metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression

Roger M. Harbord
Department of Social Medicine
University of Bristol
Bristol, UK
roger.harbord@bristol.ac.uk

Penny Whiting
Department of Social Medicine
University of Bristol
Bristol, UK

Abstract. Meta-analysis of diagnostic test accuracy presents many challenges. Even in the simplest case, when the data are summarized by a 2×2 table from each study, a statistically rigorous analysis requires hierarchical (multilevel) models that respect the binomial data structure, such as hierarchical logistic regression. We present a Stata package, `metandi`, to facilitate the fitting of such models in Stata. The commands display the results in two alternative parameterizations and produce a customizable plot. `metandi` requires either Stata 10 or above (which has the new command `xtmelogit`), or Stata 8.2 or above with `gllamm` installed.

Keywords: st0163, metandi, metandiplot, diagnosis, meta-analysis, sensitivity and specificity, hierarchical models, generalized mixed models, gllamm, xtmelogit, receiver operating characteristic (ROC), summary ROC, hierarchical summary ROC

1 Introduction

There are several existing user-written commands in Stata that are intended primarily for meta-analysis (see Sterne et al. [2007] for an overview). There is increasing interest in systematic reviews and meta-analyses of data from diagnostic accuracy studies (Deeks 2001b; Devillé et al. 2002; Tatsioni et al. 2005; Gluud and Gluud 2005; Mallett et al. 2006; Gatsonis and Paliwal 2006), which presents many additional challenges compared to more traditional meta-analysis applications, such as controlled trials. In particular, diagnostic accuracy cannot be adequately summarized by one measure; two measures are typically used, most often sensitivity and specificity or, alternatively, positive and negative likelihood ratios, and the two are correlated (Deeks 2001a). Meta-analysis of diagnostic accuracy therefore requires different and more complex methods than traditional meta-analysis applications, even in the simplest situation where the data from each primary study are summarized as a 2×2 table of test results against true disease status, both of which have been dichotomized. In addition, substantial between-study heterogeneity is commonplace, and the models must account for this (Lijmer, Bossuyt, and Heisterkamp 2002).

Several methods of meta-analyzing diagnostic accuracy data have been proposed, of which two are statistically rigorous: the hierarchical summary receiver operating characteristic (HSROC) model (Rutter and Gatsonis 2001) and the bivariate model (Reitsma et al. 2005). In the absence of covariates, these turn out to be different parameterizations of the same model (Harbord et al. 2007; Arends et al. 2008).

　　　　　　　　　　　　　　　　　　　st0163

The bivariate model can be fit in Stata by using the user-written `gllamm` command, as pointed out by Coveney (2004). In Stata 10, the same model can be fit considerably faster by using the new `xtmelogit` command. In either case, however, some data preparation is required, the syntax is complex (particularly for `gllamm`), and the output is not easy to interpret.

In this article, we present a new Stata command, `metandi`, to facilitate the fitting of these hierarchical logistic regression models for meta-analysis of diagnostic test accuracy. The `metandi` command fits the model and displays the estimates in both the HSROC and bivariate parameterizations. `metandi` also displays some familiar summary measures (sensitivity and specificity, positive and negative likelihood ratios, and the diagnostic odds ratio). However, these simple summary measures fail to describe the expected trade-off between sensitivity and specificity, which is best illustrated graphically. We have therefore included a command, `metandiplot`, to simplify the plotting of graphical summaries of the fitted model, namely, the summary receiver operating characteristic (SROC) curve and the prediction region, and also to plot the summary point and its confidence region.

The name `metandi` was chosen to indicate that, like `metan` (Bradburn, Deeks, and Altman 1998), `metandi` takes the cell counts of 2×2 tables as input but is designed for meta-analysis of diagnostic accuracy.

`metandi` is not intended to provide a comprehensive package for diagnostic meta-analysis by itself; other plots are also useful, such as forest plots showing within-study estimates and confidence intervals for sensitivity and specificity separately (Deeks 2001b).

Section 2 of this article introduces an example dataset, which we will use to illustrate the commands. Section 3 then gives some background on methods and models that have been proposed for meta-analysis of diagnostic accuracy. Sections 4 and 5 illustrate the output of `metandi` and `metandiplot` on the example dataset. Section 6, which assumes somewhat greater knowledge of both statistics and Stata, gives examples of the use of `predict` after `metandi` for model checking and identification of influential studies. Finally, sections 7 and 2, which are intended mainly as reference material, detail the formal syntax of the commands, and the methods and formulas used.

2 Example: Lymphangiography for diagnosis of lymph node metastasis

We shall illustrate the use of the `metandi` package on data from 17 studies of lymphangiography for the diagnosis of lymph node metastasis in women with cervical cancer. Lymphangiography is one of three imaging techniques in the meta-analysis of Scheidler et al. (1997), and these data have been frequently used as an example for methodological papers on meta-analysis of diagnostic accuracy (Rutter and Gatsonis 2001; Macaskill 2004; Reitsma et al. 2005; Harbord et al. 2007). These data are provided in the auxiliary file `scheidler_LAG.dta`. The total number of patients in each study ranges from 21 to 300. There is one observation in the dataset for each study.

The data needed for meta-analysis consist of the number of true positives (`tp`), false positives (`fp`), false negatives (`fn`), and true negatives (`tn`).

