A generalized meta-analysis model for binary diagnostic test performance

Ben A. Dwamena, MD

The University of Michigan & VA Medical Centers, Ann Arbor

FNASUG - November 13, 2008

University of Michigan
Medical School

Diagnostic Test Evaluation

DIAGNOSTIC TEST

Any measurement aiming to identify individuals who could potentially benefit from preventative or therapeutic intervention

This includes:

1. Elements of medical history

Diagnostic Test Evaluation

DIAGNOSTIC TEST

Any measurement aiming to identify individuals who could potentially benefit from preventative or therapeutic intervention

This includes:

1. Elements of medical history
2. Physical examination

Diagnostic Test Evaluation

DIAGNOSTIC TEST

Any measurement aiming to identify individuals who could potentially benefit from preventative or therapeutic intervention

This includes:

1. Elements of medical history
2. Physical examination
3. Imaging procedures

Diagnostic Test Evaluation

DIAGNOSTIC TEST

Any measurement aiming to identify individuals who could potentially benefit from preventative or therapeutic intervention

This includes:

1. Elements of medical history
2. Physical examination
3. Imaging procedures
4. Laboratory investigations

Diagnostic Test Evaluation

DIAGNOSTIC TEST

Any measurement aiming to identify individuals who could potentially benefit from preventative or therapeutic intervention

This includes:

1. Elements of medical history
2. Physical examination
3. Imaging procedures
4. Laboratory investigations
5. Clinical prediction rules

Diagnostic Test Evaluation

1. The performance of a diagnostic test assessed by comparison of index and reference test results on a group of subjects

Binary test data often reported as 2×2 matrix

	Reference Test Positive	Reference Test Negative
Test Positive	True Positive	False Positive
Test Negative	False Negative	True Negative

Diagnostic Test Evaluation

1. The performance of a diagnostic test assessed by comparison of index and reference test results on a group of subjects
2. Ideally these should be patients suspected of the target condition that the test is designed to detect.

Binary test data often reported as 2×2 matrix

	Reference Test Positive	Reference Test Negative
Test Positive	True Positive	False Positive
Test Negative	False Negative	True Negative

University of Michigan
Medical School

Measures of Diagnostic Performance

Sensitivity (true positive rate)	The proportion of people with disease who are correctly identified as such by test
Specificity (true negative rate)	The proportion of people without disease who are correctly identified as such by test
Positive predictive value	The proportion of test positive people who truly have disease
Negative predictive value	The proportion of test negative people who truly do not have disease

Measures of Diagnostic Performance

Likelihood ratios (LR)	The ratio of the probability of a positive (or negative) test result in the patients with disease to the probability of the same test result in the patients without the disease
Diagnostic odds ratio	The ratio of the odds of a positive test result in patients with disease compared to the odds of the same test result in patients without disease.
ROC Curve	Plot of all pairs of (1-specificity, sensitivity) as positivity threshold varies

Meta-analysis of Diagnostic Performance

Rationale

1. Evaluation of the quality and scope of available primary studies

Meta-analysis of Diagnostic Performance

Rationale

1. Evaluation of the quality and scope of available primary studies
2. Determination of the proper and efficacious use of diagnostic and screening tests in the clinical setting in order to guide patient treatment

Meta-analysis of Diagnostic Performance

Rationale

1. Evaluation of the quality and scope of available primary studies
2. Determination of the proper and efficacious use of diagnostic and screening tests in the clinical setting in order to guide patient treatment
3. Decision making about health care policy and financing

Meta-analysis of Diagnostic Performance

Rationale

1. Evaluation of the quality and scope of available primary studies
2. Determination of the proper and efficacious use of diagnostic and screening tests in the clinical setting in order to guide patient treatment
3. Decision making about health care policy and financing
4. Identification of areas for further research, development, and evaluation

Meta-analysis of Diagnostic Performance

Major steps

1. Framing objectives of the review

Meta-analysis of Diagnostic Performance

Major steps

1. Framing objectives of the review
2. Identifying the relevant literature

Meta-analysis of Diagnostic Performance

Major steps

1. Framing objectives of the review
2. Identifying the relevant literature
3. Assessment of methodological quality and applicability to the clinical problem at hand

Meta-analysis of Diagnostic Performance

Major steps

1. Framing objectives of the review
2. Identifying the relevant literature
3. Assessment of methodological quality and applicability to the clinical problem at hand
4. Summarizing the evidence qualitatively and if appropriate, quantitatively(meta-analysis)

Meta-analysis of Diagnostic Performance

Major steps

1. Framing objectives of the review
2. Identifying the relevant literature
3. Assessment of methodological quality and applicability to the clinical problem at hand
4. Summarizing the evidence qualitatively and if appropriate, quantitatively(meta-analysis)
5. Interpretation of findings and development of recommendations

Validity of Meta-analysis of Diagnostic Test Accuracy

Depends on presence, extent and sources of variability due to:

1. Methodological quality bias

Validity of Meta-analysis of Diagnostic Test Accuracy

Depends on presence, extent and sources of variability due to:

1. Methodological quality bias
2. Covariate Heterogeneity

Validity of Meta-analysis of Diagnostic Test Accuracy

Depends on presence, extent and sources of variability due to:

1. Methodological quality bias
2. Covariate Heterogeneity
3. Publication and other sample size-related bias

Validity of Meta-analysis of Diagnostic Test Accuracy

Depends on presence, extent and sources of variability due to:

1. Methodological quality bias
2. Covariate Heterogeneity
3. Publication and other sample size-related bias
4. Threshold Effects

Validity of Meta-analysis of Diagnostic Test Accuracy

Depends on presence, extent and sources of variability due to:

1. Methodological quality bias
2. Covariate Heterogeneity
3. Publication and other sample size-related bias
4. Threshold Effects
5. Unobserved heterogeneity

Extent of Heterogeneity

1. Assessed statistically using the quantity I^2 described by Higgins and Colleagues (2002).

Extent of Heterogeneity

1. Assessed statistically using the quantity I^2 described by Higgins and Colleagues (2002).
2. Defined as percentage of total variation across studies attributable to heterogeneity rather than chance.

Extent of Heterogeneity

1. Assessed statistically using the quantity I^2 described by Higgins and Colleagues (2002).
2. Defined as percentage of total variation across studies attributable to heterogeneity rather than chance.
3. I^2 is alculated as:

$$I^2 = ((Q - df)/Q) \times 100. \quad (1)$$

Q is Cochran's heterogeneity statistic; **df** equals degrees of freedom.

