Meta-Analysis of Diagnostic Accuracy with `mada`

Philipp Doebler
philipp.doebler@googlemail.com

May 25, 2012

1 Introduction

While substantial work has been conducted on methods for diagnostic meta-analysis, it has not become a routine procedure yet. One of the reasons for this is certainly the complexity of bivariate approaches, but another reason is that standard software packages for meta-analysis, for example *Comprehensive Meta-Analysis* and *RevMan* ([Bio06],[Cen11]), do not include software to fit models appropriate for diagnostic meta-analysis. For the recommended ([LDGB08]) bivariate approach of Rutter and Gatsonis ([RG01]) meta-analysts can use Bayesian approaches (for example in WinBUGS ([LTBS00]) or OpenBUGS ([LSTB09])), the stata module `metandi` ([HW10]), or the SAS macro `METADAS` ([TD11]). So currently available software is either relatively complex (WinBUGS/OpenBUGS) or proprietary (stata, SAS).

The open source R-package `mada` provides some established and some current approaches to diagnostic meta-analysis, as well as functions to produce descriptive statistics and graphics. It is hopefully complete enough to be the only tool needed for a diagnostic meta-analysis. `mada` has been developed with an R user in mind that has used standard model fitting functions already, and a lot of the output of `mada` will look familiar to such a user. While this vignette cannot provide an introduction to R, it is hopefully detailed enough to provide a novice R user with enough hints to perform diagnostic meta-analysis along the lines of it. Free introductions to R are available on the homepage of the R project. We assume that the reader is familiar with central concepts of meta-analysis, like fixed and random effects models (for example [BHHR09]) and ideas behind diagnostic accuracy meta-analysis and (S)ROC curves (a starting points could be [SAJ+00], [Wal02], [JA05] or [LDGB08]).

2 Obtaining `mada`

Once R is installed and an internet connection is available, the package can be installed from CRAN on most systems by typing

```
> install.packages("mada")
```

Development of `mada` is hosted at `http://r-forge.r-project.org/projects/mada/`; the most current version is available there[1], while only stable versions are available from CRAN. The package can then be loaded:

```
> library(mada)
```

3 Entering data

Primary diagnostic studies observe the result of a *gold standard* procedure which defines the presence or absence of a *condition*, and the result of a *diagnostic test* (typically some kind of low

[1] For example by typing `install.packages("mada", repos="http://R-Forge.R-project.org")` at an R prompt.

cost procedure, or at least one that is less invasive than the gold standard). Data from such a primary study could be reported in a 2×2 table, see Table 1.

Table 1: Data from the ith study in a 2×2 table

	with condition	without condition
Test positive	y_i	z_i
Test negative	$m_i - y_i$	$n_i - z_i$
Total	m_i	n_i

The numbers y_i and z_i are the numbers of true-positives (TP) and false positives (FP), respectively, and $m_i - y_i$ and $n_i - z_i$ are the numbers of false negatives (FN) and true negatives (TN). Often derived measures of diagnostic accuracy are calculated from 2×2 tables. Using the notation in Table 1, one can calculate

$$p_i = \text{sensitivity of } i\text{th study} = \frac{y_i}{m_i} \tag{1}$$

$$u_i = \text{false positive rate of } i\text{th study} = \frac{z_i}{n_i} \tag{2}$$

$$1 - u_i = \text{specificity of } i\text{th study} = \frac{n_i - z_i}{n_i}. \tag{3}$$

Basically all functions in the `mada` package need data from 2×2 tables. One can use R to calculate the table given specificities or sensitivities if the sample size in each group is known (sometimes there is insufficient data to reconstruct the 2×2 table). The above formulae for the sensitivity for example implies that

$$y_i = m_i p_i.$$

If a primary study reports a sensitivity of .944 and that there were 142 people with the condition, we can calculate y by

```
> y <- 142 * .944
> y
```

```
[1] 134.048
```

Since this is not an integer, we need to round it to the nearest integer

```
> round(y)
```

```
[1] 134
```

Let us now assume that the number of TP, FP, FN and TN is known for each primary study. A good way to organise information in R is to use *data frames*, which can hold different variables. In our case each row of the data frame corresponds to one primary study. As an example we enter the data from six studies from a meta-analysis of the AUDIT-C (a short screening test for alcohol problems, [KHW+08]) into a data frame

```
> AuditC6 <- data.frame(TP = c(47, 126, 19, 36, 130, 84),
+                       FN = c(9, 51, 10, 3, 19, 2),
+                       FP = c(101, 272, 12, 78, 211, 68),
+                       TN = c(738, 1543, 192, 276, 959, 89))
> AuditC6
```

```
   TP FN  FP   TN
1  47  9 101  738
2 126 51 272 1543
3  19 10  12  192
4  36  3  78  276
5 130 19 211  959
6  84  2  68   89
```

Note that many central functions in `mada` also accept four vectors of frequencies (TP, FN, FP, TN) as input. Nevertheless, it is convenient to store not only the observed frequencies, but also the study names in the same data frame. The following command shows how to do this for our shortened example:

```
> AuditC6$names <- c("Study 1","Study 2","Study 4","Study 4","Study 5","Study 6")
```

The full data set with 14 studies is part of `mada`; let's load the data set and have a look at the last six studies:

```
> data(AuditC)
> tail(AuditC)
```

```
    TP FN  FP   TN
9   59  5  55  136
10 142 50 571 2788
11 137 24 107  358
12  57  3 103  437
13  34  1  21   56
14 152 51  88  264
```

In the following we will use the `AuditC` data set as a running example.