Figure 1 shows a SROC plot of these data, generated by the official Stata commands given below. An SROC plot is similar to a conventional ROC plot (see, e.g., [R] **roc**) in that it plots sensitivity (true-positive rate) against specificity (true-negative rate), but here each symbol represents a different study rather than a different threshold within the same study. It therefore makes no sense to connect the points with a line, but it can be useful to indicate the size of each study by the symbol size. (It might be preferable to use an ellipse or rectangle to separately indicate the number of people with [`tp + fn`] and without [`tn + fp`] the disease of interest, but this is hard to achieve within the current Stata graphics system.) By convention, the specificity is plotted on a reversed scale (or equivalently, the false-positive rate is plotted on a conventional scale).

```
. use scheidler_LAG
(Lymphangiography for diagnosing lymph node metastases)
. generate sens = tp/(tp+fn)
. generate spec = tn/(tn+fp)
. label variable sens "Sensitivity"
. label variable spec "Specificity"
. local opts "xscale(reverse) xla(0(.2)1) yla(0(.2)1, nogrid) aspect(1) nodraw"
. scatter sens spec [fw=tp+fp+fn+tn], m(Oh) `opts´ name(sroccirc)
. scatter sens spec, mlabel(studyid) m(i) mlabpos(0) `opts´ name(sroclab)
. graph combine sroccirc sroclab, xsize(4.5) scale(*1.5)
```
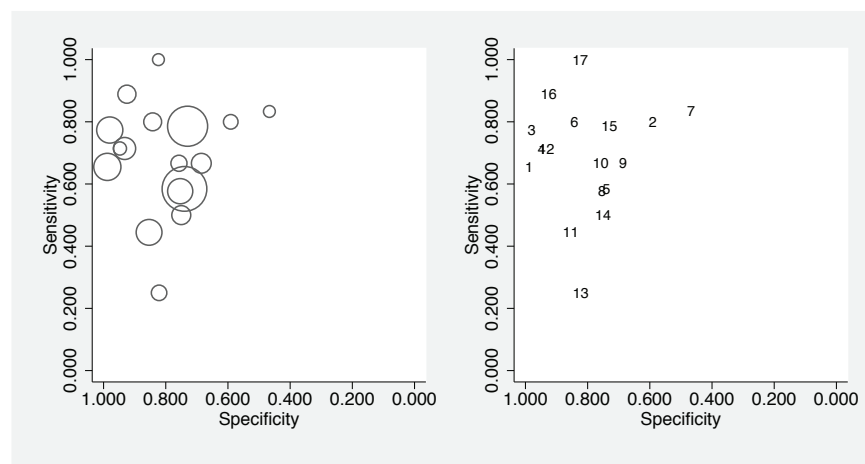

Figure 1. SROC plot of the lymphangiography data. Left panel: Studies indicated by circles sized according to the total number of individuals in each study. Right panel: Studies indicated by study ID numbers.

3   Models for meta-analysis of diagnostic accuracy

Several statistical methods for meta-analysis of data from diagnostic test accuracy studies have been proposed that account for the correlation between sensitivity and specificity (Moses, Shapiro, and Littenberg 1993; Rutter and Gatsonis 2001; Reitsma et al. 2005).

Moses, Shapiro, and Littenberg (1993) proposed a method of generating an SROC curve by using simple linear regression. This method has frequently been used, but the assumptions of simple linear regression are not met, and the method is therefore approximate. There is also uncertainty as to the most appropriate weighting of the regression (Walter 2002; Rutter and Gatsonis 2001).

Two more-complex but statistically rigorous approaches have been proposed that overcome the limitations of the linear regression method: the HSROC model (Rutter and Gatsonia 2001) and the bivariate model (Reitsma et al. 2005). Both approaches are based on hierarchical models, i.e., both approaches involve statistical distributions at two levels. At the lower level, they model the cell counts in the 2×2 tables by using binomial distributions and logistic (log-odds) transformations of proportions. Although their motivation is distinct and they allow covariates to be added to the models in different ways, it has been shown that the two models are equivalent when no covariates are fit, as well as in certain models including covariates (Harbord et al. 2007; Arends et al. 2008).

3.1   HSROC model

The HSROC model (Rutter and Gatsonis 2001) assumes that there is an underlying ROC curve in each study with parameters α and β that characterize the accuracy and asymmetry of the curve. The 2×2 table for each study then arises from dichotomizing at a positivity threshold, θ. The parameters α and θ are assumed to vary between studies; both are assumed to have normal distributions as in conventional random-effects meta-analysis. The accuracy parameter has a mean of Λ (capital lambda) and a variance of σ_α^2, while the positivity parameter θ has a mean of Θ (capital theta) and a variance of σ_θ^2. Because estimation of the shape parameter, β, requires information from more than one study, it is assumed constant across studies. When no covariates are included in an HSROC model, there are therefore five parameters: Λ, Θ, β, σ_α^2, and σ_θ^2.

3.2   Bivariate model

The bivariate model (Reitsma et al. 2005) models the sensitivity and specificity more directly. It assumes that their logit (log-odds) transforms have a bivariate normal distribution between studies. The logit-transformed sensitivities are assumed to have a mean of μ_A and a variance of σ_A^2, while the logit-transformed specificities have a mean of μ_B and a variance of σ_B^2. The trade-off between sensitivity and specificity is allowed for by also including a correlation, ρ_{AB}, that is expected to be negative. The bivariate

model, like the HSROC model, therefore has five parameters when no covariates are included: μ_A, μ_B, σ_A^2, σ_B^2, and ρ_{AB}.

