Giuseppe Biondi-Zoccai (ed.), *Diagnostic Meta-Analysis*
https://doi.org/10.1007/978-3-319-78966-8_12

12. Statistical Packages for Diagnostic Meta-Analysis and Their Application

Philipp Doebler[1] ✉, Paul-Christian Bürkner[2] ✉ and Gerta Rücker[3] ✉

(1) Department of Statistics, TU Dortmund University, Dortmund, Germany
(2) Institute of Psychology, Faculty of Psychology and Sport Sciences, University of Münster, Münster, Germany
(3) Faculty of Medicine, Institute for Medical Biometry and Statistics, Medical Center – University of Freiburg, Freiburg im Breisgau, Germany

✉ **Philipp Doebler (Corresponding author)**
Email: doebler@statistik.tu-dortmund.de

✉ **Paul-Christian Bürkner**
Email: paul.buerkner@uni-muenster.de

✉ **Gerta Rücker**
Email: ruecker@imbi.uni-freiburg.de

Keywords Diagnostic test accuracy – Meta-analysis – Software – Bivariate model – Meta-regression – SROC curve – Descriptive statistics

12.1 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 [1] or R-packages like `meta` [2]. However, the accuracy of a diagnostic test is typically evaluated in a positive and a negative arm relative to a gold standard (Chap. 3) and thus produces two end points, the sensitivity and false-positive rate (Chap. 11) calculated from a 2 × 2 table. While such a table can be boiled down to a univariate effect size and meta-analyzed (e.g., [3]), it is recommended to employ the Reitsma et al. [4] bivariate model that simultaneously analyzes the reported pairs of sensitivity and false-positive rate [5, 6]. Alternatively, the HSROC model of Rutter and Gatsonis [7] can be used, which is equivalent to the bivariate model in the absence of covariates [8]. 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 [9] 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 is powerful graphical techniques like summary receiver operating characteristic (SROC) curves (Chap. 11), which add substantially to a diagnostic meta-analysis in the presence of different (implicit or explicit) cutoff values. Note that packages for bivariate and multivariate meta-analyses such as the R-packages `metafor` [10] or `mvmeta` [11] 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 [12] 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 et al. [13] discusses in detail how to perform an analysis in R, and several recent tutorials and/or reviews exist in the medical literature of Liu et al. [14]; Kim et al. [15]; and Lee et al. [16]. The contribution by Macaskill et al. [6] 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 Sect. 12.2 and the fitting of the bivariate model in Sect. 12.3, including the calculation and plotting for SROC curves. A brief discussion wraps up the chapter and hints at computer code for advanced methods.

12.1.1 Overview of Software Packages

Table 12.1 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 12.1, 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 [17–19].

Table 12.1 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	[20, 21]
	`jags`, `rjags`	General statistical package, Bayesian, extensible, programming needed	[22]
R	`brms`	General mixed model package, Bayesian, extensible, implementation needed	[23]
	`lme4`	General mixed model package, extensible, implementation needed	[24]
	`mada`	Specialized package for DTA meta-analysis, LMM approximation to bivariate model, graphical methods	[25]
	`meta4diag`	Specialized package for DTA meta-analysis, Bayesian, graphical methods	[26]
	`metafor`	General univariate and multivariate meta-analysis package, implementation needed	[10]
	`Metatron`	Specialized package for DTA meta-analysis, multinomial processing tree models for imperfect gold standards	[27, 28]
	`mvmeta`	General multivariate meta-analysis package, implementation needed	[11]
Proprietary software			

Statistical			
Framework	**Package/macro**	**Features and notes**	**Reference(s)**
MLwiN	–	General mixed model package	[29]
SAS	`Proc NLMIXED`	General mixed model functions, implementation needed	[30, 31]
	`Proc GLIMMIX`	General mixed model functions, implementation needed	[32]
	`METADAS`	Specialized package for DTA meta-analysis	[33]
Stata	`glamm`	General mixed model package, implementation needed	[34]
	`metandi`	Specialized package for DTA meta-analysis, graphical methods	[35]
	`meqrlogit` (`xtmelogit`)	General binary mixed model package, implementation needed	[36]
	`midas`	Specialized package for DTA meta-analysis, graphical methods	[37]

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.

12.2 Sample Workflow in R

Since all three authors of this chapter are biased toward 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` [25], 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. [38] on the diagnostic accuracy of interviewer or

self-administered questionnaires to detect smoking relative to biochemical gold standards.