3.1 Zero cells

In the analysis of data in 2×2 tables zero cells often lead to problems or statistical artefacts since certain ratios do not exist. So called *continuity corrections* are added to the observed frequencies; these are small positive numbers. One suggestions in the literature is to use 0.5 as the continuity correction, which is the default value in `mada`. All relevant functions in `mada` allow user specified continuity corrections and the correction can be applied to all studies, or just to those with zero cells.

4 Descriptive statistics

Descriptive statistics for a data set include the sensitivity, specificity and false-positive rate of the primary studies and also their positive and negative likelihood ratios ($\mathrm{LR}_+, \mathrm{LR}_-$), and their diagnostic odds ratio (DOR; [GLP+03]). These are defined as

$$mathrmLR_+ = \frac{p}{u} = \frac{\text{sensitivity}}{\text{false positive rate}},$$

$$\mathrm{LR}_- = \frac{1-p}{1-u},$$

and

$$\mathrm{DOR} = \frac{\mathrm{LR}_+}{\mathrm{LR}_-} = \frac{\mathrm{TP}\cdot\mathrm{TN}}{\mathrm{FN}\cdot\mathrm{FP}}.$$

All these are easily computed using the `madad` function, together with their confidence intervals (see [Dee01] for the formulae used). `madad` also performs χ^2 tests to assess heterogeneity of sensitivities

and specificities, the null hypothesis being in both cases, that all are equal. Finally the correlation of sensitivities and false positive rates is calculated to give a hint whether the cut-off value problem is present. The following output is slightly cropped.

```
> madad(AuditC)

Descriptive summary of AuditC with 14 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,] 0.833 0.716 0.908 0.879 0.855 0.899
 [2,] 0.711 0.640 0.772 0.850 0.833 0.866
...
[14,] 0.748 0.684 0.802 0.749 0.702 0.792

Test for equality of sensitivities:
X-squared = 272.3603, df = 13, p-value = <2e-16
Test for equality of specificities:
X-squared = 2204.8, df = 13, 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,]   36.379 17.587    75.251  6.897 5.556  8.561 0.190 0.106 0.339
...
[14,]    8.850  5.949    13.165  2.982 2.448  3.632 0.337 0.264 0.430

Correlation of sensitivities and false positive rates:
   rho  2.5 % 97.5 %
 0.677  0.228  0.888
```

For the AUDIT-C data, the underlying call to `prop.test` produces a warning which should not worry us here. The `madad` function has a range of options with respect to computational details; for example one can compute 80% confidence intervals:

```
> madad(AuditC, level = 0.80)
```

Also note that all the output of `madad` is available for further computations if one assigns the output of `madad` to an object. For example the false positive rates with their confidence intervals can be extracted using the `$` construct (output cropped):

```
> AuditC.d <- madad(AuditC)
> AuditC.d$fpr

$fpr
 [1] 0.12083333 0.15005507 0.06097561 0.22112676 0.18061486 0.43354430
 [7] 0.20988806 0.52006770 0.28906250 0.17008929 0.23068670 0.19131238
[13] 0.27564103 0.25070822

$fpr.ci
            2.5%     97.5%
 [1,] 0.10050071 0.1446182
...
[14,] 0.20834216 0.2984416
```

