1 midas command

1.1 Description

midas is a comprehensive program of statistical and graphical routines for undertaking meta-analysis of diagnostic test performance in Stata. The index and reference tests (gold standard) are dichotomous. Primary data synthesis is performed within the bivariate mixed-effects regression framework focused on making inferences about average sensitivity and specificity. The bivariate approach was originally developed for treatment trial meta-analysis (Houwelingen et al. 1993; van Houwelingen et al. 2002) and modified for synthesis of diagnostic test data using an approximate normal within-study model (Reitsma et al. 2005; Riley et al. 2007a,b, 2008; Arends et al. 2008). An exact binomial rendition (Chu and Cole 2006; Riley et al. 2007b; Arends et al. 2008) of the bivariate model assumes independent binomial distributions for the true positives and true negatives conditional on the sensitivity and specificity in each study. Likelihood-based estimation of the exact binomial approach may be performed by adaptive gaussian quadrature using Stata-native **xtmelogit** command (Stata release 10) or **gllamm** (Rabe-Hesketh et al. 2004, 2002), user-written command, both with readily available post-estimation procedures for model diagnostics and empirical Bayes predictions. Additionally, **midas** facilitates exploratory analysis of heterogeneity (unobserved, threshold-related and covariate etc.), publication and other precision-related biases. Bayes' nomograms, likelihood-ratio matrices, and probability modifying plots may be derived and used to guide patient-based diagnostic decision making.

1.2 Syntax

`midas` *varlist*`(min=4 max=4)` [*if*] [*in*] `,` [`options`]

The required *varlist* is the data from the contingency tables of index and reference test results. The user provides the data in a rectangular array containing variables for the 2x2 elements *a*, *b*, *c* and *d*. Each data row contains the 2x2 data for one observation(i.e. study). The *varlist* MUST contain variables for *a*, *b*, *c* and *d* in that order:

	Reference Test Positive	Reference Test Negative
Test Positive	a = true positives	c = false negatives
Test Negative	b = false positives	d = true negatives

1.3 Options

Modeling and Post-estimation Options

nip(integer) specifies the number of integration points used for maximum likelihood estimation based on adaptive gaussian quadrature. Default is set at 1 for midas even though the default in **xtmelogit** is 7. Higher values improve accuracy at the expense of execution times. Using **xtmelogit** with nip(1), model will be estimated by Laplacian approximation. This decreases substantially computational time and yet provides reasonably valid fixed effects estimates. It may, however, produce biased estimates of the variance components.

ebpred(for|roc) generates a forest plot or roc curve of empirical Bayes versus observed estimates of sensitivity and specificity.

modchk(gof|bvn|inf|out|all) provides graphical model checking capabilities: **modchk**(gof) displays quantile plot of residual-based goodness-of fit; **modchk**(bvn) displays Chi-squared probability plot of squared Mahalanobis distances for assessment of the bivariate normality assumption; **modchk**(inf) spikeplot for checking for particularly influential observations using Cook's distance; **modchk**(out) displays a scatter plot for checking for outliers using standardized predicted random effects (standardized level-2 residuals); and **modchk**(all) provides a composite graphic of all four plots.

Quality Assessment Options

qtab(*varlist***)** creates a table showing frequency of methodological quality items.

qbar(*varlist***)** calculates study-specific quality scores and plots a bargraph of methodological quality.

qlab may be combined with **qtab(***varlist***)** or **qbar(***varlist***)** to use variable labels for table and bargraph of methodological items.

Exploratory Graphics

bivbox implements a bivariate generalization of the box plot for univariate data similar to the bivariate box plot(Rousseeuw et al. 1999). It is used to assess location, spread, correlation, skewness and tails of the data and for identifying possible outliers.

chiplot creates a chiplot(Fisher and Switzer 2001) for judging whether or nor the paired performance indices are independent by augmenting the scatter plot with an auxiliary display. In the case of independence, the points will be concentrated in the central region, in the horizontal band indicated on the plot.

Main Reporting Options

results(all) provide summary statistics for all performance indices, group-specific between-study variances, likelihood ratio test statistics and other global homogeneity tests.

results(het) provide group-specific between-study variances, likelihood ratio test statistics and other global homogeneity tests.

results(sum) provides summary statistics for all performance indices i.e. sensitivity/specificity, positive/negative likelihood ratios and diagnostic score/odds ratios.

table(dss|dlor|dlr) will create a table of study specific performance estimates with measure-specific summary estimates and results of homogeneity (chi-square) and inconsistency(I squared)tests. **dss, dlr or dlor** represent the paired performance measures sensitivity/specificity, positive/negative likelihood ratios and diagnostic score/odds ratios.

Forest Plots Options

id(*varlist***)** provides a label for studies allowing up to 4 variables.

bforest(dss|dlr|dlor) creates summary graphs with study-specific(box) and overall(diamond) point estimates and confidence intervals for each performance index pair using **graph combine** see [G] **graph: graph combine**.

uforest(dss|dlr|dlor) creates univariate summary graphs with study-specific(box) and overall(diamond) point estimates and confidence intervals allowed to extend between 0 and 1000 beyond which they are truncated and marked by a leading arrow.

fordata adds study-specific performance estimates and 95% CIs to right y-axis.

forstats adds heterogeneity statistics below summary point estimate.

ROC Curve Options

rocplane plots observed data in receiver operating characteristic space (ROC Plane) for visual assessment of threshold effect, a source of heterogeneity unique to diagnostic meta-analysis. The higher the cut-off value, the higher will be the specificity and the lower the sensitivity. This interdependence between sensitivity and specificity based

on threshold variability may be tested a priori using a rank correlation test such as Spearman's rho. In **midas**, the proportion of variation due to threshold effects is calculated as the squared correlation coefficient estimated from the between-study covariance parameter.

sroc(none|pred|conf|both) plots observed data points, summary operating sensitivity-specificity, and SROC curve without or with either or both of confidence and prediction regions at default or specified confidence level.

sroc(nnoc|pnoc|cnoc|bnoc) plots observed data points, summary operating sensitivity-specificity without or with either or both of confidence and prediction contours at default or specified confidence level. No SROC curve is plotted with these choices.

Heterogeneity Options

galb(tpr|tnr|dlor|lrp|lrn) option produces Galbraith(radial) plots of standardized logit transformed proportion (**tpr**, **tnr**) or log-transformed ratio**(lrp**, **lrn** and **dlor)** against the inverse of the its precision(x-axis). A regression line that goes through the origin is calculated, together with 95% boundaries (starting at +2 and -2 on the y-axis). Studies outside these 95% boundaries may be considered as outliers.

regvars(*varlist***)** permits univariable meta-regression analysis of one or multiple dichotomous or continuous covariables, reporting results in table and forest plot

Publication Bias Options

pubbias When this option is invoked, **midas** performs linear regression of log odds ratios on inverse root of effective sample sizes as a test for funnel plot asymmetry in diagnostic meta-analysis. A non-zero slope coefficient is suggestive of significant small study bias (p value < 0.10). The regression line is superimposed on a funnel plot (see figure 8).

