Statistical methods for the meta-analysis of full ROC curves

Oliver Kuss, Annika Hoyer

German Diabetes Center,
Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf,
Institute for Biometry and Epidemiology

October 15, 2015
Outline

Introduction

Idea I: Meta-Regression in the standard bivariate model

Idea II: Bivariate time-to-event model for interval-censored data

Discussion
Introduction
Meta-analysis for diagnostic accuracy trials

- Meta-analysis for intervention studies is well established today.
- In contrast, meta-analysis for diagnostic accuracy trials has been a vivid research area in recent years.
- Reason: Increased complexity of diagnostic trials with their bivariate outcome of sensitivity and specificity.
Meta-analysis for diagnostic accuracy trials

- Additional challenge: Single studies report a full ROC curve with several pairs of sensitivity and specificity, each one for a different threshold
- Still greater challenge: Values and numbers of thresholds can vary between single studies
Meta-analysis of diagnostic accuracy trials - An example

<table>
<thead>
<tr>
<th>Study</th>
<th>Cutpoint</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.3</td>
<td>89/115</td>
<td>1382/1684</td>
</tr>
<tr>
<td>1</td>
<td>5.5</td>
<td>72/115</td>
<td>1558/1684</td>
</tr>
<tr>
<td>1</td>
<td>5.6</td>
<td>65/115</td>
<td>1602/1684</td>
</tr>
<tr>
<td>2</td>
<td>5.3</td>
<td>207/252</td>
<td>2704/5865</td>
</tr>
<tr>
<td>2</td>
<td>5.4</td>
<td>196/252</td>
<td>3308/5865</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

...
Meta-analysis of diagnostic accuracy trials - An example

HbA$_1c$ as a screening tool for detection of Type 2 diabetes: a systematic review

C. M. Bennett, M. Guo and S. C. Dharmage

Department of Public Health, School of Population Health, The University of Melbourne, Australia

Use of high-normal levels of haemoglobin A$_1c$ and fasting plasma glucose for diabetes screening and for prediction: a meta-analysis
Meta-analysis of diagnostic accuracy trials - An example

- Population-based screening for type 2 diabetes mellitus
- Two systematic reviews (Kodama, 2013 and Bennett, 2007) report on 38 single studies (one pair of sensitivity and specificity from each study) to assess HbA1c as diagnostic marker
- **However:** Intensified search yields 124 pairs of sensitivity and specificity for 26 different HbA1c thresholds
An analysis using only the original 38 pairs discards more than 70% of the available observations.

Even worse: Using only the original 38 pairs would also ignore the different HbA1c values that were used as thresholds.
Meta-analysis of diagnostic accuracy trials - The paradox of the standard SROC

- A summary ROC curve from the original 38 pairs explicitly ignores the ROC information from the single studies and...
- ‘...is in principle unidentifiable’ [Rücker/Schumacher, 2010]
- ‘...cannot be interpreted as a kind of average curve or a curve typical for the study-specific ROC curves. It can have a shape that is very different from the study-specific shapes.’ [Arends et al., 2008]
Meta-analysis of diagnostic accuracy trials - The paradox of the standard SROC

[Chu/Guo, 2009]
... although several of these methods allow for different test thresholds to be used across the primary studies, none have been used to incorporate threshold values explicitly, a notable limitation.' (Sutton, 2008)

’If more than one threshold is reported per study, this has to be taken into account in the quantitative analyses.’ (Trikalinos, 2012)
Meta-analysis of diagnostic accuracy trials - Current approaches using the full information from each study

- Good news: Methods for the meta-analysis of full ROC curves have been proposed [Kester, 2000; Dukic, 2003; Poon, 2004; Bipat, 2007; Hamza, 2009; Putter, 2010; Martinez-Camblor, 2014; Riley, 2014].
Not so good news: Each of these methods has at least one of the following disadvantages:

- The method constitutes a two-step approach and estimation uncertainty from the first step is ignored in the second step
- The number of thresholds has to be identical across all studies
- The concrete values of the thresholds are ignored
- The method assumes a fixed-effect model
- The method is not applicable with extreme values
Idea I: Meta-Regression in the standard bivariate model
Bivariate logistic regression model with random effects

Sensitivity \( Se = \frac{TP}{TP+FN} \) and specificity \( Sp = \frac{TN}{TN+FP} \)

(TP, FN, TN, FP represent the number of true positives, false negatives, true negatives and false positives, respectively)

\[
TP_i \mid Se_i \sim \text{Binomial}(TP_i + FN_i, Se_i), \quad \text{logit}(Se_i) = \mu + \phi_i
\]

\[
TN_i \mid Sp_i \sim \text{Binomial}(TN_i + FP_i, Sp_i), \quad \text{logit}(Sp_i) = \nu + \psi_i
\]

\[
\begin{pmatrix}
\phi_i \\
\psi_i
\end{pmatrix} \sim N \left[ \begin{pmatrix}
0 \\
0
\end{pmatrix}, \begin{pmatrix}
\sigma_\phi^2 & \rho \sigma_\phi \sigma_\psi \\
\rho \sigma_\phi \sigma_\psi & \sigma_\psi^2
\end{pmatrix} \right]
\]

\[
\text{logit}(p) = \log(p/(1-p)), \mu, \nu \text{ intercepts for } \text{logit}(Se_i), \text{logit}(Sp_i)
\]
Extend to meta-regression model by including covariates