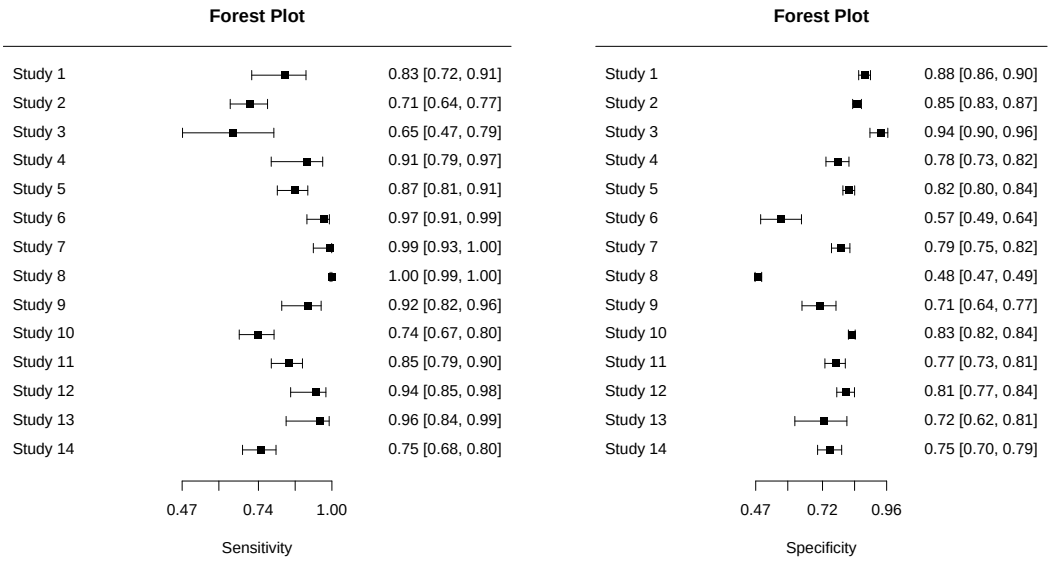


Figure 1: Paired forest plot for AUDIT-C data

4.1 Descriptive graphics

For the AUDIT-C data, the χ^2 tests already suggested heterogeneity of sensitivities and specificities. The corresponding *forest plots* confirm this:

```
> forest(madad(AuditC), type = "sens")
> forest(madad(AuditC), type = "spec")
```

These plots are shown in Figure 1.

Apart from these univariate graphics `mada` provides a variety of plots to study the data on ROC space. Note that for exploratory purposes it is often useful to employ color and other features of R's plotting system. Two high level plots are provided by `mada`: `crosshair` to produce crosshair plots ([PSS10]), and `ROCellipse`. The following is an example of a call of `crosshair` that produces (arbitrarily) colored crosshairs and makes the crosshairs wider with increased sample size; also only a portion of ROC space is plotted.

```
> ## calculate weights:
> rs <- rowSums(AuditC)
> rs <- 4 * rs/max(rs)
> crosshair(AuditC, xlim = c(0,0.6), ylim = c(0.4,1), col = 1:14, lwd = rs)
```

Figure 2 displays this plot and the next descriptive plot. `ROCellipse` plots confidence regions which describe the uncertainty of the pair of sensitivity and false positive rate. These regions are ellipses on logit ROC space, and by back-transforming them to regular ROC space the (sometimes oddly shaped) regions are produced. By default this function will also plot the point estimates. The following example is a bit contrived, but showcases the flexibility of `ROCellipse`: here the plotting of the point estimates is suppressed manipulating the `pch` argument, but then points are added in the next step.

```
> ROCellipse(AuditC, pch = "")
> points(fpr(AuditC), sens(AuditC))
```

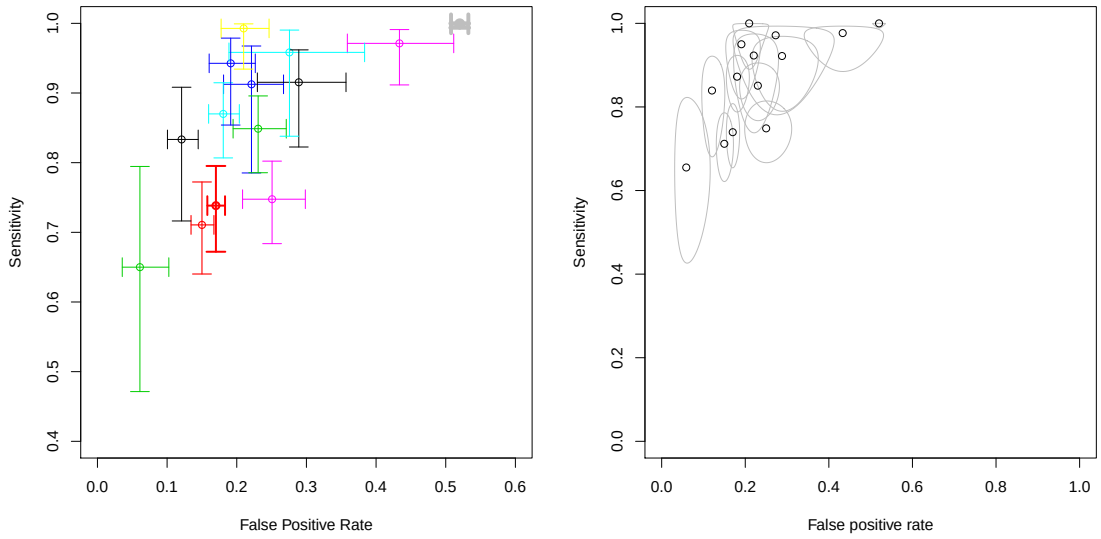


Figure 2: A "weighted" crosshair plot with (arbitrary) coloring and a plot with confidence regions for primary study estimates

5 Univariate Approaches

Before the advent of the bivariate approaches by [RG01] and [RGR+05], some univariate approaches to the meta-analysis of diagnostic accuracy were more popular. Bivariate approaches can only be recommended though, if the sample size is reasonably large; the bivariate model of [RGR+05] for example has 5 parameters, which would clearly be too much for a handful of studies. Hence `mada` provides some univariate methods. Since pooling sensitivities or specificities can be misleading ([GP06]), options for the univariate meta-analysis of these are not provided. `mada` does provide approaches for the DOR ([GLP+03]), the positive and negative likelihood ratios, and θ, the accuracy parameter of the proportional hazards model for diagnostic meta-analysis ([HBB12]). In this vignette we explain the details on the DOR methodology and the methods for θ.

