The Role of Reporting Guidelines in Promoting Reproducible Research

Presentation to the
NCI Board of Scientific Advisors

Lisa McShane, PhD
Biometric Research Branch, DCTD
National Cancer Institute

March 4, 2013
Propagation of Irreproducible Research

- Primary data generation
- Data analysis & derived results
- Results interpretation & reporting
- Results dissemination
REMARK: REporting guidelines for tumor MARKer prognostic studies


Recommended reporting elements to facilitate

• Evaluation of **appropriateness & quality** of study design, methods, and analysis
• Understanding of **context** in which conclusions apply
• **Reproducibility**
• **Comparisons** across studies, including formal meta-analyses
REMARK: Target Studies

• Studies relating marker values to clinical events (e.g., recurrence, death, response)
• NOT primarily aimed at biological discovery studies, but use encouraged to extent possible
  • Patients
  • Specimens
  • Assays
• NOT sufficient for studies developing multiplex classifiers/risk scores (e.g., derived from omics data), but applicable to studies assessing them
Purpose: To update the recommendations for the use of tumor marker tests in the prevention, screening, treatment, and surveillance of breast cancer.

“. . . primary literature is characterized by studies that included small patient numbers, that are retrospective, and that commonly perform multiple analyses until one reveals a statistically significant result. . .many tumor marker studies fail to include descriptions of how patients were treated or analyses of the marker in different treatment subgroups. The Update Committee hopes that adherence to . . . REMARK criteria will provide more informative data sets in the future.
REMARK Elements: Introduction

• State all marker(s) examined
• Study objectives
• Pre-specified hypotheses
Common Tumor Marker Study Design

What can we do with our marker on these 89 specimens?

- “Convenience” specimens
- Heterogeneous patient characteristics
- Treatments: Unknown, non-randomized, not standardized
- Insufficient sample size (underpowered)
- Uncertain specimen and data quality
REMARK Elements: Materials & Methods

• Patients
  • Inclusion/exclusion (e.g., stage, subtype), source, treatments

• Specimen characteristics
  • Format, collection, preservation, storage
  • See BRISQ criteria (Moore et al, *Cancer Cytopathology* 2011; 119:92-101)
REMARK Elements: Materials & Methods (cont.)

• Assay methods
  • Detailed protocol (reagents/kits), quantitation, scoring & reporting, reproducibility, blinding

Example: Systematic review (43 studies) of Ki67 in early breast cancer (Stuart-Harris et al, The Breast 2008; 17:323-334)
  • English publication, Jan. 1995 – Sept. 2004
  • ≥ 100 patients, OS or DFS endpoint

• Results
  • 7 different antibodies for IHC, single or combination
  • 19 different cutpoints, ranging from 0-30%
  • Significant between-study heterogeneity and publication bias
International Ki-67 Reproducibility Study (Nielsen et al, SABCS 2012 abstract)

Consecutive TMA sections, single assay batch

median: 10%

Lab E

Lab P

Lab A

Lab I

Lab U

Lab F

Lab M

Lab G

median: 28%
REMARK Elements: Materials & Methods (cont.)

• Study design
  • Case selection (e.g., random, case-control), clinical endpoints, variables considered, sample size

• Statistical analysis methods
  • Models, variable selection, handling of missing data, multiple testing adjustments, validations
Almost all articles on cancer prognostic markers report statistically significant results

Panayiotis A. Kyzas\textsuperscript{a}, Despina Denaxa-Kyza\textsuperscript{a}, John P.A. Ioannidis\textsuperscript{a,b,c,\star}

\textsuperscript{a}Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece
\textsuperscript{b}Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, Greece
\textsuperscript{c}Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Boston, USA

“If you torture the data long enough they will confess to anything.”

Source unknown
Statistical Analysis: Multiple Testing

- Multiple markers
- Multiple endpoints
- Multiple subgroups
- Multiple marker cutpoints
- Multiple models with multiple variables

Example: 8 subgroups defined by 3 binary factors

<table>
<thead>
<tr>
<th>Number of independent tests (α = 0.05 per test)</th>
<th>Probability observe ≥ 1 statistically significant (p&lt;0.05) result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>0.14</td>
</tr>
<tr>
<td>4</td>
<td>0.19</td>
</tr>
<tr>
<td>5</td>
<td>0.23</td>
</tr>
<tr>
<td>6</td>
<td>0.26</td>
</tr>
<tr>
<td>7</td>
<td>0.30</td>
</tr>
<tr>
<td>8</td>
<td>0.34</td>
</tr>
<tr>
<td>9</td>
<td>0.37</td>
</tr>
<tr>
<td>10</td>
<td>0.40</td>
</tr>
</tbody>
</table>
REMARC Elements: Results

• Data
  • Numbers of patients and events
  • Demographic characteristics
  • Standard prognostic variable distribution
  • Tumor marker distribution

• Analysis & presentation
  • Univariate analyses (marker vs. standard prognostic variables, marker vs. outcome)
  • Multivariable analyses (association of marker with outcome after adjustment for standard prognostic variables)
  • Measures of uncertainty for reported effect estimates
REMARK Elements: Results (cont.)

• Viewing in context of standard factors and treatments received

![Graphs showing 5-year survival rates for different groups of patients.](image)

- **All patients**
  - 5-yr Survival
    - POS 91%
    - NEG 63%

- **ER positive patients**
  - 5-yr Survival
    - POS 98%
    - NEG 65%

- **ER negative patients**
  - 5-yr Survival
    - POS 80%
    - NEG 60%
REMARK Elements: Discussion

• Interpretation in context of pre-specified hypotheses
• Relevance to other studies
• Limitations
• Future research
• Clinical value
REMARK Status & Future


• Plans for “before vs. after” comparisons of reporting

“Studies of ‘prognostic’ markers of no real future clinical utility and single biomarker studies will not be considered. Reports of studies into prognostic markers should be prospective and have a clear view of the practical clinical applications of the results. Retrospective analysis of biomarkers can be considered, if done within the framework of data collected from a prospective trial, with appropriate statistics and with multivariate analysis that includes established predictive/prognostic markers. Reports of prognostic tumor marker studies should follow the REMARK guidelines (available from www.equator-network.org).”

J. B. Vermorken
Editor-in-Chief
Concluding Remarks

• Poor study reporting is a significant impediment to achieving reproducible research
• Reporting will improve only with effort from all stakeholders
• Complete & transparent reporting is more fair
• Effort spent on good reporting is a smaller burden than time, effort and resources wasted on false leads