Anticipating missing reference standard data when planning diagnostic accuracy studies

Christiana Naaktgeboren
Memtab - July, 2016

BMJ 2016;352:i402
Christiana A Naaktgeboren, Joris A H de Groot, Anne W S Rutjes, Patrick M M Bossuyt, Johannes B Reitsma, Karel G M Moons
The problem: missing reference standard data

In diagnostic studies the “outcome” of interest is often missing in some patients.

This missing data often leads to biased results. There are various approaches for dealing with this bias.

QUESTION: How to approach missing reference standard data when designing a prospective diagnostic study?
Outline

• How the problem arises (verification patterns)

• Proposed solutions

• Implications for study design and planning
How the problem arises

Incidental missings

Data missing by study design

Data missing due to clinical practice

Data missing due to infeasibility
How the problem arises

Classic design

- Patients
- Index test
- Reference standard

Include all participants in analysis
How the problem arises

Partial Verification

Patients

Index test

Reference standard available

Include only some patients in analysis (selection bias)

Reference standard missing
Analytical approaches to partial verification

Complete case analysis

Inverse probability weighting ("Begg and Greenes method")

Sensitivity analysis

Multiple imputation
Another approach to partial verification

Differential Verification

 Patients

 Index test

 Preferred Reference Standard

 Alternative Reference Standard

 Include all participants in analysis
Analytical approaches to differential verification

Report results separately by reference standard used

Bayesian correction method
Proposed analytical solutions

Incidental missing data
- Complete case analysis may suffice

Data missing by research design
- Inverse Probability Weighting

Data missing due to clinical practice
- Multiple Imputation

Data missing due to infeasibility
- Bayesian correction method for Differential Verification

Complete index test dependent Differential verification
Considerations for design

- **Incidental missing data**
  - Complete case analysis may suffice

- **Data missing by research design**
  - Inverse Prob. Weight

- **Data missing due to clinical practice**
  - Multiple Imput.
  - Bayesian correction method for Dif. Verif.

- **Data missing due to infeasibility**
  - Complete index test dependent Dif. Verif.

Additional considerations:

- Consider number of patients in subgroup that will be verified
- Perform additional tests and record additional information to improve imputation
- Apply secondary reference standard and obtain and incorporate external data on its imperfection
Beyond the accuracy framework

What we really want to know: does the test improve patient outcomes?
Conclusions

• Missing reference standard results are often inevitable and may induce bias

• Analytical solutions are available
  – They require knowledge about the verification pattern
  – They are no substitute for complete data
  – **Measures can and should be taken before the study starts to facilitate correction methods**

• Clearly report missing data (STARD)
Extra slides
Analysis of incidental missings

Eligible patients
n=1000

No index test
n=0

Rapid diagnostic test (RDT)
n=1000

RDT+
n=30

RDT-
n=950

No reference
Blood sample lost n=2

Microscopy
n=28

Malaria n=20
No Malaria n=8

Microscopy
n=930

Malaria n=30
No Malaria n=900

No reference
Blood sample lost n=20
### Analysis of incidental missings

- Complete case analysis likely to be ok

<table>
<thead>
<tr>
<th>Microscopy</th>
<th>RDT +</th>
<th>RDT -</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>20</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>-</td>
<td>8</td>
<td>900</td>
<td>20</td>
</tr>
<tr>
<td>missing</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Eligible patients
N=10000

Visual inspection with acetic acid (VIA)
N=10000

VIA+
N=2000

VIA-
N=8000

Select a random subset (10%) to verify

No reference
n=7200

Colposcopy + biopsy
n=200

Cervical cancer n=300
No cervical cancer n=1700

Coposcopy + biopsy
n=800

Cervical cancer n=150
No cervical cancer n=650
Analysis of data missing by study design

- Reweight (Inverse Probability Weighing)

<table>
<thead>
<tr>
<th></th>
<th>VIA +</th>
<th>VIA -</th>
<th>missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colposcopy + biopsy</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>VIA +</td>
<td>300</td>
<td>1700</td>
<td>0</td>
</tr>
<tr>
<td>VIA -</td>
<td>150</td>
<td>650</td>
<td>7200</td>
</tr>
</tbody>
</table>
Analysis of data missing due to clinical practice

Eligible patients
n=100

Fecal calprotectin
n=100

No index test
n=0

Endoscopy + biopsy
n=25

Follow-up
n=75

No reference
n=5

Red flag during follow-up
n=5

Endoscopy + biopsy
n=25

Follow-up complete
n=75

Irritable bowel disease (IBD)  
No IBD

Irritable bowel disease (IBD)  
No IBD

Only apply endoscopy when there is a red flag symptom
Analysis of data missing due to clinical practice

- Use Multiple imputation and/or Bayesian correction method

<table>
<thead>
<tr>
<th>Endoscopy + Biopsy</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FCP +</td>
<td>FCP +</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>55</td>
</tr>
</tbody>
</table>

Incomplete data  Imputed data  Analysis results  Pooled results
Eligible patients
N=1000

Ultrasonography
N=1000

Ultrasonography+
N=150

Ultrasonography-
N=950

Biopsy
N=150

Follow-up
N=950

Cancer
n=25
No Cancer
n=125

Cancer
n=1
No Cancer
n=949

No reference
n=0

No index test
n=0

Analysis of data due to infeasibility
Analysis of data due to infeasibility

- Report two separate 2x2 tables
- Report predictive values
  - PPV = 25/(25+125) (with respect to Biopsy)
  - NPV = 949/(1+949) (with respect to follow-up)

<table>
<thead>
<tr>
<th>Biopsy</th>
<th>25</th>
<th>125</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
</tr>
<tr>
<td>-</td>
</tr>
</tbody>
</table>