5.1 Diagnostic odds ratio

In analogy to the meta-analysis of the odds ratio (OR) methods for the meta-analysis of the DOR can be developed ([GLP+03]). For the *fixed effects* case a Mantel-Haenszel (MH; see for example [Dee01]) is provided by `mada`. The underlying fixed effects model has the form

$$\mathrm{DOR}_i = \mu + \epsilon_i,$$

where μ is true underlying DOR and the ϵ_i are independent errors with mean 0 and study specific variance. The MH estimator is a weighted average of DORs observed in the primary studies and is robust to the presence of zero cells. It takes the form

$$\hat{\mu} = \sum_i \frac{\omega_i^{MH} \mathrm{DOR}_i}{\sum_i \omega_i^{MH}},$$

where $\omega_i^{MH} = \frac{z_i(m_i - y_i)}{m_i + n_i}$ are the Mantel-Haenszel weights.

One obtains an estimator for a *random effects* model following the approach of DerSimonian and Laird (DSL; [DL86]). Here the underlying model is in terms of the log DORs. One assumes

$$\log \mathrm{DOR}_i = \mu + \epsilon_i + \delta_i,$$

where μ is the mean of the log DORs, ϵ_i and δ_i are independent with mean 0; the variance σ_i^2 of ϵ_i is estimated as

$$\hat{\sigma}_i^2 = \frac{1}{y_i} + \frac{1}{m_i - y_i} + \frac{1}{z_i} + \frac{1}{n_i - z_i},$$

and the variance τ^2 of δ_i is to be estimated. The DSL estimator then is a weighted estimator, too:

$$\hat{\mu} = \sum_i \frac{\omega_i^{DSL} \mathrm{DOR}_i}{\sum_i \omega_i^{DSL}},$$

where

$$\omega_i^{DSL} = \frac{1}{\hat{\sigma}_i^2 + \tau^2}.$$

The variance τ^2 is estimated by the Cochran Q statistic trick.

The function `madauni` handles the meta-analysis of the DOR (and the negative and positive likelihood ratios). One can use `madauni` in the following fashion:

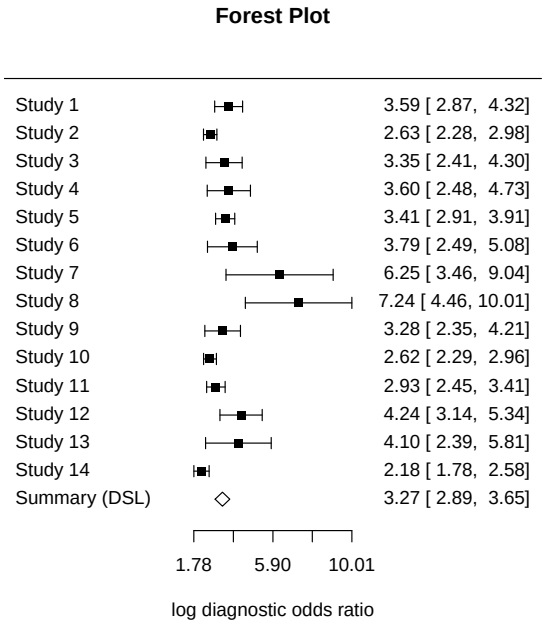
```
> (fit.DOR.DSL <- madauni(AuditC))

Call:
madauni(x = AuditC)

   DOR  tau^2
26.337  0.311

> (fit.DOR.MH <- madauni(AuditC, method = "MH"))

Call:
madauni(x = AuditC, method = "MH")

     DOR
17.93335
```

Note that the brackets around `fit.DOR.DSL <- madauni(AuditC)` are a compact way to print the fits. The `print` method for `madauni` objects is not very informative, only the point estimate is returned along with (in the random effects case) an estimate of the τ^2, the variance of the random effects. Note that estimation in the random effects case is performed on log-DOR scale, so that τ^2 of the above DSL fit is substantial. To obtain more information the `summary` method can be used:

```
> summary(fit.DOR.DSL)

Call:
madauni(x = AuditC)

Estimates:
      DSL estimate  2.5 % 97.5 %
DOR         26.337 17.971 38.596
lnDOR        3.271  2.889  3.653
tau^2        0.311  0.000  3.787
tau          0.557  0.000  1.946

Cochran's Q: 19.683 (13 df, p = 0.103)
Higgins' I^2: 33.955%
```

In addition to the confidence intervals, Cochran's Q statistic ([Coc54]) can be seen and Higgins I^2 ([HTDA03]). Producing a forest plot of the (log-)DOR values together with the summary estimate is straightforward using the `forest` method for the `madauni` class:

```
> forest(fit.DOR.DSL)
```