Clinical Utility Options

fagan(0-0.99) creates a plot showing the relationship between the prior probability specified by user, the likelihood ratio(combination of sensitivity and specificity), and posterior test probability.

pddam(lbp ubp) produces a line graph of post-test probabilities versus prior probabilities between 0 and 1 using summary likelihood ratios. Summary unconditional

predictive values based on sensitivity analysis using uniformly distributed prior probabilities between **lbp and ubp** are estimated and displayed on graph.

lrmatrix creates a scatter plot of positive and negative likelihood ratios with combined summary point. Plot is divided into quadrants based on strength-of-evidence thresholds to determine informativeness of measured test.

Miscellaneous Options

level(#) specifies the significance level for statistical tests, confidence regions, prediction regions and confidence intervals.

mscale(#) affects size of markers for point estimates on forest plots.

scheme(*string*) permits choice of scheme for graphs. The default is s2color.

textscale(#) allows choice of text size for graphs especially regarding labels for forest plots.

zcf(#) defines a fixed continuity correction in the case where a study contains a zero cell during logit or log transformations only to calculate study-specific likelihood ratios and odds ratios. By default, **midas** adds 0.5 to each cell of a study where a zero is encountered. However, the **zcf(#)** option allows the use of other constants between 0 and 1.

❑ Technical note

Although **midas** has pre-programmed graphical options, with Graph Editor (Stata Release 10 or later), you can change almost anything on your graph. You can add text, lines, arrows, and markers wherever you would like. You can right-click on any object to see a list of operations specific to the object and tool you are working with. This feature is most useful with the Pointer tool. However, there is no record of what you have done with the graph editor, so if you need to recreate the graph for some reason, you will have to redo everything that you have done with the graph editor. See [G] **graph editor**.

❑

1.4 Saved results

midas saves results as scalars in **r()** such as:

r(mtpr)	Mean sensitivity
r(mtprse)	standard error of mean sensitivity
r(mtnr)	Mean specificity
r(mtnrse)	Standard error of mean specificity
r(mlrp)	Mean likelihood ratio of a positive test result
r(mlrpse)	Standard error of mean likelihood ratio of a positive test result
r(mlrn)	Mean likelihood ratio of a negative test result
r(mlrnse)	Standard error of mean likelihood ratio of a negative test result
r(mdor)	Mean diagnostic odds ratio
r(mdorse)	Standard error of mean diagnostic odds ratio
r(AUC)	Area under summary ROC curve
r(AUClo)	Lower bound of area under summary ROC curve
r(AUChi)	Upper bound of area under summary ROC curve
r(rho)	Correlation between logits of sensitivity and specificity
r(reffs1)	Variance of logit of sensitivity
r(reffs1se)	Standard error of variance of logit of sensitivity
r(reffs2)	Variance of logit of specificity
r(reffs2se)	Standard error of variance of logit of specificity
r(Islrt)	Global inconsistency index from likelihood ratio test
r(Islrtlo)	Lower bound global inconsistency index
r(Islrthi)	Upper bound global inconsistency index

2 Example Dataset

This dataset was obtained as part of a systematic review and meta-analysis of the published literature on the staging performance of axillary positron emission tomography (FDG-PET) in breast cancer toward identification of the number, quality and scope of primary studies; quantification of overall classification performance (sensitivity and specificity), discriminatory power (diagnostic odds ratios) and informational value (diagnostic likelihood ratios); assessment of the impact of technical characteristics of test, methodological quality of primary studies and publication selection bias on estimates of diagnostic accuracy; and highlighting of any potential issues that require further research. We performed a comprehensive computer search of the English Language medical literature using primarily the PUBMED (MEDLINE) search engine and cross-citation with other databases to identify original peer-reviewed full-length human subject articles published between January 1, 1990 and January 31, 2008. Search was conducted using database-specific Boolean search strategies based on: ("breast neoplasm" OR "breast cancer" OR "breast malignancy") AND ("Positron emission tomography" OR "FDG-PET") AND ("axillary staging" OR "axillary metastases" OR "axillary node metastases" OR "axillary node staging"). Search strategies were augmented with a manual search of reference lists from identified articles and recent subject-area journals for additional articles. Details of the data set and 2 by 2 data for the first 20 studies

as shown below were generated respectively with the **describe** and **list** commands of Stata:

```
. describe

Contains data from f:\breastpet.dta
  obs:            39
 vars:            33                          7 Jan 2009 16:06
 size:         2,652 (99.9% of memory free)
```

variable name	storage type	display format	value label	variable label
author	str13	%-15s		first author
year	int	%8.0g		Year of Publication
tp	int	%8.0g		True Positive
fp	byte	%8.0g		False Positive
fn	float	%9.0g		False Negative
tn	float	%9.0g		True Negative
period	byte	%8.0g		Period of publication(before or after 1998)
prodesign	byte	%10.0g		Prospective design
ssize30	byte	%7.0g		Study size greater than 30
fulverif	byte	%8.0g		Full Verification of test results
testdescr	byte	%9.0g		Satisfactory description of index test
refdescr	byte	%9.0g		Statisfactory description of ref test
clindescr	byte	%9.0g		Clinical Information Available
report	byte	%7.0g		Satisfactory reporting of results
consel	byte	%7.0g		Consecutive selection of subjects
brdspect	byte	%10.0g		Broad spectrum of disease
blindref	byte	%7.0g		blinded reference test interpretation
sampsize	int	%8.0g		Number Data Units Per Study
qscore	byte	%9.0g		Quality score
age	byte	%8.0g		Mean Age
fast	byte	%8.0g		Prior Fasting of at leasst 4-6 hours
res	byte	%8.0g		Resolution of PET Camera
dose	int	%8.0g		Appropriate dose used
upttime	byte	%8.0g		Uptake Period specified
acquitime	byte	%8.0g		Duration of image acqusition
testcrit	byte	%8.0g		Test criteria described
multiobs	byte	%8.0g		Multiple Observers
attcor	byte	%9.0g		Attenuation correction of PET images
axonly	byte	%8.0g		Dedicated axillary imaging
pmet	float	%9.2g		Study-specfic Prevalence of axillary Metastases
Analysis	str7	%9s		Unit of Data Analysis(Patient versus Nodes)
patient	byte	%10.0g		Mode of analysis
blindtest	byte	%8.0g		blinded index test interpretation

