Random Effects Meta-analysis of Studies Reporting Pairs of Sensitivity and Specificity: a Comparison of Methods

Taye H. Hamza\textsuperscript{1}, Johannes B. Reitsma\textsuperscript{2}, Theo Stijnen\textsuperscript{1}

\textsuperscript{1}Department of Epidemiology and Biostatistics, Erasmus MC - Erasmus University Medical Center, P.O.Box 1738, 3000 DR Rotterdam, The Netherlands

\textsuperscript{2}Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, P.O.Box 22700, 1100 DE Amsterdam, The Netherlands

International Conference on Sources of Heterogeneity in Meta-Analysis of Randomized Clinical Trials. Bremen, Germany, February 9, 2007
Outline

- Introduction
- Methods Binomial
- Simulation Study
- Data Example
- Concluding Remarks
An ideal diagnostic test discriminates between diseased and non-diseased individuals without error.

True disease status for each individual is established using a reference test.

Available data:
- Two commonly reported measures:
  - Sensitivity and Specificity or
  - Two by two table
- Other reported measures:
  - Individual subject data
  - Area under the ROC curve
  - Diagnostic likelihood ratios etc.
Data considered

- Studies reporting pairs of sensitivity and specificity or a two by two table

Objective

- Compare three random effects meta-analysis methods
  - Univariate (U)
  - Approximate bivariate (AB)
  - Exact bivariate (EB)

using a simulation study
Assume N studies report a pair of sensitivity and specificity or a two by two table:

<table>
<thead>
<tr>
<th>Test Results</th>
<th>&quot;True&quot; disease status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Diseased</td>
</tr>
<tr>
<td>Positive</td>
<td>TP</td>
</tr>
<tr>
<td>Negative</td>
<td>FN</td>
</tr>
<tr>
<td>n_i</td>
<td>n_0</td>
</tr>
</tbody>
</table>

Observed true positive rate (TPR) and false positive rate (FPR):

- $TPR_i = Sensitivity = \frac{TP_i}{n_{1i}}$
- $FPR_i = 1 - Specificity = \frac{FP_i}{n_{0i}}$
Summarizing Sensitivity and Specificity

\[
TPR = \frac{TP}{N_{\text{Diseased}}}
\]

\[
FPR = \frac{FP}{N_{\text{Non-Diseased}}}
\]
Summarizing Sensitivity and Specificity

$$TPR = \frac{TP}{N_{Diseased}}$$

$$FPR = \frac{FP}{N_{Non-Diseased}}$$
Averaging Sensitivity and Specificity could be misleading

Often sensitivity and specificity are negatively correlated

The trade off between sensitivity and specificity can be explained by a summary ROC curve

How can we Model Sensitivity and Specificity jointly?
The Standard Summary ROC curve


\[ D_i = \alpha + \beta \times S_i + \text{error} \]

with \( D_i \) (logarithm of diagnostic odds ratio) = \( \logit(TPR_i) - \logit(FPR_i) \) and \( S_i \) (measure of test positivity (threshold)) = \( \logit(TPR_i) + \logit(FPR_i) \)

- study level covariates can be added
- if \( \beta = 0 \) accuracy is independent of \( S \) resulting in symmetric SROC
- if \( \beta \neq 0 \), test accuracy depends on \( S \) resulting in asymmetric SROC
- if \( |\beta| < 1 \), sensitivity and specificity are negatively correlated
- Transforming the regression line into ROC space (\( \rightarrow \) ROC curve)
The Standard Summary ROC curve (Cont.)