5.2 Proportional hazards model approach

The proportional hazards model approach (PHM; see [HBB12]) builds on the assumption of a simple form of the ROC curves. The so called *Lehmann model* ([Le06]) is assumed. Let p_i and u_i denote the ith study's sensitivity and false positive rate respectively. The relationship of p_i and u_i is then assumed to be

$$p_i = u_i^{\theta_i},$$

where $\theta_i > 0$ is a diagnostic accuracy parameter. The smaller θ, the larger the area under the ROC curve and thus the more accurate the diagnostic test. For the meta-analysis of θ the APMLE estimator is implemented in `mada` for the case of homogeneity (i.e. fixed effects) and heterogeneity (i.e. random effects). Again the standard output of the `phm` function is rather sparse:

```
> (fit.phm.homo <- phm(AuditC, hetero = FALSE))

Call:
phm.default(x = AuditC, hetero = FALSE)

Coefficients:
      theta
0.004586893

> (fit.phm.het <- phm(AuditC))

Call:
phm.default(x = AuditC)

Coefficients:
      theta     taus_sq
0.084631351 0.003706143
```

The `summary` method is more informative:

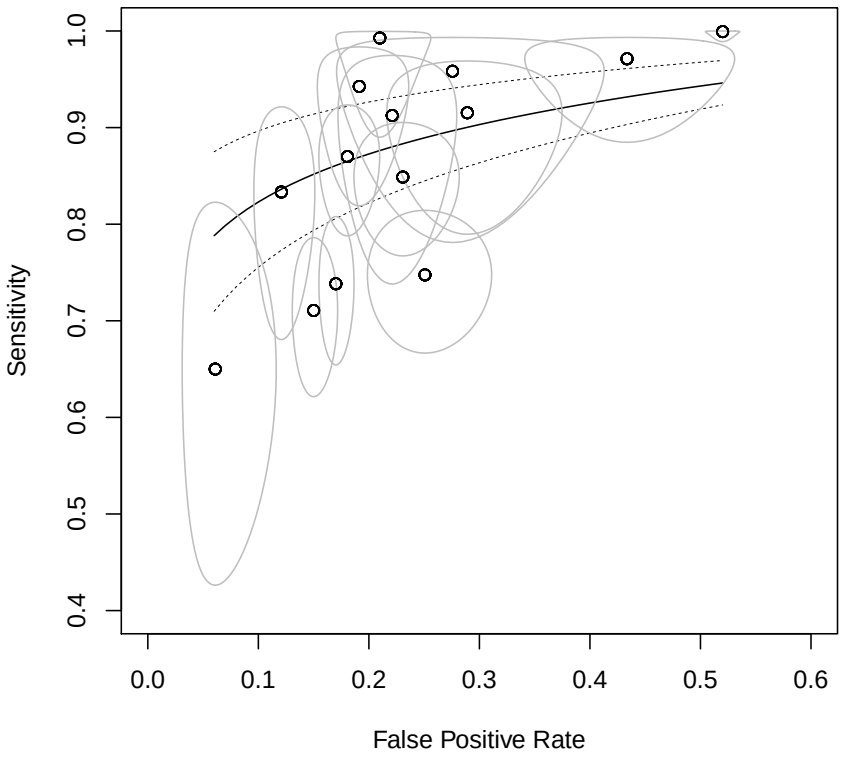
```
> summary(fit.phm.homo)

Call:
phm.default(x = AuditC, hetero = FALSE)

         Estimate       2.5 %     97.5 %
theta 0.004586893 0.003508507 0.00566528

Log-likelihood: -61.499 on 1 degrees of freedom
AIC:  125
BIC:  125.6

        Chi-square goodness of fit test (Adjusted Profile Maximum Likelihood
        under homogeneity)

data:  AuditC
Chi-square = 222.4711, df = 1, p-value < 2.2e-16
```

The χ^2 test goodness of fit test rejects the assumption of homogeneity, but the fit of the model for heterogeneity is better:

```
> summary(fit.phm.het)

Call:
phm.default(x = AuditC)

           Estimate        2.5 %      97.5 %
theta   0.084631351  0.047449859 0.121812844
taus_sq 0.003706143 -0.001277798 0.008690085

Log-likelihood: 31.121 on 2 degrees of freedom
AIC:  -58.2
BIC:  -57

        Chi-square goodness of fit test (Adjusted Profile Maximum Likelihood
        under heterogeneity)

data:  AuditC
Chi-square = 13.7264, df = 2, p-value = 0.3185
```

The estimation of θ results in an SROC curve; plotting this curve together with confidence bands obtained from the confidence interval of θ in the summary is simple (we also add the original data on ROC space with confidence regions and only plot a portion of ROC space):

```
> plot(fit.phm.het, xlim = c(0,0.6), ylim = c(0.4,1))
> ROCellipse(AuditC, add = TRUE)
```



Note that the SROC curve is not extrapolated beyond the range of the original data.