Extent of Heterogeneity

1. Assessed statistically using the quantity I^2 described by Higgins and Colleagues (2002).
2. Defined as percentage of total variation across studies attributable to heterogeneity rather than chance.
3. I^2 is alculated as:

$$I^2 = ((Q - df)/Q) \times 100. \quad (1)$$

   Q is Cochran's heterogeneity statistic; **df** equals degrees of freedom.

4. I^2 lies between 0% and 100%: 0% indicates no observed heterogeneity, greater than 50% considered substantial heterogeneity.

Extent of Heterogeneity

1. Assessed statistically using the quantity I^2 described by Higgins and Colleagues (2002).
2. Defined as percentage of total variation across studies attributable to heterogeneity rather than chance.
3. I^2 is alculated as:

$$I^2 = ((Q - df)/Q) \times 100. \quad (1)$$

   Q is Cochran's heterogeneity statistic; **df** equals degrees of freedom.

4. I^2 lies between 0% and 100%: 0% indicates no observed heterogeneity, greater than 50% considered substantial heterogeneity.
5. Advantage of I^2 : does not inherently depend on the number of the studies.

University of Michigan
Medical School

Sources of Heterogeneity: Meta-regression

1. There are different sources of heterogeneity in meta-analysis: characteristics of the study population, variations in the study design (type of design, selection procedures, sources of information, how the information is collected), different statistical methods, and different covariates adjusted for (if relevant)

Sources of Heterogeneity: Meta-regression

1. There are different sources of heterogeneity in meta-analysis: characteristics of the study population, variations in the study design (type of design, selection procedures, sources of information, how the information is collected), different statistical methods, and different covariates adjusted for (if relevant)

2. Formal investigation of sources of heterogeneity is performed by meta-regression, a collection of statistical procedures (weighted/unweighted linear, logistic regression) in which the study effect size is regressed on one or several covariates

Methodological Quality

The assessment of quality has to consider details of study design and execution such as:

1. Cogency of the research question and clinical context

Methodological Quality

The assessment of quality has to consider details of study design and execution such as:

1. Cogency of the research question and clinical context
2. Appropriateness of patient population

Methodological Quality

The assessment of quality has to consider details of study design and execution such as:

1. Cogency of the research question and clinical context
2. Appropriateness of patient population
3. Sufficient description and well-defined interpretation of index diagnostic technique(s)

Methodological Quality

The assessment of quality has to consider details of study design and execution such as:

1. Cogency of the research question and clinical context
2. Appropriateness of patient population
3. Sufficient description and well-defined interpretation of index diagnostic technique(s)
4. Appropriateness and sufficient description of reference standard information

Methodological Quality

The assessment of quality has to consider details of study design and execution such as:

1. Cogency of the research question and clinical context
2. Appropriateness of patient population
3. Sufficient description and well-defined interpretation of index diagnostic technique(s)
4. Appropriateness and sufficient description of reference standard information
5. Other factors that can affect the integrity of the study and the generalizability of the results

Methodological Quality

Methods of quality assessment may focus on:

1. Absence or presence of key qualities in the study report (checklist approach)

Methodological Quality

Methods of quality assessment may focus on:

1. Absence or presence of key qualities in the study report (checklist approach)

2. Scores developed for this purpose (scale approach)

Methodological Quality

Methods of quality assessment may focus on:

1. Absence or presence of key qualities in the study report (checklist approach)
2. Scores developed for this purpose (scale approach)
3. Levels-of-evidence methods by which a level or grade is assigned to studies fulfilling a predefined set of criteria

Threshold effects

1. Most diagnostic tests have multiple or continuous outcomes

Threshold effects

1. Most diagnostic tests have multiple or continuous outcomes
2. Dichotomization or application of cutoff value used to classify results into positive or negative

Threshold effects

1. Most diagnostic tests have multiple or continuous outcomes
2. Dichotomization or application of cutoff value used to classify results into positive or negative
3. **Implicit positivity threshold**: based on interpretation/judgement/machine calibration e.g. radiologists classifying images as normal or abnormal

Threshold effects

1. Most diagnostic tests have multiple or continuous outcomes
2. Dichotomization or application of cutoff value used to classify results into positive or negative
3. **Implicit positivity threshold**: based on interpretation/judgement/machine calibration e.g. radiologists classifying images as normal or abnormal
4. **Explicit positivity threshold**: based on a numerical threshold e.g. blood glucose level above which patient may be said to have diabetes

Threshold effects

1. The chosen threshold may vary between studies of the same test due to inter-laboratory or inter-observer variation

Threshold effects

1. The chosen threshold may vary between studies of the same test due to inter-laboratory or inter-observer variation

2. The higher the cut-off value, the higher the specificity and the lower the sensitivity

Threshold effects

1. The chosen threshold may vary between studies of the same test due to inter-laboratory or inter-observer variation
2. The higher the cut-off value, the higher the specificity and the lower the sensitivity
3. Threshold-based interdependence between sensitivity and specificity tested *a priori* using a rank correlation test such as Spearman's rho after logit transformation

Publication and Other Precision-related Biases

Publication bias Tendency for investigators, reviewers, and editors to submit or accept manuscripts for publication based on the direction or strength of the study findings.

Funnel plot Exploratory tool for investigating publication bias, plotting a measure of effect size versus a measure of study precision

1. Funnel plot should appear symmetric if no bias is present

Publication and Other Precision-related Biases

Publication bias Tendency for investigators, reviewers, and editors to submit or accept manuscripts for publication based on the direction or strength of the study findings.

Funnel plot Exploratory tool for investigating publication bias, plotting a measure of effect size versus a measure of study precision

1. Funnel plot should appear symmetric if no bias is present
2. Assessment of such a plot is very subjective.

Publication and Other Precision-related Biases

Publication bias Tendency for investigators, reviewers, and editors to submit or accept manuscripts for publication based on the direction or strength of the study findings.

Funnel plot Exploratory tool for investigating publication bias, plotting a measure of effect size versus a measure of study precision

1. Funnel plot should appear symmetric if no bias is present
2. Assessment of such a plot is very subjective.
3. Non-parametric and linear regression methods used to formally test funnel plot asymmetry.