12.2.1 Importing Data Into R

After coding data (Chap. 8), the analyst obtains a raw data file. Importing data into R is often made easier by employing graphical user interfaces (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 Excelfiles; [39]). 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. [38] smoking data are shown in Table 12.2. 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 dataset 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 (Chap. 8) 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 12.2 Selected rows of the Patrick et al. [38] smoking data

Row	Author	Study_id	Type	TP	FN	FP	TN	Population
1	Bauman and Dent (1982)	1	SAQ	21	15	28	324	S
2	Bauman and Dent (1982)	1	SAQ	90	10	120	969	S
3	Bauman and Dent (1982)	1	SAQ	104	8	26	232	G
4	Bauman and Dent (1982)	1	SAQ	332	18	92	673	G
5	Bauman et al. (1982)	2	SAQ	3	0	2	77	S
6	Bauman and 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

Row	Author	Study_id	Type	TP	FN	FP	TN	Population
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

12.2.2 Calculating Summary Statistics for Each Study

Chapter 11 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.

```
library (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: The null hypothesis is that all (true but unobservable) sensitivities are identical and similar for the specificities. These tests typically confirm 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 (Chap. 11) of this volume for details on these statistics. Sometimes it is convenient to use output of R-functions in subsequent calculations:

```
# 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).

12.2.3 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 [40] 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, 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 <- 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$population == "G", 1, 2),
   cex = 1.5)
   legend ("bottomright", c ("IAQ", "SAQ"), pch = 1:2, cex = 1.5)
   legend ("topright", c ("general population", "student
population"),
   pch = 15, col = 1:2, cex = 1.5)
   # make ROC-ellipse plot
   ROCellipse (smoking1, pch = ifelse (smoking1$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 12.1 contains the corresponding output. Note that the variable type and population have been used to set color and shape of symbols, so that trends resulting from these covariates might be recognized.