(*Continued on next page*)

```
. list author year pmet sampsize tp fp fn tn in 1/20, sep(1) ab(32) abs noo
```

author	year	pmet	sampsize	tp	fp	fn	tn
Tse	1992	.7	10	4	0	3	3
Adler1	1993	.47	18	8	0	1	10
Hoh	1993	.64	14	6	0	3	5
Crowe	1994	.5	20	9	0	1	10
Avril	1996	.47	51	19	1	5	26
Bassa	1996	.81	16	10	0	3	3
Scheidhauer	1996	.5	18	9	1	0	8
Utech	1996	.35	124	44	20	0	60
Adler2	1997	.38	50	19	11	0	20
Palmedo	1997	.3	20	5	0	1	14
Noh	1998	.54	24	12	0	1	11
Smith	1998	.42	50	19	1	2	28
Rostom	1999	.65	74	42	0	6	26
Yutani1	1999	.38	26	8	0	2	16
Hubner	2000	.27	22	6	0	0	16
Ohta	2000	.59	32	14	0	5	13
Yutani2	2000	.42	38	8	0	8	22
Greco	2001	.43	167	68	13	4	82
Schirrmeister	2001	.3	113	27	6	7	73
Yang	2001	.33	18	3	0	3	12

3 Annotated Worked Examples

3.1 Model prediction and diagnostics

Post-estimation predictions may be obtained from the estimated model, parameter estimates and empirical Bayes estimates of the random effects with standard errors computed with the delta method. The predicted logits may then be transformed to obtain predictions of sensitivity and specificity. **midas** generates a forest plot or roc curve

of empirical Bayes versus observed estimates of sensitivity and specificity using the **ebpred**(for|roc) options. Figure 1 was obtained using the syntax below:

```
. midas tp fp fn tn, eb(for)
```

Model diagnostics are rarely performed because many non-statisticians and applied researchers consider meta-analysis as a data-processing procedure rather than a model-fitting exercise (Sutton and Higgins 2008). However, with the application of complex likelihood-based meta-analytic models, it is important to evaluate possible model misspecification, goodness of fit, and to identify outlying and possibly influential data points. **midas** provides graphical model checking capabilities: quantile plot of residual-based goodness-of fit; Chi-squared probability plot of squared Mahalanobis distances for assessment of the bivariate normality assumption; spikeplot for checking for particularly influential observations using Cook's distance; a scatter plot for checking for outliers using standardized predicted random effects (standardized level-2 residuals); and composite graphic of all four plots such as figure 2 which was obtained using the syntax below:

```
. midas tp fp fn tn,  modchk(all)
```

3.2 Quality Assessment

Methodologic quality of a study is the extent to which all aspects of a study's design and conduct can be shown to protect against systematic bias, nonsystematic bias that may arise in poorly performed studies, and inferential error. The recently developed quality assessment tool for diagnostic accuracy studies (QUADAS) (Whiting et al. 2005, 2004, 2003, 2006), is a rigorously constructed and validated tool that can be used by investigators undertaking new systematic reviews. The QUADAS tool consists of 14 items that cover patient spectrum, reference standard, disease progression bias, verification and review bias, clinical review bias, incorporation bias, test execution, study withdrawals, and intermediate results. Possible methods to address quality differences are sensitivity analysis, subgroup analysis, or meta-regression analysis. Quality assessment may be summarized by stacked bars for each QUADAS item in a bar graph. **midas** provides the ability to represent the results of quality assessment by means of a bar graph. For example, figure 3 was obtained with the command syntax:

```
. midas tp fp fn tn,  qbar(prodesign fulverif testdescr refdescr clindescr report
> spectrum blindref blindtest)  qlab
```

Alternatively using the qtab(*varlist*) produces frequency table of quality scores.

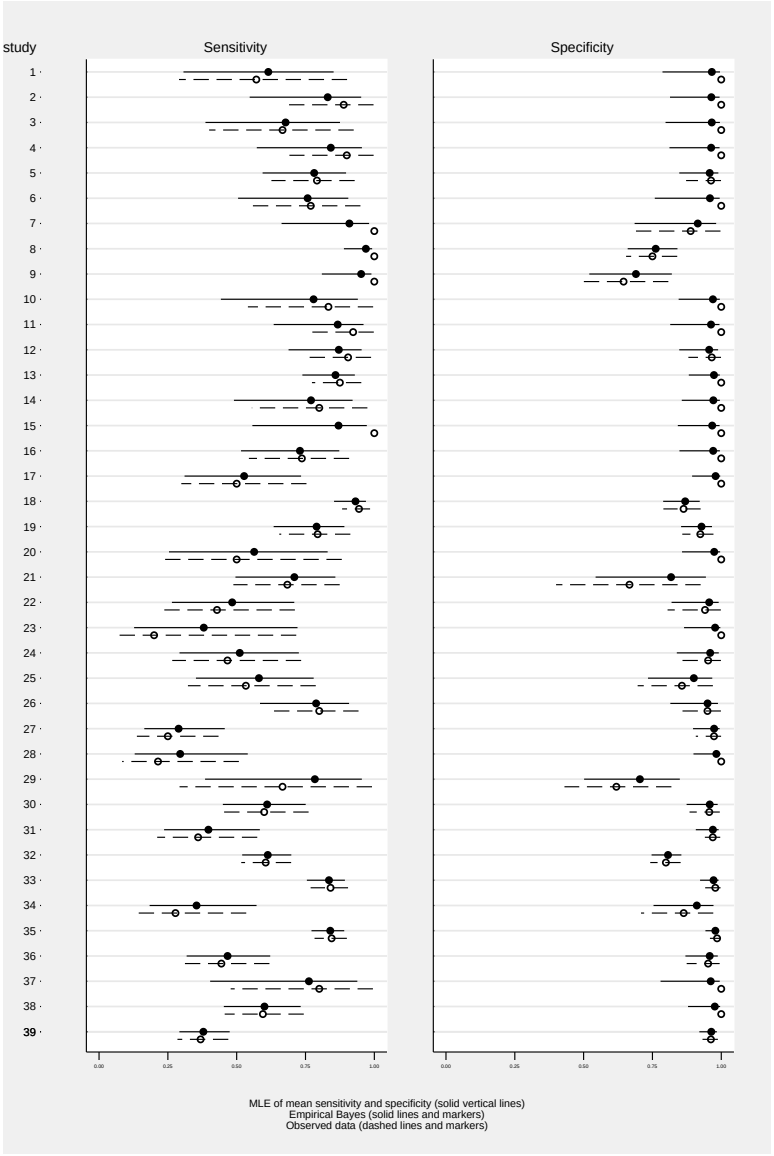

Figure 1: Paired forest plot depiction of empirical Bayes predicted versus observed sensitivity and specificity