Examples of Tests For Funnel Plot Asymmetry

(Begg 1994)	Rank correlation between standardized effect and its standard error
(Egger 1997)	Linear regression of intervention effect against its standard error weighted by inverse of the variance of intervention effect estimate
(Macaskill 2001)	Linear regression of intervention effect on sample size
(Harbord 2006)	Modified vesion of (Egger 1997) based on "score" and "score variance" of the log odds ratio
(Peters 2006)	Linear regression of intervention effect on inverse of sample size

Problems with sample size and standard error

1. The asymptotic standard error is a biased estimate of the true standard error, with larger bias for smaller cell sizes, as occurs with larger DORs and smaller studies

Problems with sample size and standard error

1. The asymptotic standard error is a biased estimate of the true standard error, with larger bias for smaller cell sizes, as occurs with larger DORs and smaller studies

2. Diagnostic studies have unequal sample sizes in diseased and non-diseased groups which reduces the precision of an estimate of test accuracy for a given sample size

Problems with sample size and standard error

1. The asymptotic standard error is a biased estimate of the true standard error, with larger bias for smaller cell sizes, as occurs with larger DORs and smaller studies

2. Diagnostic studies have unequal sample sizes in diseased and non-diseased groups which reduces the precision of an estimate of test accuracy for a given sample size

3. The standard error of the logDOR depends on proportion testing positive. However, individual studies often differ in positivity threshold leading to variability in proportion testing postive

Summary ROC Meta-analysis of Diagnostic Test Accuracy

The most commonly used and easy to implement method

1. Linear regression analysis of the relationship
   $D = \mathbf{a} + \mathbf{b}S$ where :
   D = (logit TPR) - (logit FPR) = ln DOR
   S = (logit TPR) + (logit FPR) = proxy for the threshold

Moses, Shapiro and Littenberg. Med Decis Making (1993)12:1293-1316

University of Michigan
Medical School

Summary ROC Meta-analysis of Diagnostic Test Accuracy

The most commonly used and easy to implement method

1. Linear regression analysis of the relationship
   $D = \mathbf{a} + \mathbf{b}S$ where :
   D = (logit TPR) - (logit FPR) = ln DOR
   S = (logit TPR) + (logit FPR) = proxy for the threshold
2. **a** and **b** may be estimated by weighted or unweighted least squares or robust regression, back-transformed and plotted in ROC space

Moses, Shapiro and Littenberg. Med Decis Making (1993)12:1293-1316

University of Michigan
Medical School

Summary ROC Meta-analysis of Diagnostic Test Accuracy

The most commonly used and easy to implement method

1. Linear regression analysis of the relationship
   $D = \mathbf{a} + \mathbf{b}S$ where :
   D = (logit TPR) - (logit FPR) = ln DOR
   S = (logit TPR) + (logit FPR) = proxy for the threshold
2. **a** and **b** may be estimated by weighted or unweighted least squares or robust regression, back-transformed and plotted in ROC space
3. Differences between tests or subgroups may examined by adding covariates to model

Moses, Shapiro and Littenberg. Med Decis Making (1993)12:1293-1316

University of Michigan
Medical School

Summary ROC Meta-analysis of Diagnostic Test Accuracy

1. Assumes variability in test performance due only to threshold effect and within-study variability

Summary ROC Meta-analysis of Diagnostic Test Accuracy

1. Assumes variability in test performance due only to threshold effect and within-study variability
2. Does not provide average estimates of sensitivity and specificity

Summary ROC Meta-analysis of Diagnostic Test Accuracy

1. Assumes variability in test performance due only to threshold effect and within-study variability
2. Does not provide average estimates of sensitivity and specificity
3. Continuity correction may introduce non-negligible downward bias to the estimated SROC curve

Summary ROC Meta-analysis of Diagnostic Test Accuracy

1. Assumes variability in test performance due only to threshold effect and within-study variability
2. Does not provide average estimates of sensitivity and specificity
3. Continuity correction may introduce non-negligible downward bias to the estimated SROC curve
4. Does not account for measurement error in S

Summary ROC Meta-analysis of Diagnostic Test Accuracy

1. Assumes variability in test performance due only to threshold effect and within-study variability
2. Does not provide average estimates of sensitivity and specificity
3. Continuity correction may introduce non-negligible downward bias to the estimated SROC curve
4. Does not account for measurement error in S
5. Ignores potential correlation between D and S

Summary ROC Meta-analysis of Diagnostic Test Accuracy

1. Assumes variability in test performance due only to threshold effect and within-study variability
2. Does not provide average estimates of sensitivity and specificity
3. Continuity correction may introduce non-negligible downward bias to the estimated SROC curve
4. Does not account for measurement error in S
5. Ignores potential correlation between D and S
6. Confidence intervals and p-values are likely to be inaccurate

Recent Developments

Publication Bias test for Diagnostic Meta-analysis

1. linear regression of log odds ratio on inverse square root of effective sample size

Bivariate Mixed Effects Models

Arends et al. Med Decis Making. Published online June 30, 2008

Recent Developments

Publication Bias test for Diagnostic Meta-analysis

1. linear regression of log odds ratio on inverse square root of effective sample size
2. Uses the effective sample size as weight

Bivariate Mixed Effects Models

Arends et al. Med Decis Making. Published online June 30, 2008

Recent Developments

Publication Bias test for Diagnostic Meta-analysis

1. linear regression of log odds ratio on inverse square root of effective sample size
2. Uses the effective sample size as weight
3. Effective sample size=4*(ndis*nndis)/sample size

Bivariate Mixed Effects Models

Arends et al. Med Decis Making. Published online June 30, 2008

Recent Developments

Publication Bias test for Diagnostic Meta-analysis

1. linear regression of log odds ratio on inverse square root of effective sample size
2. Uses the effective sample size as weight
3. Effective sample size=4*(ndis*nndis)/sample size

