Introduction to Diagnostic Accuracy Reviews

Hans Reitsma  MD, PhD
Dept. of Clinical Epidemiology, Biostatistics & Bioinformatics
Academic Medical Center - Amsterdam
Continental European Support Unit - Cochrane Collaboration

Amsterdam
June, 2009
Overall Goal

To perform high quality and clinical relevant reviews of diagnostic test accuracy
Good Diagnostic Accuracy Reviews

- High quality methodology
  - searching and locating
  - quality assessment
  - statistical analysis
- Continental Europe Support Unit
- Editorial process for Cochrane reviews
Relevant Diagnostic Accuracy Reviews

- Clinical useful accuracy reviews:
  - clear and focussed question
  - accuracy helpful in answering the question?

- Selecting and phrasing the right questions for diagnostic reviews is the biggest challenge
Outline

- Inherent complex nature of diagnostic research
- Role of accuracy studies
- Key features of accuracy studies
- Focussed questions for accuracy reviews
Diversity Diagnostic Research

- “Tests”: any method to obtain information on the patient’s health status

- Role of tests: making a diagnosis, but also staging, screening, selecting treatment, monitoring health
Complexity of Diagnostic Process

- Multiple tests can be performed

- Multiple diseases can be responsible for presenting symptoms

- Dynamic and multistage process
Multistage, Multidimensional Diagnostic Process

Presenting symptoms → Further Testing 1 → Further Testing 2 → Therapeutic actions → Patient Outcomes

No further testing and treatment
**Multistage, Multidimensional Diagnostic Process**

1. **Presenting symptoms**
   - Further Testing 1
   - Further Testing 2
     - Therapy
       - Patient Outcomes
     - No therapy
       - Patient Outcomes

Cave: focus on single disease
Randomized Trial of Tests

- Specific point in diagnostic work-up (final stage)

- Question: should we do Test A or B?

- Goal: directly observe whether patient outcomes are better with strategy based on test A than on test B
Relation between Tests and Patient Outcomes

Medical Test → Test Results → Clinical Decision

Clinical Decision:
- Treat → Patient Outcome
- No Treat → Patient Outcome

Patient Outcome
RCT of test-treatment combinations

Study Population

Test A
- positive result → treat X
- negative result → treat Y

Test B
- positive result → treat X
- negative result → treat Y
Randomised controlled trial of faecal-occult-blood screening for colorectal cancer

Jack D Hardcastle, Jocelyn O Chamberlain, Michael H E Robinson, Susan M Moss, Satya S Amar, Tom W Balfour, Peter D James, Christine M Mangham

**Summary**

**Background** There is growing evidence that faecal-occult-blood (FOB) screening may reduce colorectal cancer (CRC) mortality, but this reduction in CRC mortality has not been shown in an unselected population-based randomised controlled trial. The aim of this study was to assess the effect of FOB screening on CRC mortality in such a setting.

**Methods** Between February, 1981, and January, 1991, 152 850 people aged 45–74 years who lived in the Nottingham area of the UK were recruited to our study. Participants were randomly allocated FOB screening (76 466) or no screening (controls; 76 384). Controls were not told about the study and received no intervention. Screening-group participants were sent a Haemoccult FOB test kit with instructions from their family doctor. FOB tests were not rehydrated and dietary restrictions were imposed only for retesting borderline results. Individuals with negative FOB tests at the first screening, together with those who tested positive but in whom no neoplasia was found on colonoscopy, were invited to take part in further screening every 2 years. Screening was stopped in February, 1995, by which time screening-group participants had undergone 5 FOB tests at a median of three retesting.

**Results** Of 152 850 participants invited to undergo screening, 32 277 (21·1%) refused participation. The overall participation rate at the first FOB test was 70·1% (76 466/109 273). One hundred and four people died from CRC in the screening group compared with 420 in the control group—a 15% reduction in cumulative CRC mortality in the screening group (odds ratio=0·85 [95% CI 0·74–0·98], p=0·026).

**Interpretation** Our findings together with evidence from other trials suggest that consideration should be given to a national programme of FOB screening to reduce CRC mortality in the general population.