4   metandi output

The output from running metandi on the lymphangiography data is shown below (the nolog option suppresses the iteration log and is used here merely to save space):

```
. use scheidler_LAG, clear
(Lymphangiography for diagnosing lymph node metastases)

. metandi tp fp fn tn, nolog

True  positives: tp                    False positives: fp
False negatives: fn                    True  negatives: tn

Meta-analysis of diagnostic accuracy

Log likelihood   = -91.391372                    Number of studies =       17
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Bivariate						
E(logitSe)	.7266321	.1544626			.4238909	1.029373
E(logitSp)	1.638955	.2505372			1.147911	2.129999
Var(logitSe)	.1249622	.1306738			.0160943	.9702552
Var(logitSp)	.8232703	.4055446			.3135009	2.161952
Corr(logits)	.2387873	.4557706			-.6067877	.8308258
HSROC						
Lambda	2.187142	.3086554			1.582189	2.792096
Theta	.0705698	.3271092			-.5705525	.7116921
beta	.9426366	.5764601	1.64	0.102	-.1872044	2.072478
s2alpha	.7946708	.5114529			.2250873	2.805586
s2theta	.1220778	.1082908			.0214569	.6945553
Summary pt.						
Se	.6740658	.0339356			.6044139	.7367944
Sp	.8373927	.0341147			.7591292	.8937849
DOR	10.65029	3.296352			5.806411	19.53509
LR+	4.145361	.9181013			2.685598	6.398582
LR-	.389225	.0452324			.3099427	.4887875
1/LR-	2.569208	.2985712			2.045879	3.226402

```
Covariance between estimates of E(logitSe) & E(logitSp)    .0045838
```

The bivariate and HSROC parameter estimates are displayed along with their standard errors and approximate 95% confidence intervals in the standard Stata format. The bivariate location parameters, μ_A and μ_B, are denoted by E(logitSe) and E(logitSp); the variance parameters, σ_A^2 and σ_B^2, are shown as Var(logitSe) and Var(logitSp); and the correlation, σ_{AB}, is shown as Corr(logits). The HSROC parameters are denoted by using the notation of Rutter and Gatsonis (2001) given in section 3.1, spelling out Greek letters with capital initials for the capital Greek letters Λ and Θ, and showing σ_α^2 and σ_θ^2 as s2alpha and s2theta.

z statistics and p-values are not given for most of the parameters because parameter values of zero do not correspond to null hypotheses of interest. The exception is the HSROC shape (asymmetry) parameter, β (`beta`), where $\beta = 0$ corresponds to a symmetric ROC curve in which the diagnostic odds ratio does not vary along the curve.

The output also gives summary values and confidence intervals for the sensitivity (`Se`) and specificity (`Sp`) (back-transformed from `E(logitSe)` and `E(logitSp)`), as well as values for the diagnostic odds ratio (`DOR`) and the positive and negative likelihood ratios (`LR+` and `LR-`) at the summary point. The summary likelihood ratios will not, in general, be the same as would be obtained by first calculating the likelihood ratios for each study and meta-analyzing these. Such an approach has been deprecated in favor of the approach implemented here (Zwinderman and Bossuyt 2008). A summary value for the inverse of the negative likelihood ratio (`1/LR-`) is also given, because larger values of the inverse of the negative likelihood ratio indicate a more accurate test, and comparing this with the positive likelihood ratio can indicate whether a positive or negative test result has greater impact on the odds of disease.

Finally, the output shows the covariance between $\widehat{\mu}_A$ and $\widehat{\mu}_B$. This is needed to draw confidence and prediction regions, and is included to make it easier to do so in external software, such as the Cochrane Collaboration's Review Manager 5 (Nordic Cochrane Centre 2007).

❏ **Technical note**

On rare occasions, during model fitting, `gllamm` may report an error, such as "convergence not achieved: try with more quadrature points" or (less transparently) "log likelihood cannot be computed". Increasing the number of integration points beyond `metandi`'s default of 5 by using the `nip()` option (e.g., `nip(7)`) may resolve this.

❏

5   metandiplot

The `metandiplot` command produces a graph of the model fit by `metandi`, which must be the last estimation-class command executed. For convenience, the `metandi` command has a `plot` option, which produces the same graph. If `metandiplot` is not followed by a varlist, then the study-specific estimates (shown by the circles in figure 2) are not included in the graph. The `metandiplot` command has options to alter the default appearance of the graph or to turn off any of the plot elements. These options are not available when using the `plot` option to `metandi`. `metandiplot` can be run many times with different options without refitting the model with `metandi`.
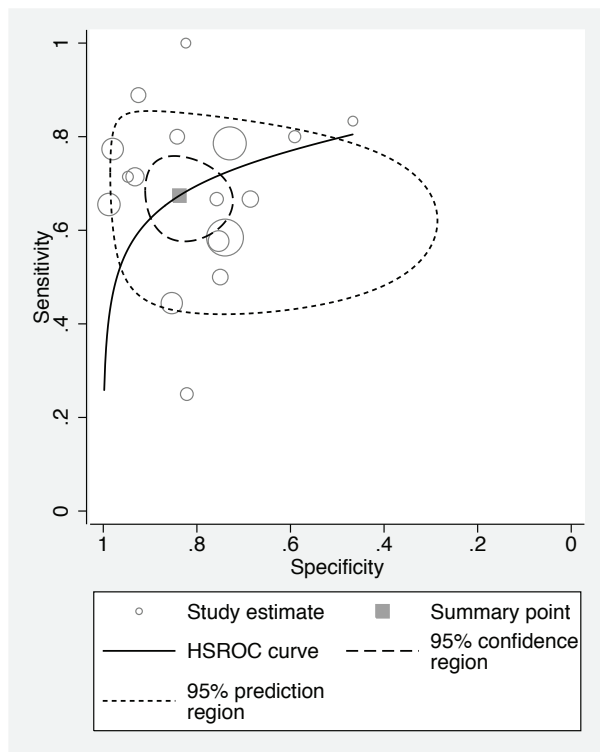
```
. metandiplot tp fp fn tn
```


Figure 2. Plot of fitted model from `metandiplot`

The resulting graph (figure 2) shows the following summaries, together with circles showing the individual study estimates:

- A summary curve from the HSROC model

- A summary operating point, i.e., summary values for sensitivity and specificity

- A 95% confidence region for the summary operating point

- A 95% prediction region (confidence region for a forecast of the true sensitivity and specificity in a future study)

The default is to include all the summaries listed above, which can result in a rather cluttered graph, so options are included to remove any of the elements; for example, `predopts(off)` turns off the prediction region. See section 7.2 for more information about `metandiplot` options.

By default, the summary HSROC curve is displayed only for sensitivities and specificities at least as large as the smallest study-specific estimates if a varlist is included.

The shape of the prediction region is dependent on the assumption of a bivariate normal distribution for the random effects and should therefore not be overinterpreted; it is intended to give a visual representation of the extent of between-study heterogeneity, which is often considerable.

6   predict after metandi

Many of Stata's standard postestimation tools will not work after `metandi` or will not work as expected, because `metandi` temporarily reshapes the data before fitting the model.

The notable exception is `predict`, which can be used to obtain posterior predictions (empirical Bayes estimates) of the sensitivity and specificity in each study (`mu`), as well as various statistics that can be useful for detecting outliers (e.g., `ustd`) and influential observations (`cooksd`).

The help file provides basic commands for examining diagnostics. We take the opportunity here to provide slightly more customized displays.

Empirical Bayes estimates give the best estimate of the true sensitivity and specificity in each study, and these estimates will be "shrunk" toward the summary point compared with the study-specific estimates shown in figure 1.