6 A bivariate approach

Typically the sensitivity and specificity of a diagnostic test depend on each other through a cut-off value: as the cut-off is varied to, say, increase the sensitivity, the specificity often decreases. So in a meta-analytic setting one will often observe (negatively) correlated sensitivities and specificities. This observation can (equivalently) also be state as a (positive) correlation of sensitivities and false positive rates. Since these two quantities are interrelated, bivariate approaches to the meta-analysis of diagnostic accuracy have been quite successful ([RG01], [VHAS02], [RGR$^+$05], [HDE$^+$07], [AHVH$^+$08]).

One typically assumes a binomial model conditional on a primary studies true sensitivity and false positive rates, and a bivariate normal model for the logit-transformed pairs of sensitivities and false positive rates. There are two ways to cast the final model: as a non-linear mixed model or as linear mixed model (see for example [AHVH$^+$08]). The latter approach is implemented in `mada`'s `reitsma` function, so we give some more details. Let p_i and u_i denote the ith study's true sensitivity and false positive rate respectively, and let $\hat{p}_i$ and $\hat{u}_i$ denote their estimates from the observed frequencies. Then, since a binomial model is assumed conditional on the true p_i, the variance of $\text{logit}(\hat{p}_i)$ can be approximated by

$$\frac{\hat{p}_i(1-\hat{p}_i)}{m_i},$$

and the variance of $\text{logit}(\hat{u}_i)$ is then

$$\frac{\hat{u}_i(1-\hat{u}_i)}{n_i}.$$

So on the within study level one assumes, conditional on p_i and u_i, that the observed variation is described by these variances and a normal model; let D_i denote a diagonal 2×2 matrix with the two variances on the diagonal. On the study level, one assumes that a global mean

$$\mu = (\mu_1, \mu_2)^\mathsf{T}$$

and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma \\ \sigma & \sigma_2^2 \end{pmatrix}$$

describe the heterogeneity of the pairs $(\text{logit}(p_i), \text{logit}(u_i))$. So the model for the ith study is then

$$(\text{logit}(\hat{p}_i), \text{logit}(\hat{u}_i))^\mathsf{T} \sim \text{N}(\mu, \Sigma + D_i).$$

Fitting this model in `mada` is similar to the other model fitting functions:

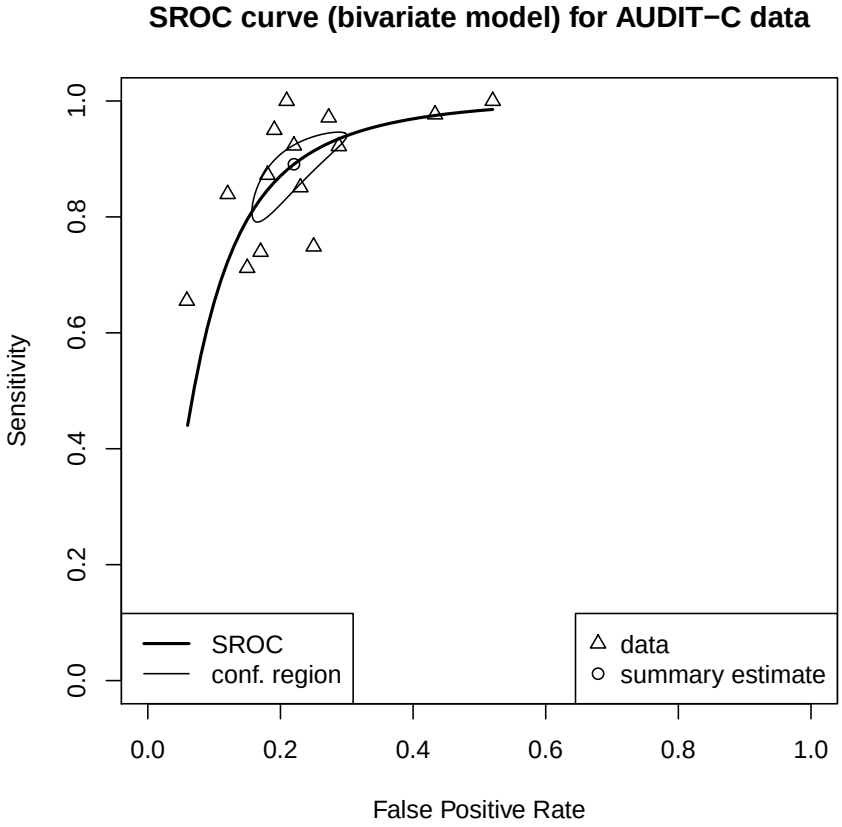
```
> (fit.reitsma <- reitsma(AuditC))

Call:
reitsma.default(X = AuditC)

Coefficients:
       mu1        mu2    Sigma11    Sigma22    Sigma21
 2.0997421 -1.2636909  1.3795810  0.4072071  0.6402931
```

The `print` method for `reitsma` objects has a scarce output. More information is offered by the `summary` method:

```
> summary(fit.reitsma)

Call:
reitsma.default(X = AuditC)

          Estimate       2.5 %     97.5 %
Sens     0.8908781  0.80199340  0.9427128
FPR      0.2203392  0.16721103  0.2845794
mu1      2.0997421  1.39880000  2.8006842
mu2     -1.2636909 -1.60552364 -0.9218582
Sigma11  1.3795810 -0.11342211  2.8725841
Sigma22  0.4072071  0.07286332  0.7415508
Sigma21  0.6402931  0.04521071  1.2353755
```