Bivariate Mixed Effects Models

1. Focused on inferences about sensitivity and specificity but SROC curve(s) can be derived from the model parameters

Arends et al. Med Decis Making. Published online June 30, 2008

University of Michigan
Medical School

Recent Developments

Publication Bias test for Diagnostic Meta-analysis

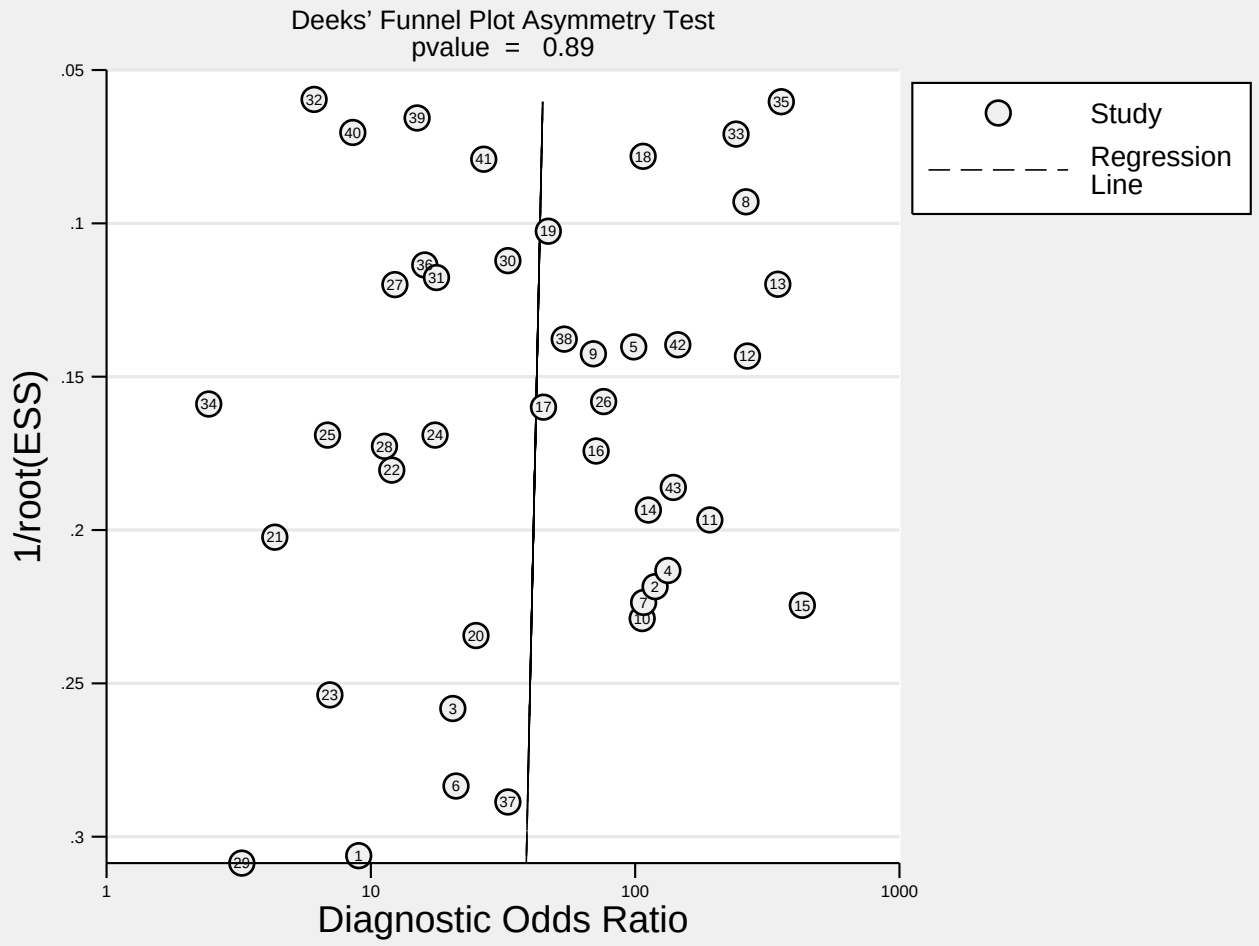
1. linear regression of log odds ratio on inverse square root of effective sample size
2. Uses the effective sample size as weight
3. Effective sample size=4*(ndis*nndis)/sample size

Bivariate Mixed Effects Models

1. Focused on inferences about sensitivity and specificity but SROC curve(s) can be derived from the model parameters
2. Generalization of the commonly used DerSimonian and Laird random effects model

Arends et al. Med Decis Making. Published online June 30, 2008

Publication Bias test for Diagnostic Meta-analysis



University of Michigan
Medical School

Bivariate Linear Mixed Model

Level 1: Within-study variability

$$\begin{pmatrix} \texttt{logit}(p_{Ai}) \\ \texttt{logit}(p_{Bi}) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix}, C_i \right)$$

$$C_i = \begin{pmatrix} s_{Ai}^2 & 0 \\ 0 & s_{Bi}^2 \end{pmatrix}$$

p_{Ai} and p_{Bi}	Sensitivity and specificity of the ith study
μ_{Ai} and μ_{Bi}	Logit-transforms of sensitivity and specificity of the ith study
C_i	Within-study variance matrix
s_{Ai}^2 and s_{Bi}^2	variances of logit-transforms of sensitivity and specificity

Reitsma JB et al. J. Clin Epidemiol (2005) 58:982-990

University of Michigan Medical School

Bivariate Linear Mixed Model

Level 2: Between-study variability

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N\left(\begin{pmatrix} M_A \\ M_B \end{pmatrix}, \Sigma_{AB}\right)$$

$$\Sigma_{AB} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

μ_{Ai} and μ_{Bi} Logit-transforms of sensitivity and specificity of the ith study

M_A and M_B Means of the normally distributed logit-transforms

Σ_{AB} Between-study variances and covariance matrix

Reitsma JB et al. J. Clin Epidemiol (2005) 58:982-990

Bivariate Binomial Mixed Model

Level 1: Within-study variability

$$y_{Ai} \sim Bin\left(n_{Ai}, p_{Ai}\right)$$

$$y_{Bi} \sim Bin\left(n_{Bi}, p_{Bi}\right)$$

n_{Ai} and n_{Bi} Number of diseased and non-diseased

y_{Ai} and y_{Bi} Number of diseased and non-diseased with true test results

p_{Ai} and p_{Bi} Sensitivity and specificity of the ith study

Chu H, Cole SR (2006) J. Clin Epidemiol 59:1331-1332

Bivariate Binomial Mixed Model

Level 2: Between-study variability

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N\left(\begin{pmatrix} M_A \\ M_B \end{pmatrix}, \Sigma_{AB}\right)$$

$$\Sigma_{AB} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

μ_{Ai} and μ_{Bi} Logit-transforms of sensitivity and specificity of the ith study