Modify the GLMM including covariates:

$$\text{logit}(Se_i) = \mu + \alpha x_j + \phi_i$$

$$\text{logit}(Sp_i) = \nu + \beta x_j + \psi_i$$

where the $x_j$ are the values of the covariate (in our case: HbA1c) that denotes the (potentially different but maybe equal) values of the threshold $j$, where the number of thresholds can be different between studies $j = 1, \ldots, J_i$. 
Parameter estimation

- The following parameters have to be estimated: $\mu, \nu, \sigma^2_\phi, \sigma^2_\psi, \rho, \alpha, \beta$
- Estimation method: PQL (SAS PROC GLIMMIX)
- However, we are not interested in the model parameters, but in the summary ROC curve
- That is, predict sensitivity and specificity at given thresholds by the BLUP principle (LSMESTIMATE statement, GLIMMIX)
Results for the HbA1c example
Idea II: Bivariate time-to-event model for interval-censored data
ROC curve from Zhang, 2010
ROC curve from Zhang, 2010
'Life-table' from Zhang, 2010
Fundamental insight

An ROC curve can be considered a bivariate time-to-event model for interval-censored data!
Bivariate time-to-event model for interval-censored data - Notation

- Given $K$ observations from a continuous variable $Y$ (in our case: HbA1c) as lying in intervals $(y_k^L, y_k^R]$ with $\Delta_k = y_k^R - y_k^L$
- $Y$ has density $f(y, \mu, \phi)$ with corresponding cdf $F(y, \mu, \theta)$; $\mu$ is a location and $\phi$ a scale parameter
- In general, three types of censoring can occur:

<table>
<thead>
<tr>
<th>Type of Censoring</th>
<th>Interval</th>
<th>Contribution to LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>$y_k^L = 0, y_k^R \neq \infty$</td>
<td>$F(y_k^R; \mu, \phi)$</td>
</tr>
<tr>
<td>Interval</td>
<td>$y_k^L \neq 0, y_k^R \neq \infty$</td>
<td>$[f(y_k; \mu, \phi)\Delta_k]$</td>
</tr>
<tr>
<td>Right</td>
<td>$y_k^L \neq 0, y_k^R = \infty$</td>
<td>$[1 - F(y_k^L; \mu, \phi)]$</td>
</tr>
</tbody>
</table>
We consider three densities for $Y$:

**Weibull**

$$f(y; \mu, \phi) = \frac{\phi y^{\phi-1} \exp\left(-\frac{y}{\mu}\right)}{\mu^\phi}$$

**Log-normal**

$$f(y; \mu, \phi) = \frac{\exp\left(-\frac{\left(\log(y) - \mu\right)^2}{2 \phi}\right)}{y \sqrt{2\pi\phi}}$$

**Log-logistic**

$$f(y; \mu, \phi) = \frac{\pi \exp\left(-\frac{\pi\left(\log(y) - \mu\right)}{(\sqrt{3}\phi)}\right)}{y \sqrt{3\phi}\left(1 + \exp\left(-\frac{\pi\left(\log(y) - \mu\right)}{(\sqrt{3}\phi)}\right)\right)}$$

(See Lindsey, 1998 for more densities)
Bivariate time-to-event model for interval-censored data - Final model

Model two location parameters $\mu_{D^+}, \mu_{D^-}$ one for the diseased ($D^+$) and one for the non-diseased ($D^-$) simultaneously, and link the groups from the same study by a bivariate random effect

$$
\log(\mu_{D^+}) = b_{D^+} + \phi_{D^+}, \log(\mu_{D^-}) = b_{D^-} + \phi_{D^-}
$$

$$
\begin{pmatrix}
\phi_{D^+} \\
\phi_{D^-}
\end{pmatrix}
\sim
N\left[
\begin{pmatrix}
0 \\
0
\end{pmatrix},
\begin{pmatrix}
\sigma_{D^+}^2 & \rho\sigma_{D^+}\sigma_{D^-} \\
\rho\sigma_{D^+}\sigma_{D^-} & \sigma_{D^-}^2
\end{pmatrix}
\right]
$$
Parameter estimation

- The following parameters have to be estimated:
  \( b_{D+}, b_{D-}, \phi_{D+}, \phi_{D-}, \sigma^2_{D+}, \sigma^2_{D-}, \rho \)
- Estimation method: Gaussian quadrature (SAS PROC NLMIXED)
- Again, we are not interested in the model parameters, but in the summary ROC curve
- That is, predict sensitivity and specificity at given thresholds by the BLUP principle (PREDICT statement, NLMIXED)
Results for the HbA1c example

Sensitivity

1-Specificity

- Weibull
- Log-normal
- Log-logistic
- GLMM
Discussion
Summary

We propose two approaches for the meta-analysis of full ROC curves that use the information for all thresholds. Both models avoid the problems of previous methods and come with the following advantages:

- Constitute a one-step approach
- Allow varying numbers and varying values of thresholds
- Explicitly model the value of the diagnostic test
- Allow heterogeneity of sensitivity and specificity across studies
- Account for correlation within studies
Summary

- Ensure proper weighting of studies’ sample size
- Work well in simulations
- Can be generalized to compare two (or even more) diagnostic tests to a common gold standard
Limitations

- Distributional assumptions for the values of the diagnostic test (Idea: Relax this by using a piecewise constant function)
- Robustness
- Idea II: How to choose the minimal value for the diagnostic test?
Thank you!

View from the IBE on Düsseldorf
Bibliography


Bibliography