- **SROC**
  - Most commonly used method
  - Easy to implement
  - Can be carried out in many statistical packages

- **Drawbacks**
  - Does not account for the between studies variability (heterogeneity)
  - Does not account the within study correlation between $D$ and $S$
  - $S$ is assumed to be error free (introduce bias in estimate of $\beta$)
  - Addition of 0.5 when there is a zero count (introduces bias)
  - Estimation technique: Weighted / Unweighted?
Univariate Random Effects Approach

Univariate Random Effects

\[ \hat{D}_i = \alpha_i + \beta \hat{S}_i + \epsilon_i \]  

with \( \alpha_i \sim N(\alpha, \sigma^2_\alpha) \), \( \epsilon_i \sim N(0, \hat{\sigma}^2_{Di}) \) and \( \hat{\sigma}^2_{Di} = \frac{1}{TP_i} + \frac{1}{FN_i} + \frac{1}{FP_i} + \frac{1}{TN_i} \)

- Accounts for the heterogeneity across studies
- Does not account for the within study correlation between \( D \) and \( S \)
- \( S \) is assumed to be error free (introduces bias in the estimate of \( \beta \))
- Addition of 0.5 when there is a zero count (introduces bias)
The Between Study Model (Reitsma et al, 2005; Arends et al, 2007)

\[
\begin{pmatrix}
\xi_i \\
\eta_i
\end{pmatrix}
\sim N
\begin{pmatrix}
\begin{pmatrix}
\bar{\xi} \\
\bar{\eta}
\end{pmatrix},
\begin{pmatrix}
\sigma^2_{\xi} & \sigma_{\xi \eta} \\
\sigma_{\xi \eta} & \sigma^2_{\eta}
\end{pmatrix}
\end{pmatrix}
\]

(2)

with \(\xi_i = true \; \text{logit}(FPR_i)\) and \(\eta_i = true \; \text{logit}(TPR_i)\)

- Accounts for the heterogeneity across studies
- The measurement error in \(S\) is accounted for
- (2) can be characterized by different SROC curves (Arends et al, 2007)
Bivariate Random Effects Approach (Cont.)

- For example:
  - Regression of $\eta$ on $\xi$
    \[
    \eta = \alpha + \beta \xi
    \]
  - $\alpha$, $\beta$ and $\sigma^2_{\eta|\xi}$ can be derived from (2)
    - $\alpha = \bar{\eta} - \frac{\sigma_{\xi \eta}}{\sigma^2_\xi} \bar{\xi}$
    - $\beta = \frac{\sigma_{\xi \eta}}{\sigma^2_\xi}$
    - $\sigma^2_{\eta|\xi} = \sigma^2_\eta - \frac{\sigma^2_{\xi \eta}}{\sigma^2_\xi}$
Bivariate Random Effects Approach (Cont.)

For example:

Regression of $D$ on $S$

\[ D = \alpha + \beta S \]

$\alpha$, $\beta$ and $\sigma^2_{D|S}$ can be derived from (2)

\[ \alpha = \bar{D} - \frac{\sigma^2_\eta - \sigma^2_\xi}{\sigma^2_\xi + \sigma^2_\eta + 2\sigma_\xi \eta} \bar{S} \quad \text{with} \quad \bar{D} = \bar{\eta} - \bar{\xi} \quad \text{and} \quad \bar{S} = \bar{\eta} + \bar{\xi} \]

\[ \beta = \frac{\sigma^2_\eta - \sigma^2_\xi}{\sigma^2_\xi + \sigma^2_\eta + 2\sigma_\xi \eta} \]

\[ \sigma^2_{D|S} = (\sigma^2_\xi + \sigma^2_\eta - 2\sigma_\xi \eta) - \frac{(\sigma^2_\eta - \sigma^2_\xi)^2}{\sigma^2_\xi + \sigma^2_\eta + 2\sigma_\xi \eta} \]
Bivariate Random Effects Approach (Cont.)

- Other choices (Arends et al; 2007)
  - Regression of $\xi$ on $\eta$
  - Rutter and Gatsonis
  - The major axis

- We considered the regression of $D$ on $S$
  - Direct extension of the traditional SROC method
Modeling the Within Study Variability

- The within study variability can be assumed to follow a
  - Normal distribution ('Approximate Method')
  - Binomial Distribution ('Exact Method')

- Normal Distribution ('Approximate Method')

\[
\hat{\xi}_i \equiv N(\xi_i, \frac{1}{x_{0i} + \frac{1}{n_{0i} - x_{0i}}}) \quad \text{and} \quad \hat{\eta}_i \equiv N(\eta_i, \frac{1}{x_{1i} + \frac{1}{n_{1i} - x_{1i}}})
\]