M_A and M_B Means of the normally distributed logit-transforms

Σ_{AB} Between-study variances and covariance matrix

Chu H, Cole SR (2006) J. Clin Epidemiol 59:1331-1332

Bivariate Mixed Models

1. Exact binomial approach preferred especially for small sample data and for avoiding continuity correction

Bivariate Mixed Models

1. Exact binomial approach preferred especially for small sample data and for avoiding continuity correction

2. The relation between logit-transformed sensitivity and specificity is given by $\mu_{Ai} = a + b \times \mu_{Bi}$ with slope $b = \sigma_{AB}/\sigma_A^2$ and intercept $a = M_A$ - $b \times M_B$

Bivariate Mixed Models

1. Exact binomial approach preferred especially for small sample data and for avoiding continuity correction

2. The relation between logit-transformed sensitivity and specificity is given by $\mu_{Ai} = a + b \times \mu_{Bi}$ with slope $b = \sigma_{AB}/\sigma_A^2$ and intercept $a = M_A$ - $b \times M_B$

3. SROC may be obtained after anti-logit transformation of the regression line

Methodological Framework

Propose a generalized framework for diagnostic meta-analysis based on a modification of the bivariate Dale model:

1. Univariate random-effects logistic models for sensitivity and specificity are associated through a log-linear model of odds ratios with effective sample size as independent variable

Methodological Framework

Propose a generalized framework for diagnostic meta-analysis based on a modification of the bivariate Dale model:

1. Univariate random-effects logistic models for sensitivity and specificity are associated through a log-linear model of odds ratios with effective sample size as independent variable

2. This unifies the estimation of summary test performance and assessment of the presence, extent, and sources of variability

Methodological Framework

Discuss specification, estimation, diagnostics, and prediction of model:

1. Using a motivating dataset of 43 studies investigating FDG-PET for staging the axilla in patients with newly diagnosed breast cancer

Methodological Framework

Discuss specification, estimation, diagnostics, and prediction of model:

1. Using a motivating dataset of 43 studies investigating FDG-PET for staging the axilla in patients with newly diagnosed breast cancer

2. Taking advantage of the ability of **gllamm** to model a mixture of discrete and continous outcomes

Bivariate Dale Model (Correlated Binary Responses)

1. Joint probabilities decomposed into two marginal distributions for the main effects
2. One log-cross-ratio for the association between two responses

```
h1{p1+(x)}=B1x;
h2{p+1(x)}=B2x;
h{(p11(x)*p22(x))/(p12(x)*p21(x))}=B3x
```

1. h1, h2, h3 are link functions in the GLM terminology
2. p1+ and p+1 are the marginal probabilities for response1=1 and response2=1 respectively
3. Most popular choice for h1=h2 is the **logit** function
4. Commonly used link function for h3 is the **natural logarithm**:

```
ln(cross-ratio)=ln{(p11(x)*p22(x))/(p12(x)*p21(x))}
```

University of Michigan
Medical School

Modified Bivariate Dale Model

Within-study variability

$$y_{Ai} \sim Bin\left(n_{Ai}, p_{Ai}\right)$$

$$y_{Bi} \sim Bin\left(n_{Bi}, p_{Bi}\right)$$

n_{Ai} and n_{Bi} Number of diseased and non-diseased

y_{Ai} and y_{Bi} Number of diseased and non-diseased with true test results

p_{Ai} and p_{Bi} Sensitivity and specificity of the ith study

Modified Bivariate Dale Model

Between-study variability

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N\left(\begin{pmatrix} M_A \\ M_B \end{pmatrix}, \Sigma_{AB} \right)$$

$$\Sigma_{AB} = \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{pmatrix}$$

μ_{Ai} and μ_{Bi} Logit-transforms of sensitivity and specificity of the ith study

M_A and M_B Means of the normally distributed logit-transforms

Σ_{AB} Between-study variances

University of Michigan
Medical School

Modified Bivariate Dale Model

Association Model

Associates the univariate random-effects logistic models for sensitivity and specificity in the form a log-linear model:

$logDOR_i = a + b \times \text{ESS}_i$

intercept a = adjusted odds ratio

and slope b = bias coefficient

Example: PET for axillary staging of breast Cancer

1. PET or Positron Emission Tomography uses radiolabeled glucose analog to evaluate tumor metabolism

Example: PET for axillary staging of breast Cancer

1. PET or Positron Emission Tomography uses radiolabeled glucose analog to evaluate tumor metabolism

2. This radiological test may be used to stage and/or examine the extent of breast cancer

Example: PET for axillary staging of breast Cancer

1. PET or Positron Emission Tomography uses radiolabeled glucose analog to evaluate tumor metabolism
2. This radiological test may be used to stage and/or examine the extent of breast cancer
3. The accuracy of axillary PET has been studied by many researchers

Example: PET for axillary staging of breast Cancer

1. PET or Positron Emission Tomography uses radiolabeled glucose analog to evaluate tumor metabolism
2. This radiological test may be used to stage and/or examine the extent of breast cancer
3. The accuracy of axillary PET has been studied by many researchers
4. We obtained, by searching PUBMED, 43 studies published between 1990 and 2008

Example: PET for axillary staging of breast Cancer

Table: Dataset

Idnum	Author	Year	TP	FP	FN	TN	SIZE
1	Tse	1992	4	0	3	3	10
2	Adler1	1993	8	0	1	10	18
3	Hoh	1993	6	0	3	5	14
4	Crowe	1994	9	0	1	10	20
5	Avril	1996	19	1	5	26	51
6	Bassa	1996	10	0	3	3	16
7	Scheidhauer	1996	9	1	0	8	18
8	Utech	1996	44	20	0	60	124
9	Adler2	1997	19	11	0	20	50
10	Palmedo	1997	5	0	1	14	20
11	Noh	1998	12	0	1	11	24
12	Smith	1998	19	1	2	28	50
13	Rostom	1999	42	0	6	26	74
14	Yutani1	1999	8	0	2	16	26
15	Hubner	2000	6	0	0	16	22
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
32	Wahl	2004	66	40	43	159	308
33	Zornoza	2004	90	2	17	91	200
34	Weir	2005	5	3	13	19	40
35	Gil-Rendo	2006	120	2	22	131	275
36	Kumar	2006	16	2	20	40	80
37	Stadnik	2006	4	0	1	5	10
38	Chung	2006	25	0	17	18	51
39	Veronesi	2006	38	5	65	128	236
40	Cermik	2008	40	15	39	125	
41	Ueda	2008	34	6	25	118	
42	Fuster	2008	14	0	6	32	
43	Heuser	2008	8	0	2	20	

