Bayesian Meta-analysis of Diagnostic Tests
Allowing for Imperfect Reference Standard

J Menten, M Boelaert, E Lesaffre

Institute of Tropical Medicine, L-Biostat KULeuven

10-May-2012
Overview

- Introduction
  - The clinical problem: meta-analysis of rapid tests for visceral leishmaniasis
  - Accuracy of diagnostic tests
  - Latent class analysis for diagnostic studies

- Meta-analysis of diagnostic studies
  - General principles
  - The bivariate model
  - Allowing for Imperfect Reference Standards

- Simulation study

- Application to VL data
Kala-Azar

Visceral leishmaniasis  
(VL / Kala-azar / Black fever)
- neglected tropical disease
- protozoal disease, transmitted by sandflies
- occurs in poor rural areas of eastern Africa, southern Asia and Latin America
- symptoms: fever, enlarged spleen
- fatal if untreated (50,000 deaths/year)
- treatment is painful, toxic, and expensive
Classical diagnosis

Microscopical examination (+/- culture) of sample from

- **Lymph node**: Low sensitivity
- **Bone Marrow**: Sensitivity: 70-80% Painful Sterilization needed
- **Spleen**: Sensitivity 95% Expertise required Risk of major bleeding

Poorly adapted to the field conditions
Accurate and easy to use rapid diagnostic tests (RDT) are needed
Rapid Diagnostic Test for Leishmaniasis

RK39-based RDT

- Dipstick/strip test on fingerprick blood
- Dichotomous read-out
- Meta-analysis of diagnostic accuracy
- Phase III studies
  - Clinical suspect patients
  - Diagnosed in primary health care
- Problem: lack of perfect reference test
Introduction
Meta-Analysis of Diagnostic Studies
Simulation Study
Application to the Leishmania Rapid Test Data

Overview
Visceral Leishmaniasis
Accuracy of Diagnostic Tests
Latent Class Analysis

Meta-Analysis of RK39 based RDT

- Reference test
  - "Gold standard": aspiration of spleen
    - not 100% sensitive
    - dangerous to perform in low resource settings
    - may be impossible to perform on certain patients

- Reference standards used in publications:
  - varies between studies:
  - may not be perfect
  - some studies use Latent Class Analysis (LCA)

- Statistical model for meta-analysis should allow:
  - for imperfect reference standards
  - for primary studies that use LCA
Summary Measures of Accuracy of Diagnostic Tests

<table>
<thead>
<tr>
<th>Index test result</th>
<th>Reference standard result</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (I+)</td>
<td>Positive (R+)</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>Negative (R-)</td>
<td>b</td>
</tr>
<tr>
<td>Negative (I-)</td>
<td></td>
<td>c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d</td>
</tr>
</tbody>
</table>

- Assuming reference test is perfect:
  - all R+ are diseased (D+)
  - all R- are not diseased (D-):
  - sensitivity $S = P(I+ | D+)$ estimated by $a/(a+c)$
  - specificity $C = P(I- | D-)$ estimated by $d/(b+d)$

- If reference test is not perfect:
  - biased estimates of S and C
Bias due to Imperfect Reference Test

Estimated Specificity Index Test

Specificity Reference Test:
- 100%
- 75%
- 50%

Sensitivity Reference Test

Estimated Specificity Index Test
Bias due to Imperfect Reference Test

- Assuming conditional independence between Index and Reference test:
  - Conditional: on true disease status
  - If $C_R \downarrow$, $\hat{S} \downarrow$
  - If $S_R \downarrow$, $\hat{C} \downarrow$

- If conditional dependence:
  - $S$ and $C$ may be over- or under-estimated

- Possible solution: Latent Class Analysis
Latent Class Analysis for Diagnostic Tests

- Assume:
  - 2 tests, A and B
  - with unknown sensitivity $S_A$ and $S_B$ and specificity $C_A$ and $C_B$
  - applied to sample of size N with disease prevalence $\pi$

- Results:

<table>
<thead>
<tr>
<th></th>
<th>A+</th>
<th>A-</th>
</tr>
</thead>
<tbody>
<tr>
<td>B+</td>
<td>$n_{A+B+}$</td>
<td>$n_{A-B+}$</td>
</tr>
<tr>
<td>B-</td>
<td>$n_{A+B-}$</td>
<td>$n_{A-B-}$</td>
</tr>
</tbody>
</table>

Counts follow a multinomial distribution:

$$n_{AjBk} \sim Mu(N, P(Aj, Bk))$$
Latent Class Analysis for Diagnostic Tests

- Assuming independence of test results given the disease status:

\[
P(A+, B+) = P(A+, B + | D+) \times P(D+) +
\]

\[
= P(A + | D+) \times P(B + | D+) \times P(D+) +
\]

\[
P(A + | D-) \times P(B + | D-) \times P(D-)
\]

\[
= \pi \times S_A \times S_B + (1 - \pi) \times (1 - C_A) \times (1 - C_B)
\]

- Similar equations for \(P(A-, B+), P(A+, B-),\) and \(P(A-, B-)\)
Solve for $S_A$, $S_B$, $C_A$, $C_B$, and $\pi$.

Problem:
- 3 independent equations
- 5 unknowns

Can be solved if:
- $\geq 3$ tests
- applied to $\geq 2$ populations with different prevalence
- using informative priors in a Bayesian setting
Meta-Analysis of Diagnostic Studies

General aspects for diagnostic study meta-analysis

- Joint modeling of S and C
- Heterogeneity between studies
- Possible correlation between S and C

Bivariate model (Reitsma, 2005)

- Hierarchical modeling of S and C
- Expanding the bivariate model:
  - Inclusion of primary publications that use LCA
  - Allow for imperfect reference standards
The Bivariate Model

- In each study $i$:
  - number of true positives $= y_{Si} \sim \text{Binomial}(n_{Si}, S_i)$, with $n_{Si}$ the number of diseased subjects
  - number of true negatives $= y_{Ci} \sim \text{Binomial}(n_{Ci}, C_i)$, with $n_{Ci}$ the number of non-diseased subjects

- At a higher level:
  - $\text{logit}(S_i) = \theta_{Si}$ and $\text{logit}(C_i) = \theta_{Ci}$ follow a bivariate normal distribution:
    $$
    \begin{pmatrix}
    \theta_{Si} \\
    \theta_{Ci}
    \end{pmatrix}
    \sim
    \mathcal{N}
    \left(
    \begin{pmatrix}
    \mu_S \\
    \mu_C
    \end{pmatrix},
    \Sigma
    \right)
    \text{ with } \Sigma =
    \begin{pmatrix}
    \sigma^2_S & \sigma_{SC} \\
    \sigma_{SC} & \sigma^2_C
    \end{pmatrix}
    $$
Adding Results from LCA

- Primary studies using LCA
- At study level, \( \hat{S}_i \) and \( \hat{C}_i \) with 95% CIs are reported:
  - we obtain: \( \text{logit}(\hat{S}_i) = \hat{\theta}_i \), \( \text{logit}(\hat{C}_i) = \hat{\theta}_i \), \( \sigma_{\hat{\theta}_i} \), \( \sigma_{\hat{\theta}_i} \)
- We assume:
  \[
  \hat{\theta}_i \sim N(\theta_i, \sigma^2_{\hat{\theta}_i})
  \]
  \[
  \hat{\theta}_i \sim N(\theta_i, \sigma^2_{\hat{\theta}_i})
  \]
- \( \theta_i \) and \( \theta_i \) follow the same bivariate normal distribution as before
Assumption: in each study R is a perfect reference test:
- $S_R = 1$ and $C_R = 1$
- often not true - especially not in field trials

Relax this assumption:
- model the 2x2 table of Index versus Reference Test for each study $i$
- estimate diagnostic accuracy of index ($S_i$, $C_i$) and reference test ($S_{Ri}$, $C_{Ri}$)
- similar to Latent Class Analysis
Modeling of Index versus Reference Test

- Results of Index vs. Reference test for study i:

<table>
<thead>
<tr>
<th></th>
<th>R+</th>
<th>R-</th>
</tr>
</thead>
<tbody>
<tr>
<td>I+</td>
<td>$y_{i11}$</td>
<td>$y_{i10}$</td>
</tr>
<tr>
<td>I-</td>
<td>$y_{i01}$</td>
<td>$y_{i00}$</td>
</tr>
</tbody>
</table>

- Counts follow a multinomial distribution:

$$y_{ijk} \sim \text{Mu}(n_i, p_{ijk})$$

with $n_i$ the sample size in study $i$
Modeling of Index versus Reference Test

- From LCA:

\[ p_{ijk} = \pi_i[S_i^j (1 - S_i)^{1-j} S_{Ri}^k (1 - S_{Ri})^{1-k} + (-1)^{j-k} \text{cov}_{i|D=1}] + (1 - \pi_i)[C_i^{1-j} (1 - C_i)^j C_{Ri}^{1-k} (1 - S_{Ri})^k + (-1)^{j-k} \text{cov}_{i|D=0}] \]

- Model \( S_i \) and \( C_i \) sing the bivariate model as before
- Constraints needed for identifiability, possibilities:
  - Constraints on covariances (e.g., \( \text{cov}_{i|D=1} \equiv \text{cov}_{i|D=0} \equiv 0 \))
  - Modeling and/or informative priors for \( S_{Ri} \) and \( C_{Ri} \)
Diagnostic accuracy \((S_{Ri}, C_{Ri})\) of reference standard

- Reference standard classified into \(J\) types
- For each type of reference standard \(j\):
  - elicit estimates of \(S_{Rj}\) and \(C_{Rj}\) from experts
  - enter this information in the model as priors

- \(S_{Ri}, C_{Ri}\) either:
  - assumed constant across studies using the same reference standard (complete pooling cfr. Gelman & Hill)
  - unmodelled (no pooling)
  - modeled using a bivariate normal for \((\text{logit}(S_{Ri}), \text{logit}(C_{Ri}))\) (partial pooling)

- Identifiability of the model?
Setup of the Simulation Study

- **Index test:**
  - average $S$: 90%, $\sigma_S$: 0.5
  - average $C$: 90%, $\sigma_C$: 0.5

- **4 Types of reference tests:**
  - Low $S_{R1}$ (85%, $\sigma_{S_{R1}} = 0.5$), Perfect $C_{R1}$ (100%)
  - Perfect $S_{R2}$ (100%), Low $C_{R2}$ (85%, $\sigma_{C_{R2}} = 0.5$)
  - Moderate $S_{R3}$ (92%, $\sigma_{S_{R3}} = 0.25$) and $C_{R3}$ (92%, $\sigma_{C_{R3}} = 0.25$)
  - Latent class analysis

- **Sample size**
  - 5 studies/reference test (20 studies in total)
  - Sample sizes: 300, 200, 200, 100, 100
  - Prevalence: 50%
  - 100 simulated data-sets
Simulated Data

Reference test:
- Low S, Perfect C
- Perfect S, Low C
- Moderate S, Moderate C
- Latent Class Analysis
Analysis of the Simulation Study

- Bivariate normal model for the index test
- Models for the reference tests:
  - complete pooling across studies
  - no pooling across studies
  - hierarchical model (partial pooling)
- 4 Priors:

<table>
<thead>
<tr>
<th></th>
<th>Perfect</th>
<th>Correct</th>
<th>Vague</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>100</td>
<td>85 (80-90)</td>
<td>90 (50-100)</td>
<td>98 (95-100)</td>
</tr>
<tr>
<td>$C_1$</td>
<td>100</td>
<td>100 (98-100)</td>
<td>90 (50-100)</td>
<td>90 (85-95)</td>
</tr>
<tr>
<td>$S_2$</td>
<td>100</td>
<td>100 (98-100)</td>
<td>90 (50-100)</td>
<td>90 (85-95)</td>
</tr>
<tr>
<td>$C_2$</td>
<td>100</td>
<td>85 (80-90)</td>
<td>90 (50-100)</td>
<td>98 (95-100)</td>
</tr>
<tr>
<td>$S_3$</td>
<td>100</td>
<td>93 (88-98)</td>
<td>90 (50-100)</td>
<td>85 (80-90)</td>
</tr>
<tr>
<td>$C_3$</td>
<td>100</td>
<td>93 (88-98)</td>
<td>90 (50-100)</td>
<td>85 (80-90)</td>
</tr>
</tbody>
</table>