- Addition of 0.5 when there is a zero count (introduces bias)
Binomial Distribution ('Exact Method')

\[ x_{1i} \sim \text{binomial}\left( \frac{e^{\eta_i}}{1 + e^{\eta_i}}, n_{1i} \right) \]

and

\[ x_{0i} \sim \text{binomial}\left( \frac{e^{\xi_i}}{1 + e^{\xi_i}}, n_{0i} \right) \]

- no need of adding a correction factor (0.5)
The three Random Effects Methods

Table: Summarizing the drawbacks of the three methods (U, AB, EB)

<table>
<thead>
<tr>
<th></th>
<th>U</th>
<th>AB</th>
<th>EB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account the between heterogeneity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Account the within study correlation between D and S</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Account the measurement error in S</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>No need of a correction for zero denominators</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
</tbody>
</table>
Generate $\xi_i$ and $\eta_i$ from a bivariate normal distribution

Calculate $\pi_{0i}$ and $\pi_{1i}$ from $\xi_i$ and $\eta_i$

$\pi_{0i} = \frac{e^{\xi_i}}{1 + e^{\xi_i}}$ and $\pi_{1i} = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$

Generate $TP_i$ and $FP_i$ from a binomial distribution

Different combination of the following true values

$(\bar{\xi}, \bar{\eta}) = (-2.20, 2.94), (-0.85, 0.62), (-0.85, 2.94)$

$(\sigma_\xi^2, \sigma_\eta^2) = (0.5, 0.5), (1.2, 1.2), (0.5, 1.2)$

$\rho_{\xi\eta} = 0.2, 0.5, 0.9$

$n = 40, 500$

$N = 25, 100$

The true values are motivated by existing meta-analysis
Simulation Study Setting

- For each scenarios 1000 simulation was carried out
- The performance of the methods compared
  - Bias
  - MSE
  - Confidence interval coverage probability
The Bias and Coverage probability do not vary largely with the

- between study variances ($\sigma^2_\xi$ and $\sigma^2_\eta$)
- the correlation between $\xi$ and $\eta$ ($\rho_{\xi\eta}$)

The Bias and Coverage probability largely depends on the

- median within study sample size ($n$)
- median sensitivity and
- median specificity
Regardless of the scenario considered the EB

- estimates $\bar{D}$ and $\beta$ with
  - negligible amount of bias
  - reasonably acceptable coverage probability
- estimates $\sigma^2_{\eta/\xi}$ with small bias compared to the U and AB
- usually gives
  - smallest MSE for $\bar{D}$ and $\sigma^2_{\eta/\xi}$
  - for $\beta$, smaller MSE than the AB
Simulation Results (Cont.)

- When $n$ is small and $\bar{\xi}$ or $\bar{\eta}$ are large the U and AB
  - estimate $\bar{D}$, $\beta$ and $\sigma^2_{\bar{\eta}/\xi}$ with
    - large amount of bias
    - low coverage probability (often far from the nominal value)

- Often the U approach gives the smallest MSE for $\beta$

- When $n$ is large and $\bar{\xi}$ and $\bar{\eta}$ are small the U and AB give comparable results with the EB in terms of bias and coverage probability
Simulation Results when $\rho \xi \eta = 0.2$, $\sigma^2_\xi = 0.5$, $\sigma^2_\eta = 1.2$

**True Parameter Values**

§ $n=40$, $\eta=2.94$, $\xi=-2.20$, $\bar{D}=5.14$  
‡ $n=500$, $\eta=0.62$, $\xi=-0.85$, $\bar{D}=1.47$

### Simulation results for $\bar{D}$

<table>
<thead>
<tr>
<th>N</th>
<th>Bias</th>
<th>Coverage Probability</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>U</td>
<td>AB</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>-0.600</td>
<td>-0.799</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.590</td>
<td>-0.791</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>-0.031</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.031</td>
<td>-0.039</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>-0.031</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.031</td>
<td>-0.039</td>
</tr>
</tbody>
</table>