Note the sensitivity and false positive rate returned in this summary are just the back-transformed μ_1 and μ_2. One can then proceed to plot the SROC curve of this model. By default the point estimate of the pair of sensitivity and false positive rate is also plotted together with a confidence region. In the following example the SROC curve is plotted a bit thicker using the `sroclwd` argument, a caption is added to the plot and also the data and a legend. By default the SROC curve is not extrapolated beyond the range of the original data.

```
> plot(fit.reitsma, sroclwd = 2,
+      main = "SROC curve (bivariate model) for AUDIT-C data")
> points(fpr(AuditC), sens(AuditC), pch = 2)
> legend("bottomright", c("data", "summary estimate"), pch = c(2,1))
> legend("bottomleft", c("SROC", "conf. region"), lwd = c(2,1))
```



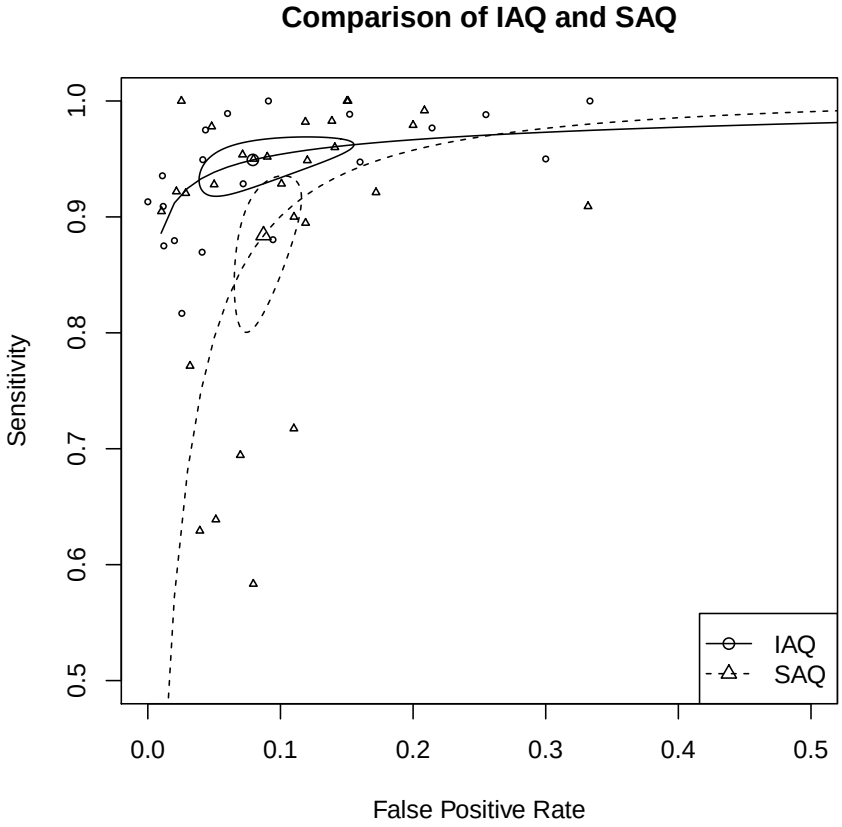
6.1 Using `mada` to compare SROC curves

We show how to compare SROC curves. [PCT+94] conducted a meta-analysis to (among other things) investigate the efficacy of self administered and interviewer administered questionnaires to detect nicotine use. The data sets `SAQ` and `IAQ` are the respective subsets of this data. First one fits bivariate models to the data sets:

```
> data(IAQ)
> data(SAQ)
> fit.IAQ <- reitsma(IAQ)
> fit.SAQ <- reitsma(SAQ)
```

Then one plots the SROC curves of these fits, beginning with the fit of the IAQ and adding the SAQ curve. Note that the `lty` arguments is used so that the curves can be distinguished.

```
> plot(fit.IAQ, xlim = c(0,.5), ylim = c(.5,1),
+      main = "Comparison of IAQ and SAQ")
> lines(sroc(fit.SAQ), lty = 2)
> ROCellipse(fit.SAQ, lty = 2, pch = 2, add = TRUE)
> ## add orginal data
> points(fpr(IAQ), sens(IAQ), cex = .5)
> points(fpr(SAQ), sens(SAQ), pch = 2, cex = 0.5)
> legend("bottomright", c("IAQ", "SAQ"), pch = 1:2, lty = 1:2)
```



The summary estimates are well separated, though the confidence regions slightly overlap. It would nevertheless be save to conclude that IAQ is a more reliable way to measure smoking than SAQ.

7 Further development of `mada`

In the future `mada` will provide functions for meta regression of diagnostic accuracy and will support the mixture approach of [HBB11] to the meta-analysis of θ as well as the bivariate approach based on the t_α transformation of [DHB12].