Supplied as normal priors for the logit $S_{Ri}$ and $C_{Ri}$
Introduction
Meta-Analysis of Diagnostic Studies
Simulation Study
Application to the Leishmania Rapid Test Data

DIC

J Menten, M Boelaert, E Lesaffre

Bayesian Meta-analysis of Diagnostic Tests
Parameter Estimates: Means

- **mu[S] Estimate**
  - Complete pooling
  - No-pooling
  - Partial pooling

- **mu[C] Estimate**
  - Complete pooling
  - No-pooling
  - Partial pooling
Standard Error of Parameter Estimates

- mu[S] SE
- mu[C] SE

- Complete pooling
- No-pooling
- Partial pooling

J Menten, M Boelaert, E Lesaffre

Bayesian Meta-analysis of Diagnostic Tests
## Coverage

Coverages of parameter estimates:

<table>
<thead>
<tr>
<th>Prior / Model</th>
<th>$\mu_S$</th>
<th>$\mu_C$</th>
<th>$\sigma_S$</th>
<th>$\sigma_C$</th>
<th>$\sigma_{SC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Reference Test</td>
<td>21</td>
<td>24</td>
<td>95</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>Correct: complete pooling</td>
<td>94</td>
<td>95</td>
<td>90</td>
<td>85</td>
<td>99</td>
</tr>
<tr>
<td>Correct: no pooling</td>
<td>96</td>
<td>94</td>
<td>87</td>
<td>84</td>
<td>100</td>
</tr>
<tr>
<td>Correct: partial pooling</td>
<td>92</td>
<td>95</td>
<td>83</td>
<td>89</td>
<td>100</td>
</tr>
<tr>
<td>Vague: complete pooling</td>
<td>96</td>
<td>96</td>
<td>89</td>
<td>83</td>
<td>100</td>
</tr>
<tr>
<td>Vague: no pooling</td>
<td>90</td>
<td>85</td>
<td>93</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>Vague: partial pooling</td>
<td>92</td>
<td>92</td>
<td>89</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>Incorrect: complete pooling</td>
<td>99</td>
<td>98</td>
<td>53</td>
<td>36</td>
<td>73</td>
</tr>
<tr>
<td>Incorrect: no pooling</td>
<td>99</td>
<td>94</td>
<td>42</td>
<td>22</td>
<td>48</td>
</tr>
<tr>
<td>Incorrect: partial pooling</td>
<td>98</td>
<td>96</td>
<td>46</td>
<td>30</td>
<td>66</td>
</tr>
</tbody>
</table>
Conclusion of simulation study

- Approach appears feasible and results in correct estimates
- Little difference between pooling/no-pooling/partial pooling
  - Use no-pooling as standard approach, unless model is non-identifiable
- Even vague priors result in identifiable models, which are an improvement from standard model
Leishmania Rapid Test Data

- **Available data:**
  - 18 studies with reference standard
  - 6 studies using LCA
  - 4 have data available for 2 reference tests

- **3 Reference standards:**
  - Spleen parasitology
  - Combined reference standards:
    - Bone marrow/lymph node parasitology and alternative diagnosis
    - Spleen parasitology and serological test

- **Main explanatory variable:** geographic region
  - Subsaharan Africa
  - Indian Subcontinent
  - Mediterranean and Latin-America
Leishmania Rapid Test Data

- Reference standard: spleen aspirate
- Combined reference standard:
- BM/LN aspirate and alternative diagnosis
- Spleen aspirate and serological test
- Latent Class Analysis

Graph showing sensitivity vs. specificity with different markers for each test type.
Link Function

Logit link

Complementary log-log link

Sensitivity vs. Specificity

J Menten, M Boelaert, E Lesaffre
Bayesian Meta-analysis of Diagnostic Tests
Imperfect Reference Test