### Simulation results for $\beta$

<table>
<thead>
<tr>
<th>N</th>
<th></th>
<th></th>
<th>Coverage Probability</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>AB</td>
</tr>
<tr>
<td>25</td>
<td>-0.137</td>
<td>-0.061</td>
<td>0.013</td>
<td>0.899</td>
</tr>
<tr>
<td></td>
<td>-0.141</td>
<td>-0.089</td>
<td>-0.003</td>
<td>0.685</td>
</tr>
<tr>
<td>25</td>
<td>-0.020</td>
<td>0.000</td>
<td>0.006</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>-0.025</td>
<td>-0.008</td>
<td>-0.001</td>
<td>0.929</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Simulation results for $\sigma^2_{D/S}$

<table>
<thead>
<tr>
<th>N</th>
<th></th>
<th></th>
<th>Coverage Probability</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>AB</td>
</tr>
<tr>
<td>25</td>
<td>-0.580</td>
<td>-0.636</td>
<td>-0.081</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>-0.584</td>
<td>-0.607</td>
<td>-0.075</td>
<td>0.411</td>
</tr>
<tr>
<td>25</td>
<td>-0.138</td>
<td>-0.148</td>
<td>-0.091</td>
<td>0.898</td>
</tr>
<tr>
<td></td>
<td>-0.082</td>
<td>-0.090</td>
<td>-0.034</td>
<td>0.907</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Simulation Results (Hamza et al)

- Our result is in agreement with Hamza et al simulation study
- They compared the approximate and exact approaches for meta-analyzing proportions from different studies
- The Exact approach (for the proportion $p$ and between studies variance)
  - quite unbiased estimate
  - reasonable coverage probability
  - the profile likelihood based confidence interval appeared to improve the coverage probability particularly when $N$ is small (for example $N=10$)
  - better than the approximate even if the distribution of the true parameter is skewed
Simulation Results (Hamza et al)

- The approximate method
  - underestimate when \( n \) is small or \( p \) is large/small
  - low coverage probability when \( n \) is small or \( p \) is large/small
- They recommend the use of exact approach whenever it is feasible
We reanalyzed a published meta-analysis data (Oei et al; 2003)

They conduct a MEDLINE search from January 1991 - December 2000

Article were included if

- at least 30 patients were studied
- arthroscopy was the reference standard
- the magnetic field strength was reported
- positivity criterion was defined
- the absolute number of TP, FN, TN and FP results were available or derivable

We used the medial meniscal tears data which included 27 studies
Data Example: Diagnostic Performance of Magnetic Resonance Imaging (MRI)(Cont.)

- the models were fitted in the SAS procedure NLMIXED

<table>
<thead>
<tr>
<th>Parameter Estimates for MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>Parameter estimates</td>
</tr>
<tr>
<td>$\alpha$</td>
</tr>
<tr>
<td>$\beta$</td>
</tr>
<tr>
<td>$\sigma^2_{\alpha}$</td>
</tr>
</tbody>
</table>

| 95% confidence interval     |        |      |      |       |
| $\alpha$                   | [3.92, 5.05] | [3.20, 5.53] | [4.29, 5.67] |
| $\beta$                    | [-0.81, 0.27] | [-2.57, 1.60] | [-1.66, 0.33] |
| $\sigma^2_{\alpha}$        | [0.51, 2.63] | [0.39, 2.42] | [0.36, 3.00] |
Summarizing Sensitivity and Specificity

**Figure:** ROC curves from the three approaches for the MRI data
Compared to the EB approach

- the U underestimates
  - $\alpha$ by 0.492 and $\beta$ by 0.394
- the AB underestimates
  - $\alpha$ by 0.616 and $\beta$ by 0.182
Concluding Remarks

- The EB approach outperformed the U and AB
- The Models can easily be fitted in commercially available statistical packages, for example SAS NLMIXED
- We recommend the use of EXACT BIVARIATE approach
- Different options are available when there is a convergence problem in NLMIXED (SAS Institute Inc 2004. SAS/STAT User Guide)


Hamza TH, van Houwelingen H, Stijnen T. Random effects meta-analysis of proportions: how to model the within study variability? Journal of Clinical Epidemiology(Provisionally Accepted)