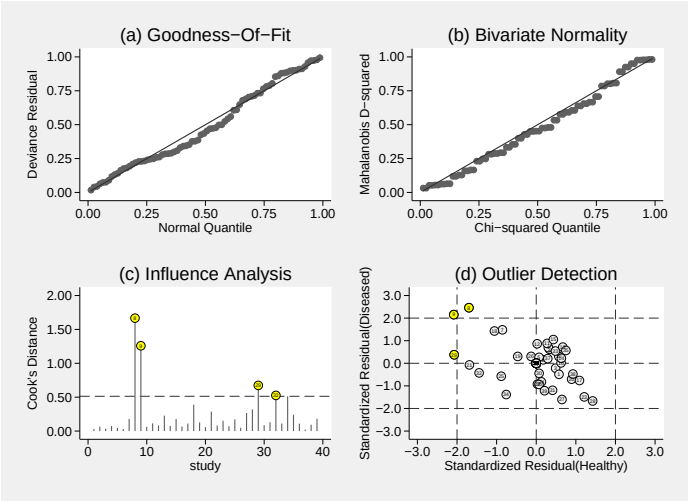

Figure 2: Graphical depiction of residual-based goodness-of-fit, bivariate normality, influence and outlier detection analyses

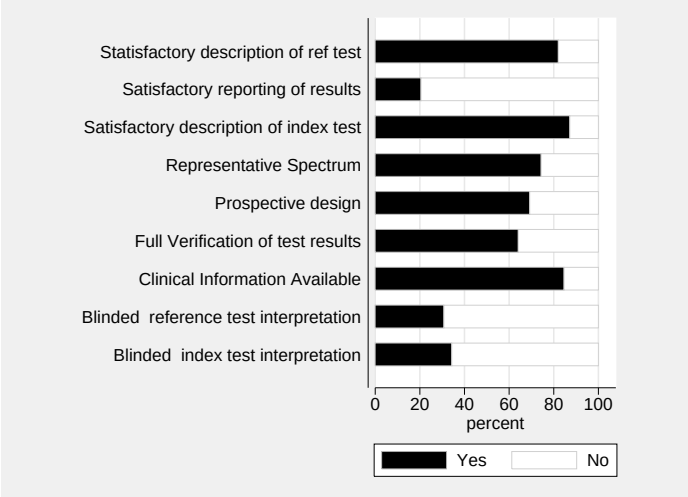

Figure 3: Bar graph of quality assessment displaying stacked bars for each quality item including modified QUADAS criteria

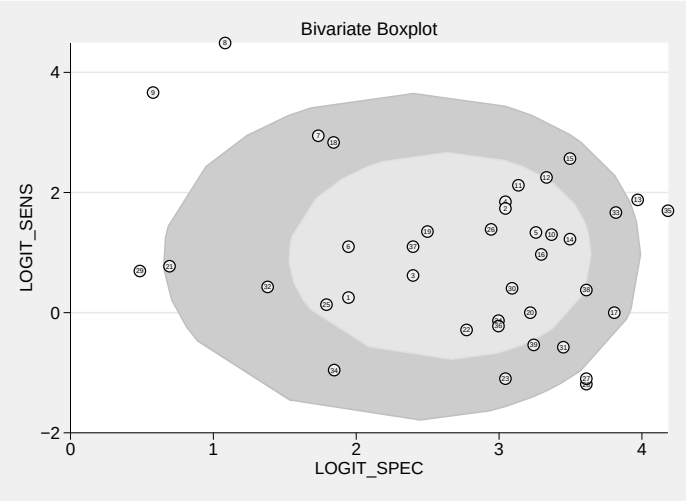

Figure 4: Bivariate box plot with most studies clustering within the median distribution with seven outliers suggesting indirectly a lower degree of heterogeneity

3.3 Bivariate Association

midas uses a bivariate random effects modeling of sensitivity and specificity, therefore, it is expected that this pair of performance measures will be interdependent (see Figure 3). The bivariate box plot describes the degree of interdependence including the central location and identification of any outliers. The inner oval represents the median distribution of the data points. The outer oval represents the 95% confidence bound. We obtained figure 4 with the syntax:

```
. midas tp fp fn tn, bivbox
```

It demonstrates a skewedness of the test performance measures toward a higher specificity with lower sensitivity, providing indirect evidence of some threshold variability.

3.4 Summary Performance Estimates

Summary estimates of sensitivity and specificity and their 95% confidence intervals can be calculated after anti-logit transformation of of the mean logit sensitivity and logit specificity and respective standard errors. These intervals take into account the heterogeneity beyond chance between studies (random effects model).

```
. midas tp fp fn tn, res(sum) nip(1)
```

(*Continued on next page*)

```
SUMMARY PERFORMANCE ESTIMATES
Parameter                    Estimate          95% CI
Sensitivity                      0.73 [    0.63,    0.80]
Specificity                      0.96 [    0.92,    0.97]
Positive Likelihood Ratio        16.2 [     9.7,    27.3]
Negative Likelihood Ratio        0.29 [    0.21,    0.39]
Diagnostic Odds Ratio              57 [      30,     107]
```

