The STARD Guidelines: Reporting Studies of Diagnostic Accuracy

Mary M. Christopher
University of California-Davis
Co-Editor-in-Chief, Veterinary Clinical Pathology

Goals of This Presentation

- To emphasize the role of scientific journals in asserting terminology and reporting standards
- To describe the STARD Statement for reporting studies of diagnostic accuracy
- To consider the relevance of STARD for veterinary journals

Who Defines Terminology and Reporting Standards?

- Professional organizations
  - Specialty colleges
  - Consensus groups and committees (SNOMED)
- Government agencies
  - National Library of Medicine (MEDLINE, MeSH)
  - World Health Organization
- Scientific journals
  - Specialty vs general
  - Authors, peer reviewers, editors

Terms Used to Report Diagnostic Test Performance

- The ability of a test to measure what it is supposed to measure
  - Accuracy
  - Sensitivity & specificity
  - Criterion
  - Validity
  - Diagnostic utility
- The ability of a test to provide similar results in all applications
  - Reliability
  - Repeatability
  - Reproducibility
  - Agreement
  - Correlation
  - Inter- and intra-observer variation

Michel et al. J Med Libr Assoc 2006;94:221

MeSH Term Equivalents

<table>
<thead>
<tr>
<th>Concepts</th>
<th>MeSH Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity</td>
<td>Sensitivity and specificity</td>
</tr>
<tr>
<td></td>
<td>Reproducibility of results</td>
</tr>
<tr>
<td>Reliability</td>
<td>Reproducibility of results, observer variation</td>
</tr>
<tr>
<td></td>
<td>Diagnostic errors (false negative and false negative reactions)</td>
</tr>
</tbody>
</table>

- Concepts of validity (accuracy) and reliability are confused
- Many concepts not represented

The STARD Initiative (2003)

- Standards for Reporting of Diagnostic Accuracy
  - International working group of epidemiologists, statisticians, clinical pathologists, journal editors
- Follow-up on the Cochrane Diagnostic and Screening Test Methods Working Group
STARD: Rationale
- New diagnostic tests are developed at a fast rate and are essential to medicine
- Survey of journal articles (1978-1993) showed absence of critical information about design and conduct of diagnostic studies
- Exaggerated and biased results from poorly designed and reported studies
- Premature dissemination
- Incorrect diagnosis and treatment decisions

STARD: Objective
- To improve the accuracy and completeness of reporting of studies of diagnostic accuracy
  - Allow readers to assess the potential for bias (internal validity)
  - Allow readers to evaluate its generalizability (external validity)
- To improve the value of published research for clinicians, scientists, reviewers, journals, and the public

STARD: Methods
- Searched the literature for published guidelines on the appropriate conduct and reporting of diagnostic studies
- Definitions
  - Index test: the test under evaluation
  - Reference (gold) standard: best available method
  - Diagnostic accuracy: amount of agreement between index and reference methods
    - Sensitivity and specificity, likelihood ratios, odds ratios, area under the ROC curve

The STARD Initiative
- Consensus checklist (25 items) arranged under the usual headings of a medical research article
- Flow diagram
- Published online: www.consort-statement.org/stardstatement.htm

The STARD Statement
- Explanatory document to facilitate the use, understanding, and dissemination of the checklist
- Summary of available evidence on bias and applicability

The STARD Checklist
- TITLE: Identify as a study of “diagnostic accuracy”; use the MeSH heading “sensitivity and specificity”
- ABSTRACT: State the research question or study aims; avoid general descriptions, such as “diagnostic value” or “clinical usefulness”
- KEY WORDS: Identify as a study of “diagnostic accuracy”; use the MeSH heading “sensitivity and specificity”
- INTRODUCTION: State the research question or study aims; avoid general descriptions, such as “diagnostic value” or “clinical usefulness”
The STARD Checklist

**METHODS**

<table>
<thead>
<tr>
<th>Study population</th>
<th>Inclusion/exclusion criteria; prospective or retrospective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test methods</td>
<td>Reference (gold) standard, technical specifications, units, personnel, blinding</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>Methods for quantifying accuracy and uncertainty</td>
</tr>
</tbody>
</table>

The Study Population

- Readers must know whether the study excluded patients with conditions known to affect the way a test works, which could inflate diagnostic accuracy
- Tests may perform differently in primary, secondary, and tertiary care settings
- Performance may differ if the test is used for screening rather than confirmation

Test Methods

- Sufficient detail to allow other researchers to replicate the study
- Define categories of results before the study is done to avoid biasing test performance
- Subjective test results more likely interpreted as 'abnormal' in settings with a higher prevalence of the target condition (context bias)
- Knowledge of the reference test results can influence reading of the index test (and vice versa)

Statistical Methods

- Quantify the amount of statistical uncertainty around the observed value
- Poor reproducibility (of the index or reference test) adversely affects diagnostic accuracy
- For quantitative assays, report imprecision as the CV at two or more specified mean values near clinical decision points

STARD: Results

<table>
<thead>
<tr>
<th>RESULTS</th>
<th>Study population</th>
<th>Dates of study, population demographics, number included and excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test results</td>
<td>Time between tests, other diseases, cross-tabulation</td>
</tr>
<tr>
<td></td>
<td>Statistical results</td>
<td>Estimates of diagnostic accuracy and uncertainty, outliers, reproducibility</td>
</tr>
</tbody>
</table>

Flow Diagram of the Study

- Diagram representing the flow of the study with various nodes and connections indicating the process and outcomes.
Population Results

- Demographic and clinical features of the study population can affect measures of diagnostic accuracy (spectrum bias).
- Results of the index test should not influence the decision to order the reference standard test.
- Index and reference tests should be performed on the same patient at the same time.
- Test sensitivity is often higher in studies with a higher proportion of patients with advanced disease.

Cross Tabulation

<table>
<thead>
<tr>
<th>THYROID CYTOLOGY</th>
<th>Non-diagnostic</th>
<th>Normal</th>
<th>Suspect</th>
<th>Malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papillary cancer</td>
<td>12</td>
<td>30</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Medullary cancer</td>
<td>18</td>
<td>15</td>
<td>5</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>45</td>
<td>23</td>
<td>45</td>
</tr>
</tbody>
</table>

- Authors should present results in absolute numbers.
- Allows scientific colleagues to (re)calculate measures of diagnostic accuracy or perform alternative (meta)analyses.

Distributions of Continuous Results

Statistical Results

- Values presented should be taken as estimates.
- Many journals require or strongly encourage the use of confidence intervals as measures of precision.
- Uninterpretable, indeterminate, and intermediate test results pose a problem.
- Possible sources of heterogeneity in results should be explored (plan subgroup analyses before the start of the study).

STARD: Checklist

- First reported data.
- Practical uses of the assay.
- Settings in which the assay would be expected to perform well (vs settings not tested).
- Methodological shortcomings.

STARD: Summary

- One checklist may more likely be accepted by authors, peer reviewers, and journal editors.
- Published in several leading biomedical journals.
- Planned evaluations and updates.
- Applicability to veterinary research.
Conclusions

- Terminology and reporting standards are essential for effective communication and quality databases
- Terminology and reporting standards impact the discipline and profession
  - Improved quality of study design
  - Improved quality & value of published articles
  - Improved quality of healthcare