University of Michigan
Medical School

Recode Data for gllamm

```
gen dor = (tp*tn)/(fp*fn)
gen ldor = ln(dor)
gen ldorvar = (1/fn)+(1/tn)+(1/fp)+(1/tp)
gen ldorse = sqrt((1/fn)+(1/tn)+(1/fp)+(1/tp))
tempvar  n1 n2 ESS zero thetai sethetai
gen ‘n1’ = tp + fn
gen ‘n2 ’= tn + fp
gen ‘ESS’ =(4 * ‘n1’ * ‘n2’)/(‘n1’ + ‘n2’)
gen ‘thetai’=(tp * tn)/(fp * fn)
replace ‘thetai’=log(‘thetai’)
gen ‘sethetai’=sqrt(‘ESS’)
gen size =1/‘sethetai’
```

Recode Data for gllamm

```
gen ttruth1 = tn   /* number truly disease-free */
gen ttruth2 = tp   /* number truly diseased */
gen ttruth3 = `thetai'
gen num1 = tn+fp   /* total disease-free */
gen num2 = tp+fn   /* total diseased */
gen num3 =  1
reshape long num ttruth, i(study) j(dtruth) string
qui tabulate dtruth, generate(disgrp)
eq disgrp1: disgrp1
eq disgrp2: disgrp2
eq disgrp3: disgrp3
gen gvar = .
replace gvar = 1 if dtruth == "1"
replace gvar = 2 if dtruth == "2"
replace gvar = 3 if dtruth == "3"
 forvalues i=1/3 {
     g size_`i' = disgrp`i'* size
   }
}
```

University of Michigan
Medical School

Bivariate Binomial Mixed Model

```
gllamm ttruth disgrp1 disgrp2 if dtruth !="3", nocons ///
i(study) nrf(2) eqs(disgrp1 disgrp2)  ///
f(bin) l(logit) denom(num) ip(m) adapt
```

Table: Estimation results

Variable	**Coefficient**	**(Std. Err.)**
Fixed Effects		
logitsen	3.084	(0.260)
logitspe	0.925	(0.197)
Random-Effects		
logitsen	1.144	(0.232)
logitspe	1.109	(0.174)
Correlation	-0.319	(0.256)