References

[AHVH+08] L.R. Arends, T.H. Hamza, J.C. Van Houwelingen, M.H. Heijenbrok-Kal, M.G.M. Hunink, and T. Stijnen. Bivariate Random Effects Meta-Analysis of ROC Curves. *Medical Decision Making*, 28:621–638, 2008.

[BHHR09] M. Borenstein, L.V. Hedges, J.P.T. Higgins, and H.R. Rothstein. *Introductionto Meta-Analysis*. Wiley, 2009.

[Bio06] Inc. Biostat. Comprehensive Meta-Analysis (CMA), Version 2, 2006. Computer program.

[Cen11] The Nordic Cochrane Centre. Review Manager (RevMan), Version 5.1, 2011. Computer program.

[Coc54] W.G. Cochran. The combination of estimates from different experiments. *Biometrics*, 10:101–129, 1954.

[Dee01] J.J. Deeks. Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal*, 323:157–162, 2001.

[DHB12] P. Doebler, H. Holling, and D. Böhning. A Mixed Model Approach to Meta-Analysis of Diagnostic Studies With Binary Test Outcome. *Psychological Methods*, 2012.

[DL86] R. DerSimonian and N. Laird. Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7:177–188, 1986.

[GLP+03] A.S. Glas, J.G. Lijmer, M.H. Prins, G.J. Bonsel, and P.M.M. Bossuyt. The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology*, 56:1129–1135, 2003.

[GP06] C. Gatsonis and P. Paliwal. Meta-Analysis of Diagnostic and Screening Test Accuracy Evaluations: Methodologic Primer. *American Journal of Roentgenology*, 187:271–281, 2006.

[HBB11] H. Holling, W. Böhning, and D. Böhning. Likelihood-based clustering of meta-analytic sroc curves. *Psychometrika*, pages 1–21, 2011.

[HBB12] H. Holling, W. Böhning, and D. Böhning. Meta-Analysis of Diagnostic Studies based upon SROC-Curves: a Mixed Model Approach using a Proportional Hazards Model. *Statistical Modelling - an International Journal*, 2012.

[HDE+07] R.M. Harbord, J.J. Deeks, M. Egger, P. Whiting, and J.A.C. Sterne. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*, 8:239–251, 2007.

[HTDA03] J. Higgins, S.G. Thompson, J.J. Deeks, and D.G. Altman. Measuring inconsistency in meta-analyses. *British Medical Journal*, 327:557–560, 2003.

[HW10] R.M. Harbord and P. Whiting. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata Journal*, 9:211–229, 2010.

[JA05] C.M. Jones and T. Athanasiou. Summary Receiver Operating Characteristic Curve Analysis Techniques in the Evaluation of Diagnostic Tests. *The Annals of Thoracic Surgery*, 79:16–20, 2005.

[KHW+08] L. Kriston, L. Hölzel, A.K. Weiser, M.M. Berner, and M. Härter. Meta-analysis: Are 3 Questions Enough to Detect Unhealthy Alcohol Use? *Annals of Internal Medicine*, 149:879–888, 2008.

[LDGB08] M.M.G. Leeflang, J.J. Deeks, C. Gatsonis, and P.M.M. Bossuyt. Systematic Reviews of Diagnostic Test Accuracy. *Annals of Internal Medicine*, 149:889–897, 2008.

[Le06] C.T. Le. A solution for the most basic optimization problem associated with an ROC curve. *Statistical Methods in Medical Research*, 15:571–584, 2006.

[LSTB09] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique and future directions. *Statistics in medicine*, 28(25):3049–3067, 2009.

[LTBS00] D.J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. Winbugs – a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10:325–337, 2000.

[PCT+94] D.L. Patrick, A. Cheadle, D.C. Thompson, P. Diehr, T. Koepsell, and S. Kinne. The validity of self-reported smoking: a review and meta-analysis. *American Journal of Public Health*, 84:1086–1093, 1994.

[PSS10] B. Phillips, L.A. Stewart, and A.J. Sutton. âĂŸcross hairsâĂŹ plots for diagnostic meta-analysis. *Research Synthesis Methods*, 1:308–315, 2010.

[RG01] C.M. Rutter and C.A. Gatsonis. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*, 20:2865–2884, 2001.

[RGR+05] J.B. Reitsma, A.S. Glas, A.W.S. Rutjes, R.J.P.M. Scholten, P.M. Bossuyt, and A.H. Zwinderman. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58:982–990, 2005.

[SAJ+00] A.J. Sutton, K.R. Abrams, D.R. Jones, T.A. Sheldon, and F. Song. *Methods for Meta-Analysis in Medical Research*. Wiley, 2000.

[TD11] Y. Takwoingi and JJ Deeks. METADAS: an SAS macro for meta-analysis of diagnostic accuracy studies, Version 1.3, 2011. Computer program.

[VHAS02] H.C. Van Houwelingen, L.R. Arends, and T. Stijnen. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*, 21:589–624, 2002.

[Wal02] S.D. Walter. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Statistics in Medicine*, 21:1237–1256, 2002.