- Priors obtained through expert opinion (preliminary)
  - Data from 6 experts
  - Average S/C + plausible range (95 out of 100 studies)
  - Bivariate normal model fitted to the (cloglog-transformed) expert data
  - Spleen parasitology
    - $S_1$: 93% (87-98); $C_1$: 98% (90-100)
  - Bone marrow/lymph node parasitology and alternative diagnosis
    - $S_2$: 84% (81-88); $C_2$: 96% (80-100)
  - Spleen parasitology and serological test
    - $S_3$: 98% (97-99); $C_3$: 94% (80-99)
Imperfect Reference Test

Analysis:

- Model choices:
  - Complete pooling: expert opinion used as prior for $\mu_{S_j}$ and $\mu_{C_j}$
  - No pooling: expert opinion used as prior for the study specific $S_{Ri}$ and $C_{Ri}$
  - Partial pooling: expert opinion used as priors for $\mu_{S_j}$, $\mu_{C_j}$, $\sigma_{S_j}$, and $\sigma_{C_j}$

- Priors:
  - Assume perfect reference standard
  - Priors from expert opinion
  - Vague priors
## Model Fit

**DIC:**

<table>
<thead>
<tr>
<th>Prior:</th>
<th>Perfect RT</th>
<th>Expert Opinion</th>
<th>Vague</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Pooling</td>
<td>224.5</td>
<td>247.7</td>
<td>220.7</td>
</tr>
<tr>
<td>No Pooling</td>
<td>279.6</td>
<td>314.7</td>
<td></td>
</tr>
<tr>
<td>Partial Pooling</td>
<td>260.0</td>
<td>219.7</td>
<td></td>
</tr>
</tbody>
</table>

**Model fit:**

- Poor with priors obtained from expert opinion
- Better with vague prior
- Poor with no pooling model (identifiability ?)
Diagnostic Accuracy of Index Tests

Reference standards perfect

Expert opinion

Sensitivity

Specificity

Indian Subcontinent
East-Africa
Latin-America & Mediterranean

J Menten, M Boelaert, E Lesaffre
Bayesian Meta-analysis of Diagnostic Tests
### Diagnostic Accuracy of Reference Test

#### Sensitivity and specificity (Partial pooling model):

<table>
<thead>
<tr>
<th></th>
<th>Supplied Mean (95% PI)</th>
<th>Estimated from model with Expert Prior Estimate (95% CI)</th>
<th>Vague Prior Estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Spleen parasitology</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>93.4 (87.2, 97.9)</td>
<td>94.1 (92.0, 95.9)</td>
<td>98.1 (92.9, 100)</td>
</tr>
<tr>
<td>C</td>
<td>98.3 (89.8, 99.9)</td>
<td>98.9 (97.7, 99.6)</td>
<td>99.9 (98.7, 100)</td>
</tr>
<tr>
<td><strong>Bone marrow/lymph node parasitology and alternative diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>84.1 (81.4, 87.8)</td>
<td>86.6 (84.6, 88.5)</td>
<td>98.3 (95.5, 99.9)</td>
</tr>
<tr>
<td>C</td>
<td>96.3 (80.1, 100)</td>
<td>93.2 (86.9, 98.6)</td>
<td>91.4 (83.7, 99.1)</td>
</tr>
<tr>
<td><strong>Spleen parasitology and serological test</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>97.7 (97.0, 98.8)</td>
<td>98.0 (97.2, 98.6)</td>
<td>99.4 (97.5, 100)</td>
</tr>
<tr>
<td>C</td>
<td>94.0 (79.9, 99.3)</td>
<td>95.6 (92.6, 97.7)</td>
<td>99.6 (95.5, 100)</td>
</tr>
</tbody>
</table>
Conclusions and Further Work

Conclusions

- Correcting for imperfect reference tests results may result in important changes in estimates for S or C of the index test
- In the application, there appears to be some conflict between supplied prior and observed data

Further work

- Incorporate conditional dependencies between reference and index test
- More appropriate pooling of expert opinions
- Further assess influence of prior