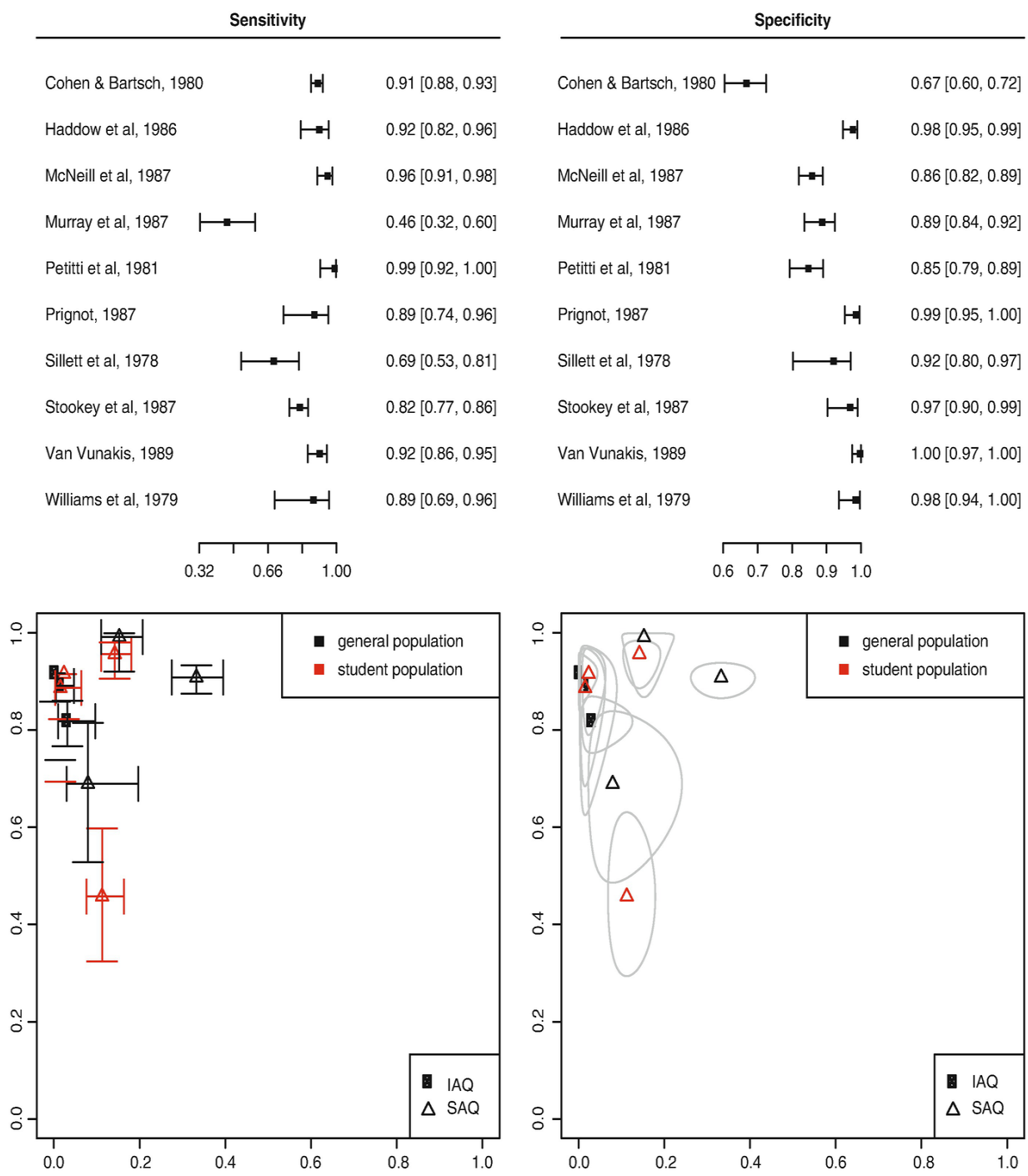

Fig. 12.1 Paired forest plots (top), "crosshairs" plot (bottom left), and confidence ellipses plot (bottom right) for a subset of the smoking data

Here, one learns about outliers (easily observed in the sensitivity forest plot or in the crosshair and ellipse plots) and heterogeneity (nonintersecting 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.

12.2.4 Fitting the Bivariate Model

The bivariate model of Reitsma et al. [4] has been introduced in (Chaps.10 and 11) 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 Chaps. 10 and 11. All packages mentioned in Table 12.1 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 entail Markov-Chain-Monte-Carlo (MCMC) techniques. 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.

12.2.5 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. [4] with restricted maximum likelihood estimation (REML). Chu and Cole [30] caution that this approximation is slightly biased. It can be improved upon by fitting a generalized linear mixed model, a point made more precise in the recent simulation study of Vogelsang et al. [41]. 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 12.2. 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, result_id == 1 & type == "SAQ")
library (mada) # LMM-approximation to the bivariate model
# 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 significantly different from 0.50. 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 (0.50 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 et al. [42]. The remainder of the output contains an estimate of the area under the SROC curve (AUC) with a value of 0.949, which is close to optimal (though the partial AUC of 0.869 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 backtransformed values are given (0.84 and 0.08 here). The between-study standard deviations of the random effects follow (1.80 and 0.90 here) together with their correlation (0.50 here), from which the covariance can be computed if necessary.

12.2.6 SROC Curves

In the majority of areas of application of DTA meta-analysis, explicit or implicit cutoff values are used to dichotomize the result of the screening test. Authors of primary studies choose cutoff 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 versus sensitivity is known as the receiver operating characteristic (ROC) curve.

Variation in the cutoff 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 [7] 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 12.2 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%.