3.5 Heterogeneity Statistics

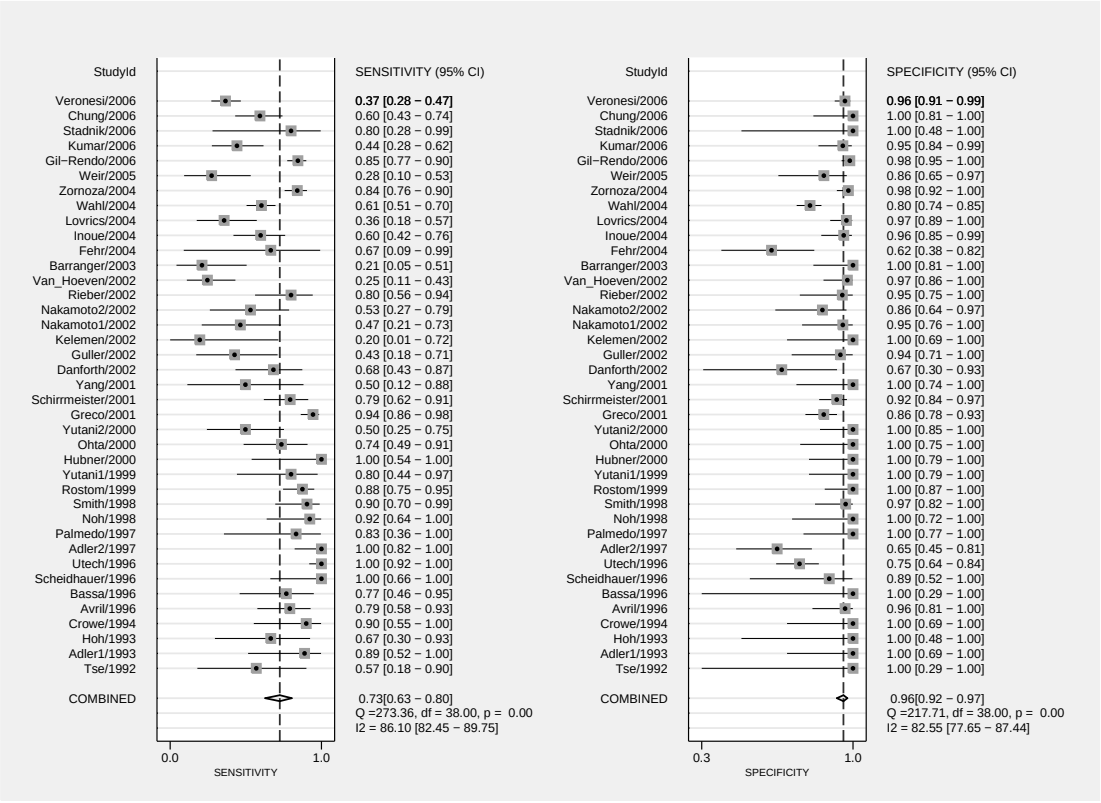

Figure 5: Forest plot showing study-specific (right-axis) and mean sensitivity and specificity with corresponding heterogeneity statistics

The impact of unobserved heterogeneity is traditionally assessed statistically using the quantity I^2. It describes the percentage of total variation across studies that is attributable to the heterogeneity rather than chance(Higgins and Thompson 2002).

I^2 can be calculated from heterogeneity statistic and the degrees of freedom (Higgins and Thompson 2002). Alternatively, for mixed models, it is the intraclass correlation coefficient expressed as a percentage with similar interpretation and is implemented in **midas** separately for sensitivity and specificity. I^2 lies between 0% and 100%. A value of 0% indicates no observed heterogeneity, and values greater than 50% may be considered substantial heterogeneity. The main advantage of I^2 is that it does not inherently depend on the number of studies in the meta-analysis (Higgins and Thompson 2002). The **midas** output of heterogeneity statistics is shown below and was generated from the syntax:

```
. midas tp fp fn tn, res(het) nip(1)
HETEROGENEITY STATISTICS
Heterogeneity (Chi-square): LRT_Q = 146.822, df =2.00, LRT_p =0.000
Inconsistency (I-square): LRT_I2 =  99, 95% CI = [ 98- 99]
Proportion of heterogeneity likely due to threshold effect = 0.11
Interstudy variation in Sensitivity: ICC_SEN =  0.31, 95% CI = [ 0.18- 0.45]
Interstudy variation in Sensitivity: MED_SEN =  0.76, 95% CI = [ 0.70- 0.83]
Interstudy variation in Specificity: ICC_SPE =  0.29, 95% CI = [ 0.12- 0.45]
Interstudy variation in Specificity: MED_SPE =  0.75, 95% CI = [ 0.68- 0.84]
```

3.6 Forest plot to demonstrate study-specific sensitivity and specificity on right y-axis

Within **midas**, forest plots can be created for each test performance parameter individually or may be displayed as paired plots, for example, sensitivity paired with specificity. The user has the option of displaying heterogeneity statistics within the forest plots using the option **forstats**. Figure 5 was obtained using the command syntax:

```
. midas tp fp fn tn, texts(0.60) bfor(dss) id(author year) ford fors
```

3.7 Summary ROC Curve

Based on parameters estimated by the bivariate model, several summary ROC linear regression lines based on either the regression of logit sensitivity on specificity, the regression of logit specificity on sensitivity, or an orthogonal regression line by minimizing the perpendicular distances may be derived. These lines can be transformed back to the original ROC scale to obtain a summary ROC curve. In **midas**, derived logit estimates of sensitivity, specificity and respective variances are used to construct a hierarchical summary ROC curve. **midas** has several options for the display of the ROC curve. For example, figure 6 was obtained by using the syntax below:

```
. midas tp fp fn tn, sroc(both)
```

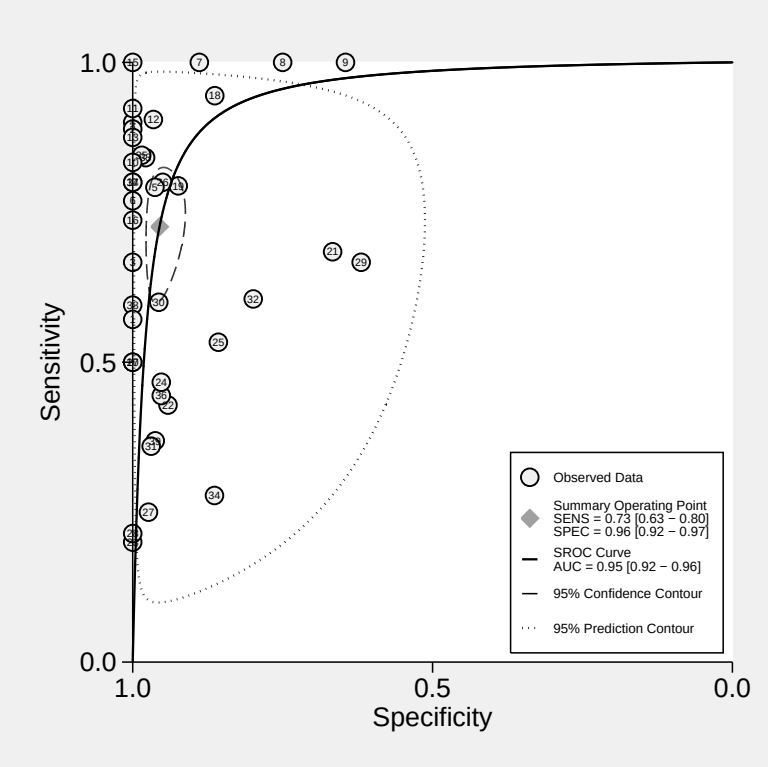

Figure 6: Summary ROC curve with confidence and prediction regions around mean operating sensitivity and specificity point

The summary ROC curve is displayed along with the observed study data. The dashed line around the summary point estimate, represents the 95% confidence region. The area under the curve (AUROC), serves as a global measure of test performance. The AUROC is the average TPR over the entire range of FPR values. The following guidelines have been suggested for interpretation of intermediate AUROC values: low (0.5>= AUC <= 0.7), moderate (0.7 >= AUC <= 0.9), or high (0.9 >= AUC <= 1) accuracy (Swets 1988).