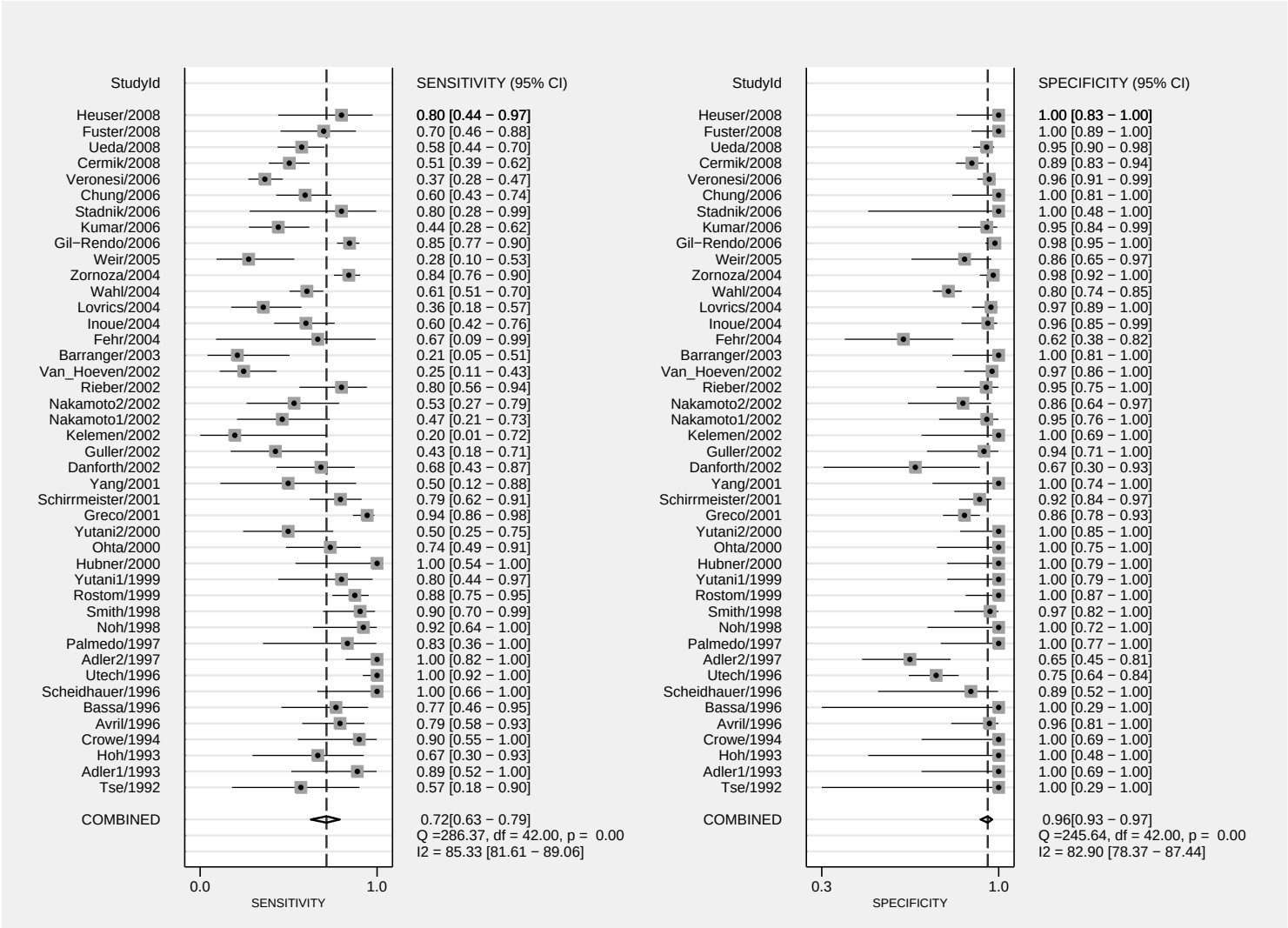
University of Michigan
Medical School

Bivariate Binomial Mixed Model

Table: Summary estimates

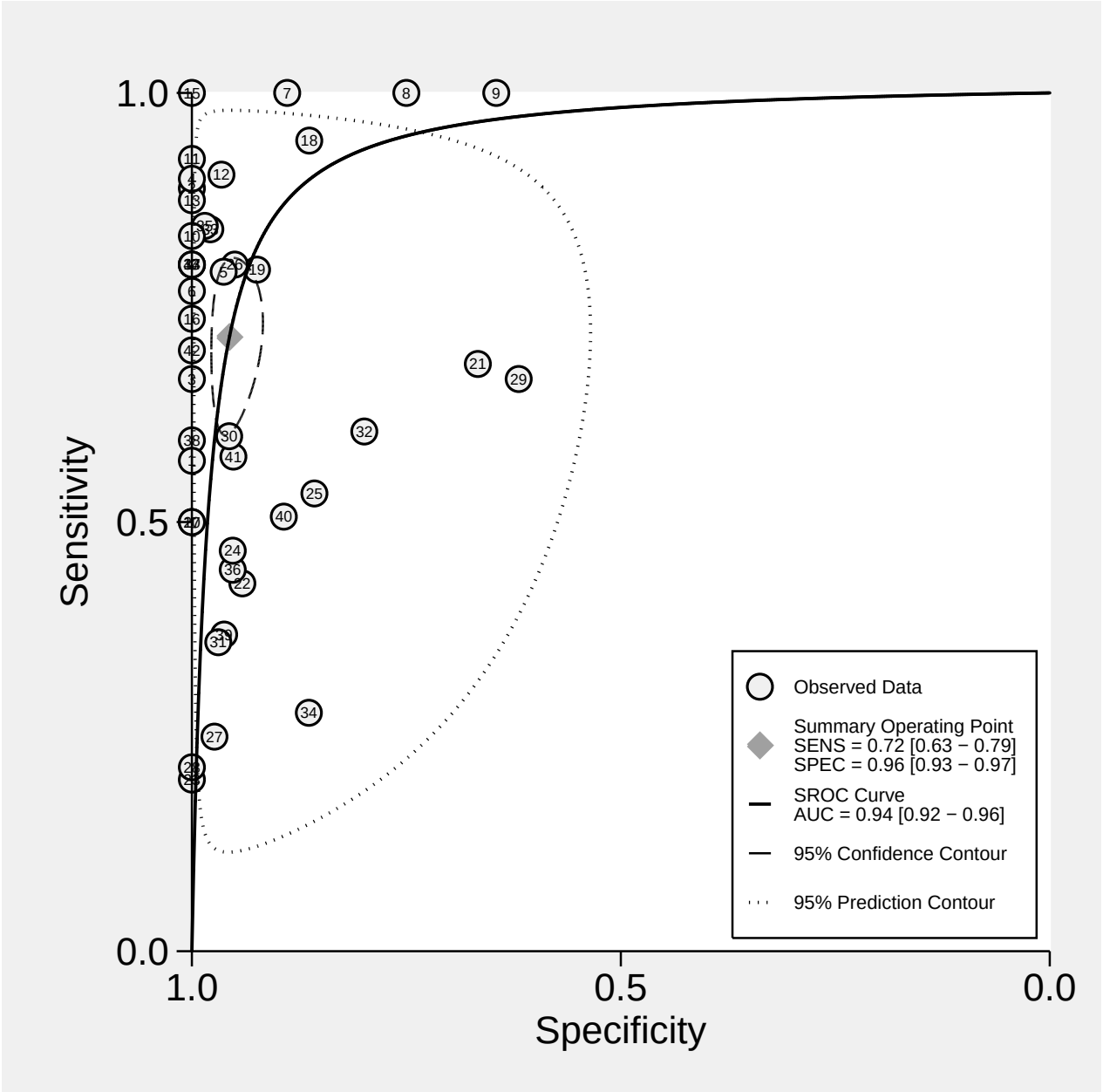
Variable	**Coefficient**	**(Std. Err.)**
sens	0.716	(0.040)
spec	0.956	(0.011)
ldor	4.009	(0.305)
lrp	16.362	(4.047)
lrn	0.297	(0.042)

Forest Plot



University of Michigan
Medical School

SROC Curve



University of Michigan
Medical School

No bias Uncorrelated Random-Effects

```
gllamm ttruth disgrp1 disgrp2 disgrp3, nocons nocor ///
i(study) nrf(2) eqs(disgrp1 disgrp2) f(bin bin gauss) ///
l(logit logit id) denom(num) ip(m) adapt fv(gvar) lv(gvar)
```

Table: Estimation results

Variable	**Coefficient**	**(Std. Err.)**
Fixed effects		
logitsen	3.119	(0.265)
logitspe	0.921	(0.193)
logdor	3.694	(0.211)
Random effects		
logitsen	1.196	(0.246)
logitspe	1.143	(0.173)

University of Michigan
Medical School

No bias Uncorrelated Random-Effects

Table: Summary estimates

Variable	**Coefficient**	**(Std. Err.)**
sens	0.715	(0.039)
spec	0.958	(0.011)
ldor	3.694	(0.211)
lrp	16.888	(4.384)
lrn	0.297	(0.041)

Bias Correlated Random-Effects

```
gllamm ttruth disgrp1 disgrp2 disgrp3 size_3, nocons ///
i(study) nrf(2) eqs(disgrp1 disgrp2) f(bin bin gauss) ///
l(logit logit id) denom(num) ip(m) adapt fv(gvar) lv(gvar)
```

Table: Estimation results

Variable	**Coefficient**	**(Std. Err.)**
Fixed Effects		
logitsen	3.084	(0.260)
logitspe	0.925	(0.197)
logdor	4.324	(0.543)
bias	-3.801	(3.032)
Random-effects		
logitsens	1.144	(0.232)
logitspe	1.109	(0.174)
Correlation	-0.319	(0.256)

Bias Correlated Random-Effects

Table: Summary estimates

Variable	**Coefficient**	**(Std. Err.)**
sens	0.716	(0.040)
spec	0.956	(0.011)
ldor	4.324	(0.543)
lrp	16.362	(4.046)
lrn	0.297	(0.042)