```
. predict eb
(option mu assumed; posterior predicted Se & Sp)

. metandiplot, addplot(scatter eb1 eb0, msymbol(o))
> legend(label(5 "Empirical Bayes"))
```
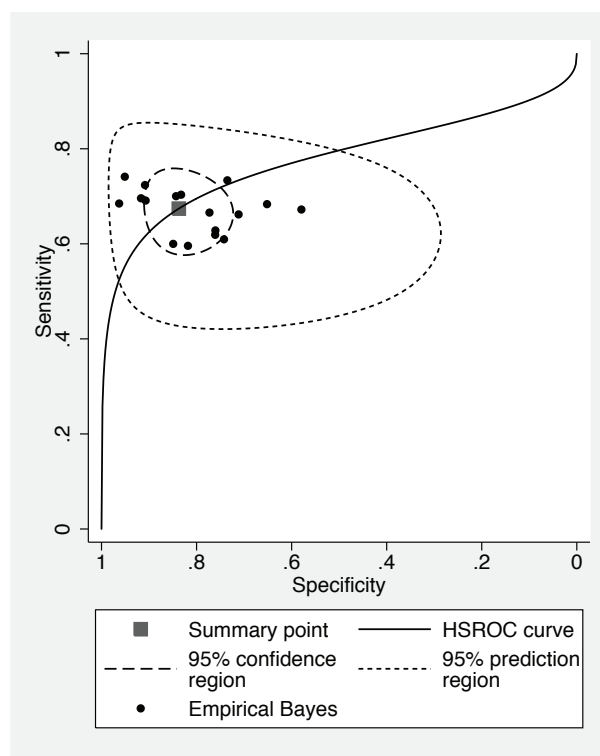

Figure 3. Empirical Bayes estimates

Comparing figure 3 with figure 2 shows that the shrinkage is generally greater for sensitivity than for specificity in this example, reflecting both the smaller variance of sensitivity (on the logit scale) and the fact that most studies have fewer participants with disease than without disease, leading to more precise estimates of specificity than of sensitivity.

Cook's distance is a measure of the influence of a study on the model parameters and can be used to check for particularly influential studies. Cook's distance is calculated using `gllapred` and so is available in Stata 10 only if the `gllamm` option was used with `metandi`. `gllapred` calculates Cook's distance to measure influence on all model parameters including the variance parameters (Skrondal and Rabe-Hesketh 2004, sec. 8.6.6). To check for outliers, standardized predicted random effects can be interpreted as standardized study-level residuals.