3.8 Meta-regression

Meta-regression, the use of regression methods to incorporate the effect of covarying factors on summary measures of performance, has been used to explore between-study heterogeneity in therapeutic studies. In diagnostic studies, likewise, heterogeneity in sensitivity and specificity can result from many causes related to definitions of the test and reference standards, operating characteristics of the test, methods of data collection, and patient characteristics. Covariates may be introduced into a regression with any test performance measure as the dependent variable. As with any meta-regression, however, the sample size will correspond to the number of studies in the analysis with small number of studies limiting the power of regression to detect significant effects. The syntax below produces both tabular and graphical output (Figure 7):

```
. midas tp fp fn tn,  reg(prodesign age qscore sampsize fulverif testdescr refd
> escr clindescr report spectrum blindref blindtest)
```

(*output omitted*)

Parameter	category	LRTChi2	Pvalue	I2	I2lo	I2hi
prodesign	Yes	2.73	0.26	27	0	100
	No	.	.	.	.	.
age		7.18	0.03	72	38	100
qscore		4.84	0.09	59	7	100
sampsize		0.83	0.66	0	0	100
fulverif	Yes	0.03	0.99	0	0	100
	No	.	.	.	.	.
testdescr	Yes	1.79	0.41	0	0	100
	No	.	.	.	.	.
refdescr	Yes	0.61	0.74	0	0	100
	No	.	.	.	.	.
clindescr	Yes	6.75	0.03	70	34	100
	No	.	.	.	.	.
report	Yes	2.75	0.25	27	0	100
	No	.	.	.	.	.
spectrum	Yes	1.15	0.56	0	0	100
	No	.	.	.	.	.
blindref	Yes	8.35	0.02	76	48	100
	No	.	.	.	.	.
blindtest	Yes	25.80	0.00	92	85	99
	No	.	.	.	.	.

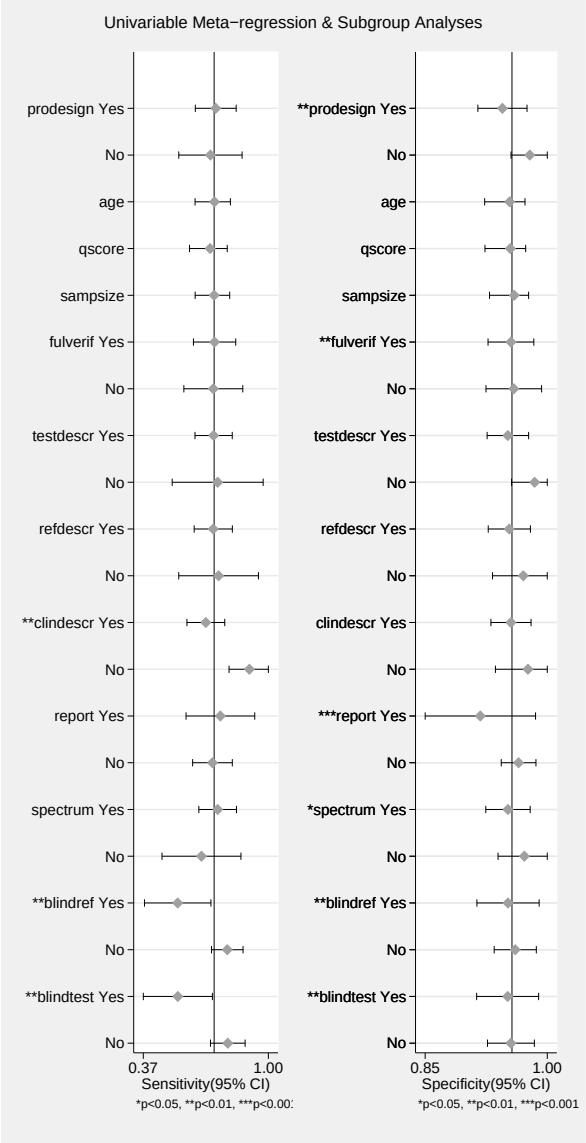

Figure 7: Forest plot of multiple univariable meta-regression and subgroup analyses

3.9 Linear regression test of funnel plot asymmetry

The funnel plot is generally considered a good exploratory tool for investigating publication bias (Light and Pillemer 1984), plotting a measure of effect size against a measure of study precision, appearing symmetric if no bias is present. However, assessment of such a plot is very subjective. Thus, non-parametric and parametric linear regression methods have been developed to formally test for such funnel plot asymmetry (Begg and Mazumdar 1994; Egger and Smith 1998; Harbord et al. 2006; Macaskill et al. 2001; Peters et al. 2006; Rucker et al. 2008; Schwarzer et al. 2007, 2002). Using these methods for assessing publication bias in diagnostic test studies may produce misleading results (Song et al. 2002; Deeks et al. 2005). Formal testing for publication bias may be conducted by a regression of diagnostic log odds ratio against 1/sqrt(effective sample size), weighting by effective sample size (Deeks et al. 2005), with P $<$ .10 for the slope coefficient indicating significant asymmetry.

```
. midas tp fp fn tn, pubbias
```

yb	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Bias	-3.206976	4.332084	-0.74	0.464	-11.98461	5.57066
Intercept	4.255449	.5420326	7.85	0.000	3.157187	5.353711

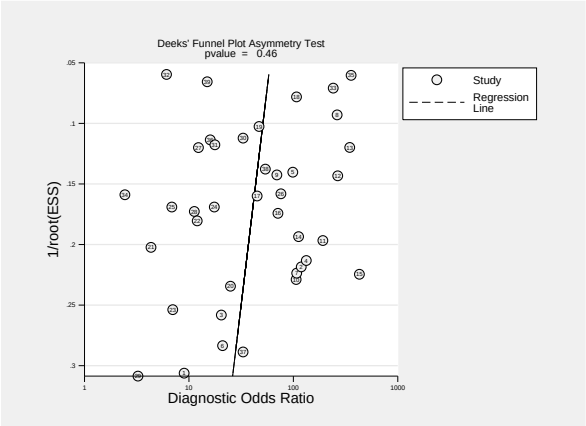

Figure 8: Funnel plot with superimposed regression line

Using the syntax above yields funnel plot with superimposed regression line as shown in figure 8. The statistically non-significant p-value (0.89) for the slope coefficient suggests symmetry in the data and a low likelihood of publication bias. However, the test is known to have low power (Deeks et al. 2005).