See Commentary page 1463

**Introduction**

Colorectal cancer (CRC) is the second commonest cause of death from malignant disease in England and Wales, and resulted in about 16 000 deaths in 1993.1 Although there have been advances in the management of symptomatic CRC, there has been little overall reduction in CRC mortality during the past 30 years. Tumour stage is an important determinant of outcome; 24–28% of patients have metastatic disease at presentation and the tumour is confined to the bowel wall in only 6–10% (Dukes’ stage A).2–4 Early diagnosis before the
Figure 1: **Trial profile**
RCT of Testing

- Best evidence of effectiveness, but rare
- Usually need large sample sizes
- Dynamic, multi-stage nature of diagnostic process leads to infinite number of trials

- Intervention handbook of Cochrane can be used to review such trials
Multistage, Multidimensional Diagnostic Process

Presenting symptoms → Further Testing 1 → Further Testing 2 → Therapy

No therapy → Patient Outcomes

Cave: focus on single disease
Classical Diagnostic Thinking

- Multivariable Approach:
  - probability disease being present given clinical profile
  - multiple pieces of information
  - estimate independent contribution (weight)
  - multivariable analysis
  - clinical decision rules
## Wells Rule for DVT

**Table 1. The Wells Rule To Estimate the Probability of Deep Venous Thrombosis**

<table>
<thead>
<tr>
<th>Clinical Feature</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active cancer</td>
<td>1</td>
</tr>
<tr>
<td>Paralysis, paresis, or recent plaster immobilization of the lower extremity</td>
<td>1</td>
</tr>
<tr>
<td>Recently bedridden for more than 3 days or major surgery within 4 weeks</td>
<td>1</td>
</tr>
<tr>
<td>Localized tenderness along the distribution of the deep venous system</td>
<td>1</td>
</tr>
<tr>
<td>Entire leg swollen</td>
<td>1</td>
</tr>
<tr>
<td>Calf swelling by more than 3 cm when compared with the asymptomatic leg</td>
<td>1</td>
</tr>
<tr>
<td>Pitting edema (greater in the symptomatic leg)</td>
<td>1</td>
</tr>
<tr>
<td>Collateral superficial veins (nonvaricose)</td>
<td>1</td>
</tr>
<tr>
<td>Alternative diagnosis as likely or more possible than that of deep venous thrombosis</td>
<td>−2</td>
</tr>
</tbody>
</table>
Possible Review Questions in Rules

- What is the optimal weight for a specific item?
- Selection of best subset of items
- Focus is on multivariable adjusted weights. Not possible in Cochrane
# Landscape of Diagnostic Research

## "Multivariable approach"
- probability disease given profile
- multiple pieces of information
- estimate independent contribution
- multivariable analysis
- clinical decision rules

## Other Forms of Test Evaluation
- technical evaluation
- diagnostic accuracy

## Randomised Trials of Testing
- randomise between test-treatment combinations
- compare clinical outcomes

**Review:**
pooling of regression weights

**Cochrane:** NO

**Review of RCT’s**

**Cochrane:** YES
but use framework of intervention reviews
Test Evaluation

- Early technical evaluation: repeatability & reproducibility, inter-observer variability
- Diagnostic accuracy
- Patient outcome / cost-effectiveness
Diagnostic Accuracy Studies

- Test under evaluation (index test) and reference standard in all patients
- Compare the results of the index test(s) with the results of reference standard: cross sectional relationship
- Accuracy refers to the amount of agreement
Basic Design

Series of patients

Index test

Reference standard

Cross-classification
### Reference Standard

<table>
<thead>
<tr>
<th>Condition present</th>
<th>Condition absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Negative</td>
<td>Negative</td>
</tr>
</tbody>
</table>

#### 2 by 2 Table

<table>
<thead>
<tr>
<th></th>
<th>Reference Standard</th>
<th>Index test</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>a</td>
<td>Positive</td>
</tr>
<tr>
<td>FP</td>
<td>b</td>
<td>Negative</td>
</tr>
<tr>
<td>FN</td>
<td>c</td>
<td></td>
</tr>
<tr>
<td>TN</td>
<td>d</td>
<td></td>
</tr>
</tbody>
</table>

- TP: True Positive
- FP: False Positive
- FN: False Negative
- TN: True Negative
Clinical Problem