```
. metandi tp fp fn tn, gllamm nolog
  (output omitted)
. predict cooksd, cooksd
(Cook´s distance may take a few seconds...)
. predict ustd_Se ustd_Sp, ustd
. local opts "mlabel(studyid) mlabpos(0) m(i) nodraw"
. scatter cooksd studyid,`opts´ name(cooksd)
. scatter ustd_Se ustd_Sp, xscale(rev) xla(, grid) xline(0) yline(0) `opts´
> name(ustd)
. graph combine cooksd ustd, xsize(5) scale(*1.5)
```
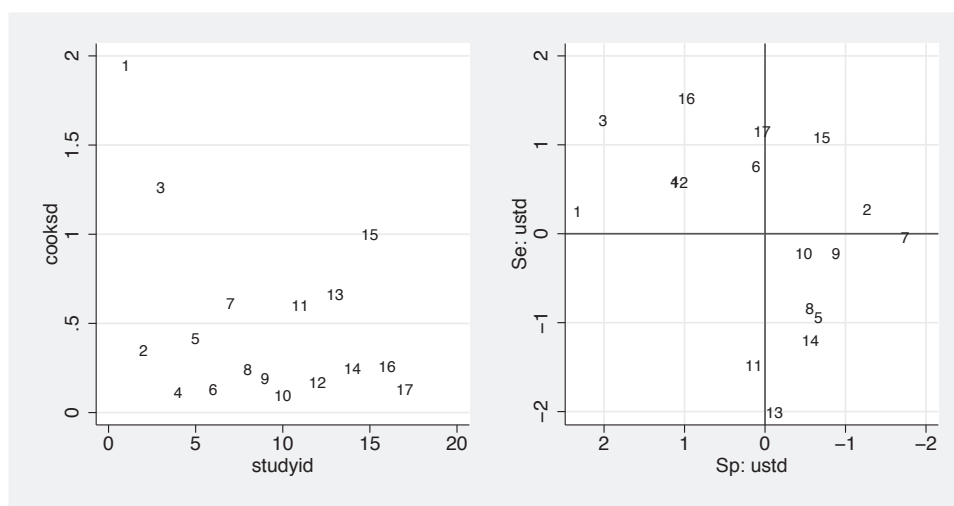

Figure 4. Left panel: Cook's distance. Right panel: Standardized residuals (standardized predicted random effects).

Figure 4 shows both Cook's distance and the standardized residuals. (The residual corresponding to specificity has been plotted on a reversed axis to correspond with the convention for ROC plots used in figure 1.) These two graphs are best read in combination. Cook's distance shows which studies are influential, while the standardized residuals give some insight into why. According to Skrondal and Rabe-Hesketh (2004), a typical cutpoint for declaring a value of Cook's D to be "large" is four times the number of parameters divided by the number of clusters (here studies). (Definitions of Cook's D differ, hence so does the cutpoint—the definition used by Stata in [R] **regress postestimation** divides by the number of parameters.) Because there are five parameters in this model, this suggests a cutpoint of 20 divided by the number of studies for interpreting Cook's D after metandi, giving $20/17 \approx 1.2$ for the lymphangiography meta-analysis. Here, study 1 is particularly influential, followed by study 3. Studies 1 and 3 have high standardized residuals for specificity, leading to influence on both the mean and variance of logit-transformed specificity. Study 13 has a large (negative) standardized residual for sensitivity but does not appear to be so influential as judged

by its Cook's distance. Further investigation of the effect of individual studies on the model could be undertaken by refitting the model and leaving out each study in turn.

7 Syntax and options for commands

7.1 The metandi command

Syntax

metandi *tp fp fn tn* $\big[\,if\,\big]$ $\big[\,in\,\big]$ $\big[$, <u>p</u>lot <u>g</u>llamm force ip(g|m) <u>ni</u>p(#)
  <u>nob</u>ivariate <u>noh</u>sroc <u>nos</u>ummarypt detail <u>l</u>evel(#) <u>tr</u>ace <u>nolog</u>$\big]$

  by is allowed with metandi; see [D] **by**.

Options

plot requests a plot of the results on an SROC plot. This is a convenience option equivalent to executing the metandiplot command after metandi with the same list of variables, *tp*, *fp*, *fn*, and *tn* (and the same *if* and *in* qualifiers, if specified). Greater control of the plot is available through the options of the metandiplot command when issued as a separate command after metandi.

gllamm specifies that the model be fit using gllamm. This is the default in Stata 8 and 9, so the option is of use only in Stata 10, in which the model is fit using xtmelogit by default.

force forces metandi to attempt to fit data where one or more studies have $tp + fn = 0$ (or $tn + fp = 0$), i.e., where there are no individuals that are positive (negative) for the reference standard. Without this option, metandi exits with an error when such data exist. Problems may be encountered in fitting such data, particularly when the model is fit using xtmelogit. Sensitivity (specificity) cannot be estimated within such studies, so they are not included in the plot produced by metandiplot.

ip(g|m) specifies the quadrature (numerical integration) method used to integrate out the random effects: ip(g), the default, gives Cartesian product quadrature, while ip(m) gives spherical quadrature, which is available in gllamm but not in xtmelogit. Spherical quadrature can be more efficient, though its properties are less well known and it can sometimes cause the adaptive quadrature step to take longer to converge. See Rabe-Hesketh, Skrondal, and Pickles (2005).

nip(#) specifies the number of integration points used for quadrature. Higher values should result in greater accuracy but typically at the expense of longer execution times. Specifying too small a value can lead to convergence problems or even failure of adaptive quadrature; if you receive the error "log likelihood cannot be computed", try increasing nip(). For Cartesian product quadrature, nip() specifies the number of points for each of the two random effects; the default is nip(5). For spher-

ical quadrature, `nip()` specifies the degree, d, of the approximation; the default is
`nip(9)`, and the only values currently supported by `gllamm` are 5, 7, 9, 11, and 15.
These defaults give approximately the same accuracy, because degree d for spherical
quadrature approximately corresponds in accuracy to $(d+1)/2$ points per random
effect for Cartesian product quadrature (Rabe-Hesketh, Skrondal, and Pickles 2005,
app. B).

`nobivariate`, `nohsroc`, and `nosummarypt` suppress reporting of the bivariate param-
eter estimates, the HSROC parameter estimates, or the summary point estimates,
respectively.

`detail` displays the output of all `gllamm` or `xtmelogit` commands issued.

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. The
default is `level(95)` or as set by `set level`; see [U] **20.7 Specifying the width
of confidence intervals**.

`trace` adds a display of the current parameter vector to the iteration log.

`nolog` suppresses display of the iteration log.

7.2   The metandiplot command

Syntax