3.10 Fagan plot (Bayes Nomogram)

The clinical or patient-relevant utility of a diagnostic test is evaluated using the likelihood ratios to calculate post-test probability(PTP) based on Bayes' theorem as follows: Pretest Probability=Prevalence of target condition PTP= LR × pretest probability/[(1-pretest probability)× (1-LR)] This concept is depicted visually with Fagan's nomograms (Fagan 1975). When Bayes theorem is expressed in terms of log-odds, the posterior log-odds are linear functions of the prior log-odds and the log likelihood ratios. A Fagan plot (see figure 9) consists of a vertical axis on the left with the prior log-odds, an axis in the middle representing the log-likelihood ratio and an vertical axis on the right representing the posterior log-odds. Lines are then drawn from the prior probability on the left through the likelihood ratios in the center and extended to the posterior probabilities on the right.

Figure 9 demonstrates that FDG-PET is very informative raising probability of axil-

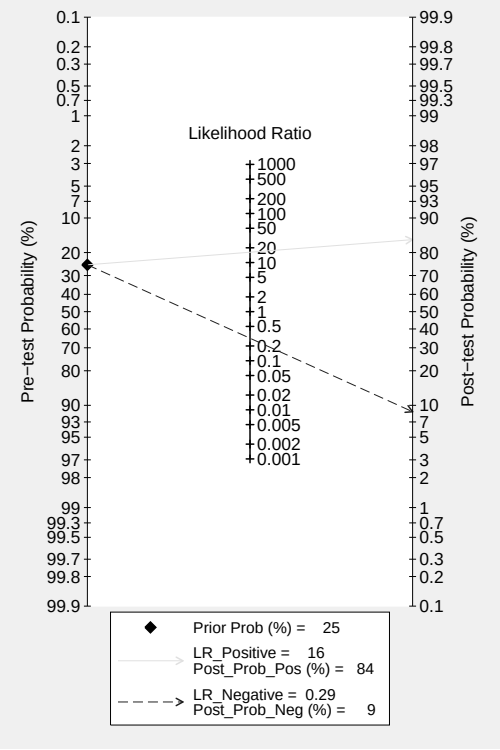

Figure 9: Fagan plot

lary breast metastases over 3-fold when positive from 25% and lowering the probability of disease to as low as 9% when negative. It was generated with the syntax:

```
. midas tp fp fn tn, fagan(0.25)
```

3.11 Likelihood ratio scattergram

Informativeness may also be represented graphically by a likelihood ratio scattergram or matrix (Stengel et al. 2003). It defines quadrants of informativeness based on established evidence-based thresholds:

1. Left Upper Quadrant, Likelihood Ratio Positive > 10, Likelihood Ratio Negative <0.1: Exclusion & Confirmation
2. Right Upper Quadrant, Likelihood Ratio Positive >10, Likelihood Ratio Negative >0.1: Confirmation Only
3. Left Lower Quadrant, Likelihood Ratio Positive <10, Likelihood Ratio Negative <0.1: Exclusion Only
4. Right Lower Quadrant, Likelihood Ratio Positive <10, Likelihood Ratio Negative >0.1: No Exclusion or Confirmation

The likelihood ratio scattergram (figure 10) shows summary point of likelihood ratios obtained as functions of mean sensitivity and specificity (Leeflang et al. 2008) in the right upper quadrant suggesting that FDG-PET is useful for confirmation of presence of axillary metastatic disease (when positive) and not for its exclusion (when negative). This figure was generated with the **midas** syntax:

```
. midas tp fp fn tn, lrmat
```

3.12 Predictive Values and Probability Modifying Plot

The conditional probability of disease given a positive OR negative test, the so-called positive (negative) predictive values are critically important to clinical application of a diagnostic procedure. They depend not only on sensitivity and specificity, but also on disease prevalence (p). The probability modifying plot is a graphical sensitivity analysis of predictive value across a prevalence continuum defining low to high-risk populations. It depicts separate curves for positive and negative tests. The user draws a vertical line from the selected pre-test probability to the appropriate likelihood ratio line and then reads the post-test probability off the vertical scale. General summary statistics have also been introduced (Li et al. 2007) for when it may be of interest to evaluate the effect of p on predictive values: unconditional positive and negative predictive values, which permit prevalence heterogeneity. These measures are obtained by integrating their corresponding conditional (on p) versions with respect to a prior distribution for p. The prior posits assumptions about the risk level in a hypothetical population of interest, e.g. low, high, moderate risk, as well as the heterogeneity in the population.

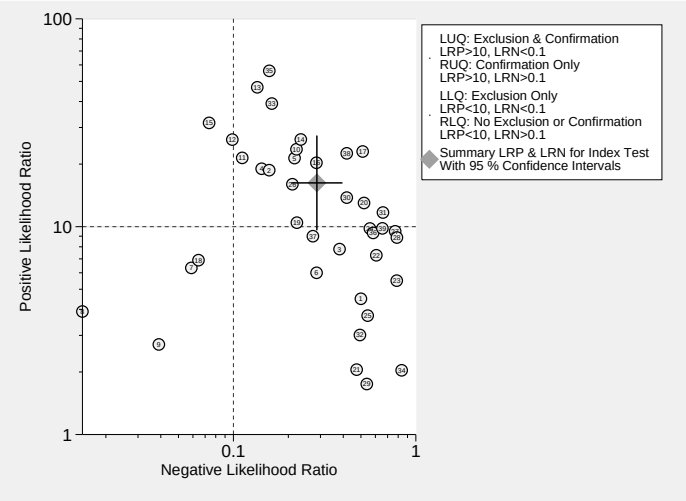

Figure 10: Likelihood ratio scattergram

```
. midas tp fp fn tn, pddam(0.25 0.75)
```

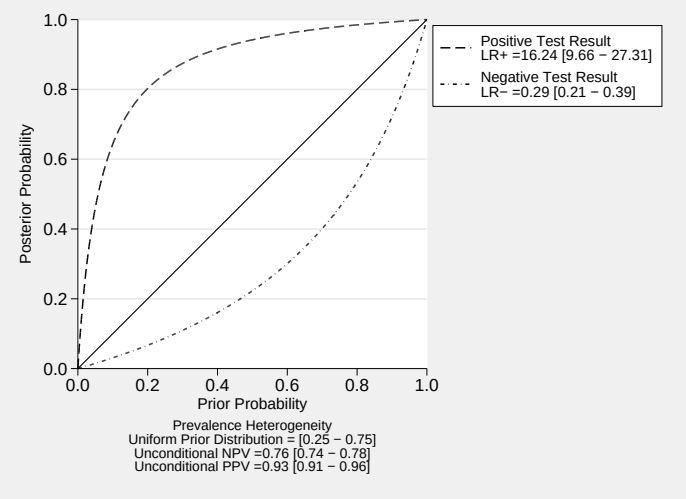

Figure 11: Probability Modifying Plot

Using the syntax above generates figure 11, which plots the relationship between pre- and post-test probability based on the likelihood of a positive (above diagonal line) or negative (below diagonal line) test result over the 0-1 range of pre-test probabilities.

