accuracy studies accounting for disease prevalence. *Statistical Methods in Medical Research*. doi:10 . 1177 / 0962280215596769

Partlett, C. & Takwoingi, Y. (2016). Meta-analysis of test accuracy studies in R: a summary of user-written programs and step-by-step guide to using glmer. version 1.0. Available from: http://methods.cochrane.org/sdt/.

Patrick, D., Cheadle, A., Thompson, D., Diehr, P., Koepsell, T., & Kinne, S. (1994). The validity of self-reported smoking: a review and meta-analysis. *American Journal of Public Health*, *84*, 1086–1093.

Phillips, B., Stewart, L., & Sutton, A. (2010). 'cross hairs' plots for diagnostic meta-analysis. *Research Synthesis Methods*, *1*, 308–315.

Plummer, M. (2016). *rjags: Bayesian graphical models using MCMC. R package version 4-6*. Retrieved from https://CRAN.R-project.org/package=rjags

Putter, H., Fiocco, M., & Stijnen, T. (2010). Meta-analysis of diagnostic test accuracy studies with multiple thresholds using survival methods. *Biometrical Journal*, *52*(1), 95–110.

R Core Team. (2017). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*(2), 167–190.

Reitsma, J., Glas, A., Rutjes, A., Scholten, R., Bossuyt, P., & Zwinderman, A. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, *58*, 982–990.

Richter, F., Schlattmann, P., & Dewey, M. (to appear). The evaluation of bivariate mixed models in meta-analyses of diagnostic accuracy studies with SAS, Stata and R. *Methods of Information in Medicine*.

Riley, R., Ahmed, I., Ensor, J., Takwoingi, Y., Kirkham, A., Morris, R., ... Deeks, J. (2015). Meta-analysis of test accuracy studies: an exploratory method for investigating the impact of missing thresholds. *Systematic Reviews*, *4*(1), 12.

Riley, R., Takwoingi, Y., Trikalinos, T., Guha, A., Biswas, A., Ensor, J., ... Deeks, J. (2014). Meta-analysis of test accuracy studies with multiple and missing thresholds: a multivariate-normal model. *Journal of Biometrics & Biostatistics*, *5*(3), 196. doi:10.4172/2155-6180.1000196

Rücker, G. & Schumacher, M. (2010). Summary ROC curve based on a weighted Youden index for selecting an optimal cutpoint in meta-analysis of diagnostic accuracy. *Statistics in Medicine*, *29*, 3069–3078.

Rutter, C. & Gatsonis, C. [C.A.]. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*, *20*, 2865–2884.

Schiller, I. & Dendukuri, N. (2015). *HSROC: joint meta-analysis of diagnostic test sensitivity and specificity with or without a gold standard reference test*. R package version 2.1.8.

Schlattmann, P., Höhne, J., & Verba, M. (2016). *CAMAN: finite mixture models and meta-analysis tools - based on C.A.MAN*. R package version 0.74. Retrieved from https://CRAN.R-project.org/package=CAMAN

Schlattmann, P., Verba, M., Dewey, M., & Walther, M. (2015). Mixture models in diagnostic meta-analyses – clustering summary receiver operating characteristic curves accounted for heterogeneity and correlation. *Journal of Clinical Epidemiology*, *68*(1), 61–72.

Schwarzer, G. (2007). meta: an R package for meta-analysis. *R News*, *7*(3), 40–45.

Schwarzer, G., Carpenter, J., & Rücker, G. (2015). *Meta-analysis with r*. UseR! New York: Springer.

Simoneau, G., Levis, B., Cuijpers, P., Ioannidis, J., Patten, S., Shrier, I., … Benedetti, A. (to appear). A comparison of bivariate, multivariate random-effects, and Poisson correlated gamma-frailty models to meta-analyze individual patient data of ordinal scale diagnostic tests. *Biometrical Journal*. doi:10.1002/bimj.201600184

Steinhauser, S., Schumacher, M., & Rücker, G. (2016). Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Medical Research Methodology*, *16*(1), 97. doi:10.1186/s12874-016-0196-1

Takwoingi, Y. (2016). Meta-analysis of test accuracy studies in Stata: a bivariate model approach. version 1.1. Available from: http://methods.cochrane.org/sdt/.

Takwoingi, Y. & Deeks, J. (2011). METADAS: an SAS macro for meta-analysis of diagnostic accuracy studies, version 1.3. Computer program.

The Nordic Cochrane Centre. (2014). Review Manager (RevMan) [Computer Program], Version 5.3. The Cochrane Collaboration, Copenhagen.

Verde, P. E. (2017). *bamdit: bayesian meta-analysis of diagnostic test data*. R package version 3.1.0. Retrieved from https://CRAN.R-project.org/package=bamdit

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. Retrieved from http://www.jstatsoft.org/v36/i03/

Wickham, H. & Bryan, J. (2017). *readxl: read excel files*. R package version 1.0.0. Retrieved from https://CRAN.R-project.org/package=readxl