`metandiplot` $\big[\,tp\ fp\ fn\ tn\,\big]$ $\big[\,if\,\big]$ $\big[\,in\,\big]$ $\big[\,weight\,\big]$ $\big[$, <u>notr</u>uncate <u>level</u>(*#*)
    <u>predlevel</u>(*numlist*) <u>npoints</u>(*#*) *subplot_options* `addplot`(*plot*)
    twoway_options $\big]$

Options

`notruncate` specifies that the HSROC curve will not be truncated outside the region
of the data. By default, the HSROC curve is not shown when the sensitivity or
specificity is less than its smallest study estimate.

`level(#)` specifies the confidence level, as a percentage, for the confidence contour.
The default is `level(95)` or as set by `set level`; see [U] **20.7 Specifying the
width of confidence intervals**.

`predlevel`(*numlist*) specifies the levels, as a percentage, for the prediction contour(s).
The default is one contour at the same probability level as the confidence region.
Up to five prediction contours are allowed.

`npoints(#)` specifies the number of points to use in drawing the outlines of the confi-
dence and prediction regions. The default is `npoints(500)`.

subplot_options, which are <u>summ</u>opts(), <u>conf</u>opts(), <u>pred</u>opts(), <u>curve</u>opts(), and
    <u>stud</u>yopts(), control the display of the summary point, confidence contour, predic-
tion contour(s), HSROC curve, and study symbols, respectively. The options within

each set of parentheses are simply passed through to the appropriate `twoway` plot. Any of the plots can be turned off by specifying, for example, `summopts(off)`.

`addplot(`*plot*`)` allows adding additional `graph twoway` plots to the graph; see [G] ***ad-dplot_option***. For example, empirical Bayes predictions could be generated by using `predict` after `metandi` and then added to the graph. See section 6.

twoway_options are most of the options documented in [G] ***twoway_options***, including options for titles, axes, labels, schemes, and saving the graph to disk. However, the `by()` option is not allowed.

7.3   The predict command after metandi

Syntax

`predict` [*type*] *newvarlist* [*if*] [*in*] [`,` *statistic*]

statistic	description
`mu`	posterior predicted (empirical Bayes) sensitivity and specificity; the default
`u`	posterior means (empirical Bayes predictions, BLUPs) of random effects
`sdu`	posterior standard deviations of random effects
`ustd`	standardized posterior means of random effects
<u>lin</u>pred	linear predictor with empirical Bayes predictions plugged in: $\text{linpred} = \text{xb} + \text{u}$
<u>cook</u>sd	Cook's distance for each study; available only when model was fit using `gllamm`

Most of the above statistics require *newvarlist* to consist of two new variables to store them: one for the statistic associated with sensitivity and one for the statistic associated with specificity. If *newvarlist* contains only one *newvar*, the statistics associated with sensitivity and specificity will be stored in *newvar*1 and *newvar*0, respectively. `cooksd`, however, is computed once for each study and therefore requires only one *newvar*. See section 6 for examples.

(*Continued on next page*)

7.4 Saved results

`metandi` saves the following results in `e()`:

Scalars
    `e(N)`   number of studies   `e(ll)`   log likelihood

Macros
    `e(cmd)`   `metandi`   `e(predict)`   program used to implement
    `e(tpfpfntn)`   names of *tp fp fn tn*   predict
                      variables   `e(properties)` b V
    `e(cmd)`   `metareg`

Matrices
    `e(b)`   bivariate coefficient vector   `e(V)`   variance–covariance matrix of
    `e(b_hsroc)`   HSROC coefficient vector   the bivariate estimators
                                              `e(V_hsroc)`   variance–covariance matrix of
                                                    the HSROC estimators

Functions
    `e(sample)`   marks estimation sample

8   Methods and formulas

It is possible to use routines for linear mixed models to fit an approximate version of the bivariate model obtained by using empirical logit transforms of the estimated sensitivity and specificity in each study together with their estimated standard errors (Reitsma et al. 2005). However, the small cell counts common in diagnostic accuracy studies can lead to poor performance of such approximations. Generalized mixed models, in particular, hierarchical (mixed-effects) logistic regression, can handle the binomial nature of the data directly and are therefore preferable (Chu and Cole 2006; Riley et al. 2007).

Such models are complex to fit, however, because they require numerical integration (quadrature) to integrate out the random effects. `metandi` uses `gllamm` or `xtmelogit` to fit the bivariate model by using adaptive quadrature, then transforms the parameter estimates to those of the HSROC model by using the delta method (Cox 1998).

Because the bivariate model can sometimes prove difficult to fit, some care has been taken to provide good starting values. First, two separate univariate models are fit to sensitivity and specificity. These provide excellent starting values for the two mean and two variance parameters of the bivariate model. A reasonable starting value for the correlation parameter is obtained from the correlation between the posterior means (empirical Bayes predictions) of the two univariate models.

We now give the mathematical forms of the bivariate and HSROC models in the absence of covariates. See Rutter and Gatsonis (2001); Reitsma et al. (2005); and Harbord et al. (2007) for information on the models with covariates, which are not currently supported by `metandi`.