Tests with more informative positive results have curves tending toward the (0,1) location while tests with more informative negative results produce curves toward the (1,0) location.

4 References

Arends, L. R., T. H. Hamza, J. C. Houwelingen, M. H. Heijenbrok-Kal, M. G. M. Hunink, and T. Stijnen. 2008. Bivariate Random Effects Meta-Analysis of ROC Curves. *Med Decis Making* 28: 621–628. http://dx.doi.org/10.1177/0272989X08319957.

Begg, C. B., and M. Mazumdar. 1994. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50: 1088–1101.

Chu, H., and S. R. Cole. 2006. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 59(12): 1331–2; author reply 1332–3. http://dx.doi.org/10.1016/j.jclinepi.2006.06.011.

Deeks, J. J., P. Macaskill, and L. Irwig. 2005. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 58(9): 882–893. http://dx.doi.org/10.1016/j.jclinepi.2005.01.016.

Egger, M., and G. D. Smith. 1998. Bias in location and selection of studies. *BMJ* 316(7124): 61–66.

Fagan, T. J. 1975. Letter: Nomogram for Bayes theorem. *N Engl J Med* 293(5): 257.

Fisher, N., and P. Switzer. 2001. Graphical assessment of dependence: Is a picture worth a 100 tests? *American Statistician* 55: 233–239.

Harbord, R. M., M. Egger, and J. A. C. Sterne. 2006. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 25(20): 3443–3457. http://dx.doi.org/10.1002/sim.2380.

Higgins, J. P. T., and S. G. Thompson. 2002. Quantifying heterogeneity in a meta-analysis. *Stat Med* 21(11): 1539–1558. http://dx.doi.org/10.1002/sim.1186.

van Houwelingen, H. C., L. R. Arends, and T. Stijnen. 2002. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 21(4): 589–624.

Houwelingen, H. C. V., K. H. Zwinderman, and T. Stijnen. 1993. A bivariate approach to meta-analysis. *Stat Med* 12(24): 2273–2284.

Leeflang, M., J. Deeks, C. Gatsonis, and P. Bossuyt. 2008. Systematic Reviews of Diagnostic Test Accuracy. *Ann Intern Med* 149: 889–897.

Li, J., J. Fine, and N. Safdar. 2007. Prevalence-dependent diagnostic accuracy measures. *Statistics in Medicine* 26: 3258–3273.

Light, L., and D. Pillemer. 1984. *Summing Up: The science of reviewing reaearch*. Cambridge, MA: Harvard University Press.

Macaskill, P., S. D. Walter, and L. Irwig. 2001. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 20(4): 641–654. http://dx.doi.org/10.1002/sim.698.

Peters, J. L., A. J. Sutton, D. R. Jones, K. R. Abrams, and L. Rushton. 2006. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 295(6): 676–680. http://dx.doi.org/10.1001/jama.295.6.676.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2002. Reliable estimation of generalized linear mixed models using adaptive quadrature. *Stata Journal* 2: 1–21.

———. 2004. *GLLAMM Manual*. University of California–Berkeley, Division of Biostatistics, Working Paper Series. Paper No. 160. http://www.bepress.com/ucbbiostat/paper160/.

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. *J Clin Epidemiol* 58(10): 982–990. http://dx.doi.org/10.1016/j.jclinepi.2005.02.022.

Riley, R. D., K. R. Abrams, P. C. Lambert, A. J. Sutton, and J. R. Thompson. 2007a. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med* 26(1): 78–97. http://dx.doi.org/10.1002/sim.2524.

Riley, R. D., K. R. Abrams, A. J. Sutton, P. C. Lambert, and J. R. Thompson. 2007b. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol* 7: 3. http://dx.doi.org/10.1186/1471-2288-7-3.

Riley, R. D., J. R. Thompson, and K. R. Abrams. 2008. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* 9(1): 172–186. http://dx.doi.org/10.1093/biostatistics/kxm023.

Rousseeuw, P., I. Ruts, and J. W. Tukey. 1999. The bagplot, a bivariate boxplot. *American Statistician* 53: 382–387.

Rucker, G., G. Schwarzer, and J. Carpenter. 2008. Arcsine test for publication bias in meta-analyses with binary outcomes. *Stat Med* 27(5): 746–763. http://dx.doi.org/10.1002/sim.2971.

Schwarzer, G., G. Antes, and M. Schumacher. 2002. Inflation of type I error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. *Stat Med* 21(17): 2465–2477. http://dx.doi.org/10.1002/sim.1224.

———. 2007. A test for publication bias in meta-analysis with sparse binary data. *Stat Med* 26(4): 721–733. http://dx.doi.org/10.1002/sim.2588.

Song, F., K. S. Khan, J. Dinnes, and A. J. Sutton. 2002. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *Int J Epidemiol* 31(1): 88–95.

Stengel, D., K. Bauwens, J. Sehouli, A. Ekkernkamp, and F. Porzsolt. 2003. A likelihood ratio approach to meta-analysis of diagnostic studies. *J Med Screen* 10(1): 47–51.

Sutton, A. J., and J. P. T. Higgins. 2008. Recent Developments in Meta-Analysis. *Stat Med* 27: 625–650.

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285–1293.

Whiting, P., R. Harbord, and J. Kleijnen. 2005. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol* 5: 19. http://dx.doi.org/10.1186/1471-2288-5-19.

Whiting, P., A. W. S. Rutjes, J. Dinnes, J. Reitsma, P. M. M. Bossuyt, and J. Kleijnen. 2004. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 8(25): iii, 1–iii234.

Whiting, P., A. W. S. Rutjes, J. B. Reitsma, P. M. M. Bossuyt, and J. Kleijnen. 2003. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 3: 25. http://dx.doi.org/10.1186/1471-2288-3-25.

Whiting, P. F., M. E. Weswood, A. W. S. Rutjes, J. B. Reitsma, P. N. M. Bossuyt, and J. Kleijnen. 2006. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 6: 9. http://dx.doi.org/8-6-9.