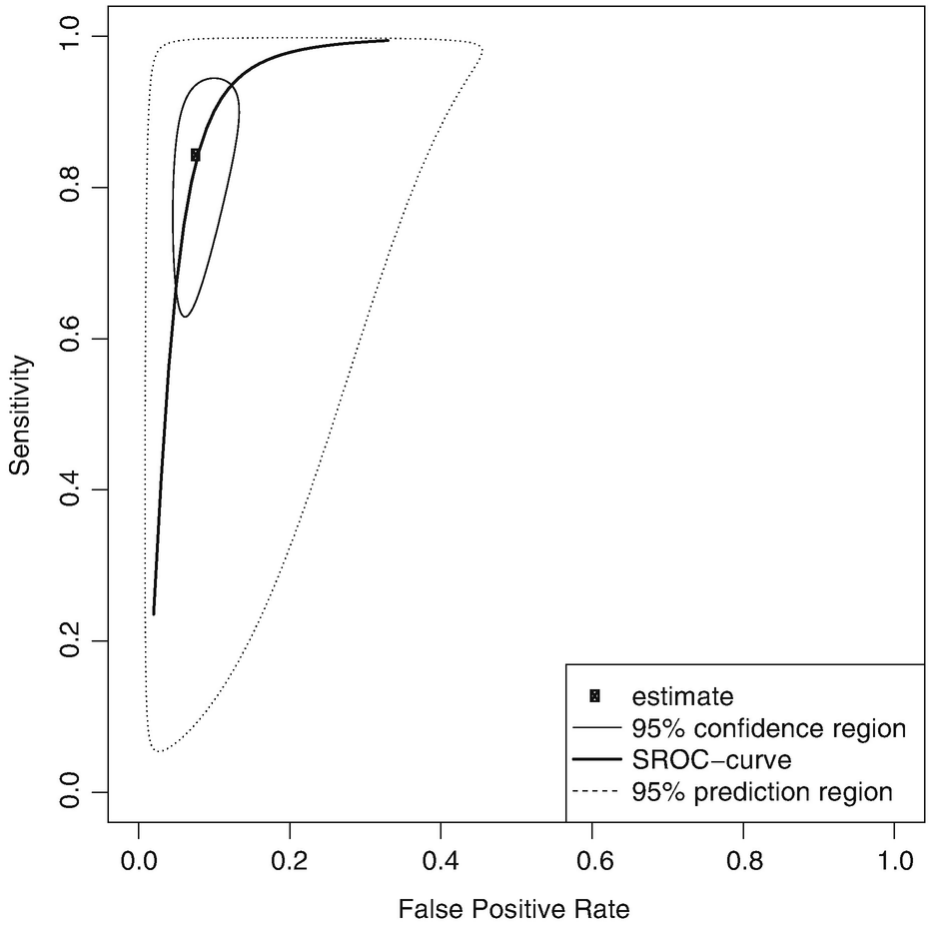

Fig. 12.2 Graphical representation of a bivariate model fit

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

12.2.7 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 11 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 modeling 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.

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 12.1 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. [43].

```
## 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 on fit measures then follows, similar to the case without covariates. Note that no 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
```

12.2.8 Fitting Strategies for Advanced Models

The bivariate model is a special case of a generalized linear mixed model (GLMM; e.g., [44, 45]), 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 (Chap. 11) or trivariate extensions including the observed prevalence [46–48]. 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 multipurpose R-package `brms` is used in our sample implementation, as it is a (relatively) convenient open-source alternative to commercial packages for GLMMs [23] and also because its Bayesian approach to parameter estimation allows to include prior information, which is in contrast to frequentist packages like `lme4` [49]. 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),
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 [23].

12.3 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.

12.3.1 SROC Models

12.3.1.1 Moses-Littenberg SROC Approach

The Moses-Littenberg SROC curve [50] 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 (Chap. 10) covers them. 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.

12.3.1.2 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 (Chap. 11).

Holling et al. [51] 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 et al. [52] cluster the Lehmann family curves with semiparametric mixtures, and the approach could be implemented in R using the R-package `CAMAN`[7] [53], 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 [54] could be implemented, again with some programming on the side of the user. The variant of Charoensawat et al. [55] 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 [56], implemented in `mada` in the `rsSROC` function. An extension of this approach by Steinhauser et al. [57] is discussed in the subsequent section of models for multiple thresholds.

12.3.2 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 cutoff 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 see the discussion by Macaskill et al. [6] and empirical work on the introduced bias by Levis et al. [58]. This problem has led 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 [59], generalizing the HSROC model of Rutter and Gatsonis [7], 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 et al. [60] extend the bivariate model of Reitsma et al. [4] 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 et al. [61] 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. [62] compared to an approach with the bivariate model.

Riley et al. [63] 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. [57] build on ideas of Rücker and Schumacher [56] 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 et al. [64] 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 [65, 66], we are not aware of readily available implementations.

12.3.3 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 ((Chap. 11); [67]), 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 and prior experience with mixed regression models and time budget, the meta-analyst is advised to compromise between flexibility and extensibility of the package on the one hand and ease of use on the other hand.

Key Messages

- 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

1. The Nordic Cochrane Centre. Review manager (RevMan) [Computer Program], version 5.3. The Cochrane Collaboration, Copenhagen; 2014.

2. Schwarzer G. meta: an R package for meta-analysis. R News. 2007;7:40–5.

3. Glas A, Lijmer J, Prins M, Bonsel G, Bossuyt P. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. 2003;56:1129–35.
[Crossref]

4. Reitsma J, Glas A, Rutjes A, Scholten R, Bossuyt P, Zwinderman A. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58:982–90.
[Crossref]

5. Leeflang M, Deeks J, Gatsonis C, Bossuyt P. Systematic reviews of diagnostic test accuracy. Ann Intern Med. 2008;149:889–97.