8.1   The bivariate model

Following Reitsma et al. (2005), we denote the sensitivity in the ith study by p_{Ai} and the specificity by p_{Bi}, and base analysis on their logit transforms:

$$\mu_{Ai} = \text{logit}(p_{Ai})$$

$$\mu_{Bi} = \text{logit}(p_{Bi})$$

(We use the letter μ where Reitsma et al. (2005) used θ to avoid a clash of notation with the HSROC model defined in the next section.)

The bivariate model is a random-effects model in which the logit transforms of the true sensitivity and true specificity in each study have a bivariate normal distribution across studies, thereby allowing for the possibility of correlation between them (Reitsma et al. 2005):

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \Sigma_{AB} \right\} \quad \text{with} \quad \Sigma_{AB} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

8.2   The HSROC model

The HSROC model (Rutter and Gatsonis 2001) was originally formulated in terms of the probability, π_{ij}, that a patient in study i with disease status j has a positive test result, where $j = 0$ for a patient without the disease and $j = 1$ for a patient with the disease. Therefore, sensitivity $p_{Ai} = \pi_{i1}$ and specificity $p_{Bi} = 1 - \pi_{i0}$.

The HSROC model for study i takes the form

$$\text{logit}(\pi_{ij}) = (\theta_i + \alpha_i X_{ij}) \exp(-\beta X_{ij}) \tag{1}$$

where $X_{ij} = -1/2$ for those without disease ($j = 0$) and $+1/2$ for those with disease ($j = 1$). Both θ_i and α_i are allowed to vary between studies. In the model without covariates fit by `metandi`, they are assumed to have independent normal distributions with $\theta_i \sim N(\Theta, \sigma_\theta^2)$ and $\alpha_i \sim N(\Lambda, \sigma_\alpha^2)$. The model is nonlinear in the parameter β and therefore cannot be fit in `gllamm` directly.

We can rewrite (1) as two separate equations for the logit transforms of sensitivity p_{Ai} and specificity p_{Bi}, thus connecting to the parameters μ_{Ai} and μ_{Bi} of the bivariate model above:

$$\mu_{Ai} = \text{logit}(p_{Ai}) = b^{-1}(\theta_i + \frac{1}{2}\alpha_i)$$

$$\mu_{Bi} = \text{logit}(p_{Bi}) = -b(\theta_i - \frac{1}{2}\alpha_i)$$

This tells us that μ_{Ai} and μ_{Bi} are linear combinations of two random variables, θ_i and α_i, with independent normal distributions, and that they therefore must have a bivariate normal distribution. Some straightforward further algebra gives the explicit relationship between the parameters of the two models (Harbord et al. 2007; Arends et al.

2008), enabling HSROC parameter estimates to be obtained by transforming the bivariate parameter estimates. Standard errors for the transformed parameter estimates are obtained by the delta method, which gives the same standard errors that would be obtained from standard maximum-likelihood methods if the HSROC model were fit directly (Cox 1998).

8.3   Methods and formulas for metandiplot

HSROC curve

The HSROC model gives rise to an SROC curve by allowing the threshold parameter, θ_i, to vary while holding the accuracy parameter, α_i, fixed at its mean, Λ. The expected sensitivity for a given specificity is then given by (Rutter and Gatsonis 2001; Macaskill 2004)

$$\text{logit(sensitivity)} = \Lambda e^{-\beta/2} - e^{-\beta}\text{logit(specificity)}$$

Bivariate confidence and prediction regions

Confidence and prediction regions in SROC space can be constructed by using the estimates from the bivariate model (Reitsma et al. 2005; Harbord et al. 2007). An elliptical joint confidence region for μ_A and μ_B is most easily specified by using a parametric representation (Douglas 1993)

$$\mu_A = \widehat{\mu}_A + s_A \, c \, \cos t \tag{2}$$

$$\mu_B = \widehat{\mu}_B + s_B \, c \, \cos(t + \arccos r) \tag{3}$$

where s_A and s_B are the estimated standard errors of $\widehat{\mu}_A$ and $\widehat{\mu}_B$, r is the estimate of their correlation, and varying t from 0 to 2π generates the boundary of the ellipse. The constant c has been called the boundary constant of the ellipse (Alexandersson 2004); $c = \sqrt{2f_{2,n-2;\alpha}}$, where n is the number of studies and $f_{2,n-2;\alpha}$ is the upper $100\alpha\%$ point of the F distribution with degrees of freedom 2 and $n-2$ (Douglas 1993; Chew 1966). This ellipse is then back-transformed to conventional ROC space to give a confidence region for the summary operating point.

A prediction region giving the region that has a given probability (e.g., 95%) of including the *true* sensitivity and specificity of a future study is generated similarly. The covariance matrix for the true logit sensitivity and logit specificity in a future study is

$$\Sigma_{AB} + \text{Var}\begin{pmatrix}\widehat{\mu}_A\\\widehat{\mu}_B\end{pmatrix}$$

In practice, both terms are estimated by fitting the model to the data. The parameters s_A, s_B, and r in (2) and (3) can then be replaced by the corresponding quantities derived from this covariance matrix to give the prediction ellipse in logit ROC space, which is then back-transformed to a prediction region for the true sensitivity and specificity of a future study in conventional ROC space.

8.4 Methods and formulas for predict

If `metandi` fit the model by using `gllamm`, then `predict` after `metandi` uses `gllapred`; see Rabe-Hesketh, Skrondal, and Pickles (2004). If `metandi` fit the model by using `xtmelogit`, `predict` after `metandi` uses the prediction facilities of `xtmelogit`; see [XT] **xtmelogit postestimation**.