- Patient with chest pain suggestive for acute myocardial infarction (AMI)

- Index test: creatine kinase (CK) measurement

- Do low values of CK measurement rule out myocardial infarction (triage)?
Anatomy of Accuracy Study

- Target population: patients with chest pain
- Index test: CK measurement
- Target condition: acute myocardial infarction
- Final diagnosis based on WHO criteria (reference standard):
  - clinical outcome
  - ECG-changes
  - enzym values
  - autopsy
Example

Patients with chest pain

CK measurement

WHO criteria for AMI

Cross-classification
# Results of CK Study

<table>
<thead>
<tr>
<th>CK</th>
<th>AMI</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
<td>Absent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>high (&gt;80)</td>
<td>215</td>
<td>16</td>
<td>231</td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>15</td>
<td>114</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>130</td>
<td>360</td>
<td></td>
</tr>
</tbody>
</table>
### Test Accuracy Measures

- **sensitivity** $215 / 230 = 93\%$
- **specificity** $114 / 130 = 88\%$
- **odds ratio**: $(215/15) / (16/114) = 102$

<table>
<thead>
<tr>
<th>CK</th>
<th>AMI</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high (&gt;80)</td>
<td></td>
<td></td>
<td>215</td>
<td>16</td>
</tr>
<tr>
<td>low</td>
<td></td>
<td></td>
<td>15</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>230</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>360</td>
<td></td>
</tr>
</tbody>
</table>
Role of Accuracy Studies

Relevant and focused questions
View on Accuracy

- Accuracy can be seen as a surrogate for clinical outcomes

- Evidence from accuracy studies combined with existing data about effectiveness can predict the outcome of hypothetical trial
Test A or B?

Test A

- True positives \([p \times se_A]\)
- False positives \([(1-p) \times (1 - spec_A)]\)
- False negatives \([p \times (1 - se_A)]\)
- True negatives \([(1-p) \times sp_A]\)

Treat

- Risk reduction as observed in trial + risk side effects
- Risk side effects, no treatment effect
- Risk untreated
- None

Test B

- True positives \([p \times se_B]\)
- False positives \([(1-p) \times (1 - spec_B)]\)
- False negatives \([p \times (1 - se_B)]\)
- True negatives \([(1-p) \times sp_B]\)

Treat

- Risk reduction as observed in trial + risk side effects
- Risk side effects, no treatment effect
- Risk untreated
- None

\(p = \text{prevalence}\)
\(se = \text{sensitivity}\)
\(sp = \text{specificity}\)

*Sutton et al.* Integration of meta-analysis and economic decision modelling for evaluating tests. MDM 2008
Relevant Accuracy Questions

- Describe existing diagnostic pathway

- Formulate the intended role of the index test (triage, replacement, add-on)
Intended Role of Test

- Describing the change in the diagnostic pathway is helpful for:
  - patient selection
  - selecting the right comparator test
  - choice and interpretation of accuracy measures
## Landscape of Diagnostic Research

<table>
<thead>
<tr>
<th>“Multivariable approach”</th>
<th>Other Forms of Test Evaluation</th>
<th>Randomised Trials of Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>· multiple pieces of information</td>
<td>Phased approach</td>
<td>· randomise between test-treatment combinations</td>
</tr>
<tr>
<td>· estimate independent contribution</td>
<td>· technical evaluation</td>
<td>· compare clinical outcomes</td>
</tr>
<tr>
<td>· multivariable analysis</td>
<td>· diagnostic accuracy</td>
<td></td>
</tr>
<tr>
<td>· probability disease given profile</td>
<td>· clinical outcomes</td>
<td></td>
</tr>
<tr>
<td>· clinical decision rules</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key question:
Is accuracy sufficient?

Review:
pooling of regression weights

Cochrane: NO

Review of accuracy studies

Cochrane: YES

Review of RCT’s

Cochrane: YES
but use framework of intervention reviews
Take Home Messages

- Focus of accuracy reviews is on test evaluation
- The ultimate goal is to determine whether change in testing improves patient outcomes
- Randomized trials of test-treatment combinations provide the best direct evidence
- Accuracy studies meaningful if their design match the intended role