Bias Uncorrelated Random-Effects

```
gllamm ttruth disgrp1 disgrp2 disgrp3 size_3, nocons nocor ///
i(study) nrf(2) eqs(disgrp1 disgrp2) f(bin bin gauss) ///
l(logit logit id) denom(num) ip(m) adapt fv(gvar) lv(gvar)
```

Table: Estimation results

Variable	**Coefficient**	**(Std. Err.)**
Fixed effects		
logitsen	3.119	(0.265)
logitspe	0.921	(0.193)
logdor	4.324	(0.543)
bias	-3.801	(3.032)
Random effects		
logitsen	1.196	(0.246)
logitspe	1.144	(0.173)

Bias Uncorrelated Random-Effects

Table: Summary estimates

Variable	**Coefficient**	**(Std. Err.)**
sens	0.715	(0.039)
spec	0.958	(0.011)
ldor	4.324	(0.543)
lrp	16.888	(4.384)
lrn	0.297	(0.041)

Comparative Results

Table: **Fit and Complexity Measures**

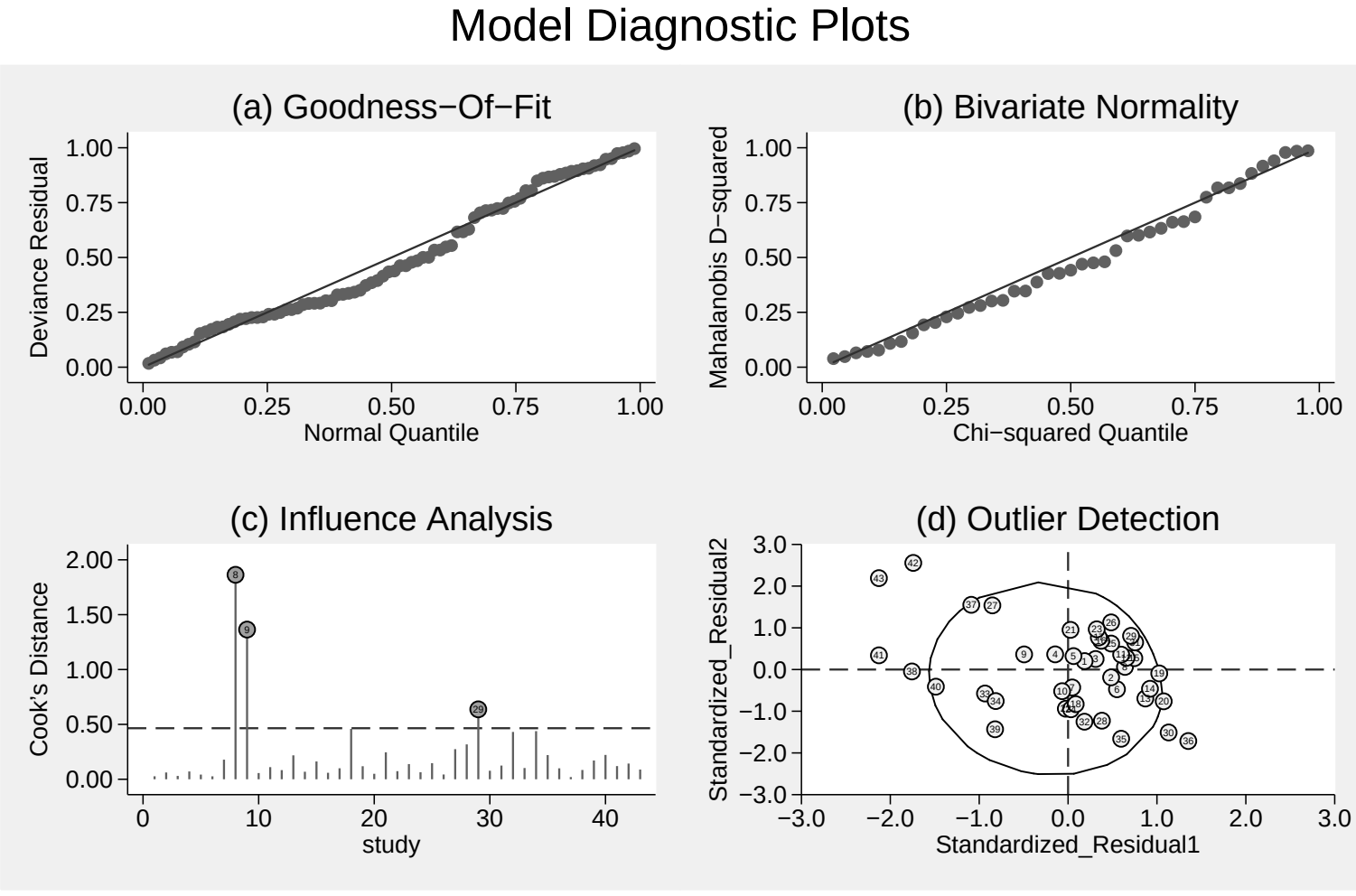
Model	**nparm**	**Deviance**	**BIC**
No Bias	7	548.42	582.44
Bias Correlated Random-effects	8	548.42	587.30
Bias Uncorrelated Random-effects	7	548.37	582.39

Table: **Sensitivity and Specificity**

Model	**Sens**	**Spec**
No Bias	0.716 (0.638 - 0.795)	0.956 (0.935 - 0.978)
Bias Correlated RE	0.716 (0.638 - 0.795)	0.956 (0.935 - 0.978)
Bias Uncorrelated RE	0.715 (0.638 - 0.792)	0.958 (0.937 - 0.979)

Prediction and Diagnostics

May use **gllapred** for empirical bayes predictions, residual analysis, influence analysis, normality testing etc



University of Michigan
Medical School

Conclusions

1. The preferred model is the **Bias Uncorrelated Random-effects Model**

Conclusions

1. The preferred model is the **Bias Uncorrelated Random-effects Model**
2. If interest is in diagnostic performance only, then the **Bivariate binomial mixed** and **modified bivariate Dale** models are equivalent.

Conclusions

1. The preferred model is the **Bias Uncorrelated Random-effects Model**
2. If interest is in diagnostic performance only, then the **Bivariate binomial mixed** and **modified bivariate Dale** models are equivalent.
3. The **modified bivariate Dale** models may be extended further to include study-level covariates to assess impact on summary test performance jointly or separately.

University of Michigan
Medical School