9   Acknowledgments

Joseph Coveney first worked out how to fit the bivariate model by using `gllamm` and posted the syntax on Statalist in response to a query from Ben Dwamena; our thanks to Joe for generous email correspondence. We thank the authors of `gllamm` for all their work, and Sophia Rabe-Hesketh in particular for helpful email correspondence. Our thanks also to Susan Mallett and Jon Deeks for useful feedback on earlier versions of `metandi`.

10   References

Alexandersson, A. 2004. Graphing confidence ellipses: An update of ellip for Stata 8. *Stata Journal* 4: 242–256.

Arends, L. R., T. H. Hamza, J. C. van Houwelingen, M. H. Heijenbrok-Kal, M. G. M. Hunink, and T. Stijnen. 2008. Bivariate random effects meta-analysis of ROC curves. *Medical Decision Making* 28: 621–638.

Bradburn, M. J., J. J. Deeks, and D. G. Altman. 1998. sbe24: metan—an alternative meta-analysis command. *Stata Technical Bulletin* 44: 4–15. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 86–100. College Station, TX: Stata Press.

Chew, V. 1966. Confidence, prediction, and tolerance regions for the multivariate normal distribution. *Journal of the American Statistical Association* 61: 605–617.

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

Coveney, J. 2004. Re: st: bivariate random effects meta-analysis of diagnostic test. Statalist archive. Available at http://www.stata.com/statalist/archive/2004-04/msg00820.html.

Cox, C. 1998. Delta method. In *Encyclopedia of Biostatistics*, ed. P. Armitage and T. Colton, 1125–1127. New York: Wiley.

Cox, D. R., and E. J. Snell. 1989. *Analysis of Binary Data*. 2nd ed. London: Chapman & Hall.

Deeks, J. J. 2001a. Systematic reviews of evaluations of diagnostic and screening tests. In *Systematic Reviews in Health Care: Meta-Analysis in Context*, 2nd edition, ed. M. Egger, G. Davey Smith, and D. G. Altman, 248–282. London: BMJ Books.

———. 2001b. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal* 323: 157–162.

Devillé, W. L., F. Buntinx, L. M. Bouter, V. M. Montori, H. C. W. de Vet, D. A. W. M. van der Windt, and P. D. Bezemer. 2002. Conducting systematic reviews of diagnostic studies: Didactic guidelines. *BMC Medical Research Methodology* 2: 9.

Douglas, J. B. 1993. Confidence regions for parameter pairs. *American Statistician* 47: 43–45.

Gatsonis, C., and P. Paliwal. 2006. Meta-analysis of diagnostic and screening test accuracy evaluations: Methodologic primer. *American Journal of Roentgenology* 187: 271–281.

Gluud, C., and L. L. Gluud. 2005. Evidence based diagnostics. *British Medical Journal* 330: 724–726.

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

Lijmer, J. G., P. M. M. Bossuyt, and S. H. Heisterkamp. 2002. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Statistics in Medicine* 21: 1525–1537.

Macaskill, P. 2004. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *Journal of Clinical Epidemiology* 57: 925–932.

Mallett, S., J. J. Deeks, S. Halligan, S. Hopewell, V. Cornelius, and D. G. Altman. 2006. Systematic reviews of diagnostic tests in cancer: Review of methods and reporting. *British Medical Journal* 333: 413–416.

Moses, L. E., D. Shapiro, and B. Littenberg. 1993. Combining independent studies of a diagnostic test into a summary ROC curve: Data-analytic approaches and some additional considerations. *Statistics in Medicine* 12: 1293–1316.

Nordic Cochrane Centre. 2007. *Review Manager (RevMan): Version 5*. Software program. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2004. GLLAMM manual. Working Paper 160, Division of Biostatistics, University of California–Berkeley. Downloadable from http://www.bepress.com/ucbbiostat/paper160/.

———. 2005. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* 128: 301–323.

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

Riley, R. D., K. R. Abrams, A. J. Sutton, P. C. Lambert, and J. R. Thompson. 2007. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology* 7: 3.

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

Scheidler, J., H. Hricak, K. K. Yu, L. Subak, and M. R. Segal. 1997. Radiological evaluation of lymph node metastases in patients with cervical cancer: A meta-analysis. *Journal of the American Medical Association* 278: 1096–1101.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Boca Raton, FL: Chapman & Hall/CRC.

Sterne, J. A. C., R. J. Harris, R. M. Harbord, and T. J. Steichen. 2007. What meta-analysis features are available in Stata? Stata FAQ. Available at http://www.stata.com/support/faqs/stat/meta.html.

Tatsioni, A., D. A. Zarin, N. Aronson, D. J. Samson, C. R. Flamm, C. Schmid, and J. Lau. 2005. Challenges in systematic reviews of diagnostic technologies. *Annals of Internal Medicine* 142: 1048–1055.

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

Zwinderman, A. H., and P. M. Bossuyt. 2008. We should not pool diagnostic likelihood ratios in systematic reviews. *Statistics in Medicine* 27: 687–697.

About the authors

Roger Harbord is a research associate in medical statistics in the Department of Social Medicine at the University of Bristol, UK. He is one of three co-convenors of the Cochrane Collaboration's Screening and Diagnostic Tests Methods Group.

Penny Whiting is a research fellow in the same department who specializes in methodology and conduct of diagnostic systematic reviews. She led the development of QUADAS, a 14-item evidence-based tool for assessing the quality of diagnostic accuracy studies included in systematic reviews (Cox and Snell 1989).