Toward a Common Goal: Publication and Meta-Analysis

Paul E Kinahan, PhD, FIEEE
University of Washington

Summary

• Quantitative imaging biomarkers have a key role in realizing the potential of precision medicine
• Evidence-based medicine and cost containment mandate use of pooled data in meta-analysis
• A large fraction of scientific research is not useable or reproducible, often due to poor reporting
• There are reporting standards (e.g. STARD) that can be adopted for quantitative imaging biomarkers
• Efforts are underway for draft PET guidelines
• Means of adoption and sustainability require careful thought

Quantitative imaging can characterize hallmarks of disease and response to therapy

Response to therapy of liver met GIST

PET/CT

CT

PET

SUV: 5 to 1.8

1 wk imatinib therapy

Castell and Cook, British J Cancer 2008
Quantitative Imaging Biomarkers

- Measurements of anatomical, physiological, and biochemical characteristics of the body through medical imaging
- Are becoming increasingly used in clinical research for drug and medical device development and clinical decision-making.

Imaging biomarker examples

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor volume</td>
<td>CT, MRI</td>
</tr>
<tr>
<td>β-amyloid</td>
<td>PET</td>
</tr>
<tr>
<td>Tumor proliferation</td>
<td>PET</td>
</tr>
<tr>
<td>Bone mineral density</td>
<td>DXA, CT</td>
</tr>
<tr>
<td>Receptor occupancy</td>
<td>PET</td>
</tr>
<tr>
<td>Plaque composition</td>
<td>US, IR, MRI, PET</td>
</tr>
</tbody>
</table>

Jeff Evelhoch, Merck Research Laboratories

Biomarkers are essential to the success of Precision medicine

- Precision medicine
  = Predictive + Prognostic + Personalized
- Biomarkers can be serum, tissue, or imaging
- Predictive biomarkers
  - For therapy selection
- Early response biomarkers
  - Reject ineffective therapies as soon as possible
- Prognostic biomarkers
  - Inform about an event independent of specific treatment

*Modified from Hedvig Hricak*

The need for evidence

Over the past decade, the cost of health care in the United States has increased at a rate greater than twice the general rate of inflation

During the past decade, imaging services and their costs have grown at about twice the rate of other technologies in health care (eg, laboratory procedures and pharmaceuticals)

*Hendee, Radiology 2010*

Pressures on reimbursement

- Using the ACR appropriateness criteria and evidence-based medical literature, CareCore National denies 15-20% of the four million imaging requests from physicians each year
  - Sura & Ho, J Clin Imaging Sci 2011
- 2010 ACR survey: If imaging centers were cut 50%, 29% would drop out of Medicare, 41% would limit access to Medicare beneficiaries, 46% would close imaging centers, 75% would reduce the number of staff and reduce overhead, and 78% would forgo technology upgrades
Problems with reproducibility

- Scientists at Amgen tried to replicate 53 studies considered landmarks in the basic science of cancer ... they were able to reproduce the original results in just six.

- Prinz and colleagues at Bayer HealthCare reported that they had successfully reproduced the published results in just a quarter of 67 seminal studies.

The Economist, October 2013

Waste in research  
Chalmers, Lancet 2009

<table>
<thead>
<tr>
<th>Questions relevant to clinicians and patients?</th>
<th>Appropriate design and methods?</th>
<th>Accessible full publication?</th>
<th>Unbiased and usable report?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low priority questions addressed</td>
<td>Over 50% of studies designed without reference to systematic review of existing evidence</td>
<td>Over 50% of studies never published in full</td>
<td>Over 30% of trial interventions not sufficiently described</td>
</tr>
<tr>
<td>Important outcomes not assessed</td>
<td>Over 50% of studies fail to take adequate steps to reduce bias – eg, uncontrolled treatment allocation</td>
<td>Over 50% of studies never published in full</td>
<td>Over 50% of planned study outcomes not reported</td>
</tr>
<tr>
<td>Clinicians and patients not involved in setting research agenda</td>
<td>Over 50% of studies fail to take adequate steps to reduce bias – eg, uncontrolled treatment allocation</td>
<td>Based under-reporting of studies with disappointing results</td>
<td>Most new research not interpreted in the context of systematic assessment of other relevant evidence</td>
</tr>
</tbody>
</table>

Research waste
Reducing waste from incomplete or unusable reports of biomedical research

Glasziou, Lancet 2014

“Findings from a 2009 article suggested that at least 50% of research reports were sufficiently poor or incompletely as to make them unusable.”

“Unless research is adequately reported, the time and resources invested in the conduct of research is wasted.”

<table>
<thead>
<tr>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trials: missing effect size and confidence interval (38%); no mention of adverse effects (49%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trials: 40-89% inadequate treatment descriptions</td>
</tr>
<tr>
<td>RMB studies: 23% missing number of trials and duration</td>
</tr>
<tr>
<td>Survey questions: 65% missing survey or comparisons</td>
</tr>
<tr>
<td>Figures: 31% graphs ambiguous</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical trials: outcomes missing: 52% efficacy and 65% harm outcomes per trial incompletely reported</td>
</tr>
<tr>
<td>Animal studies: number of animals and raw data missing (64%); age and weight missing (24%)</td>
</tr>
<tr>
<td>Diagnostic studies: missing age and sex (67%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trials: no systematic attempt to set new results in context of previous trials (50%)</td>
</tr>
</tbody>
</table>

Data

| Trials: most data never made available, author-held data lost at about 7% per year |

Figures and tables are often incomplete or un-interpretable

- 31% of all graphs published in JAMA in 1999-2000 could not be interpreted unambiguously


The Lack of Evidence for PET or PET/CT Surveillance of Patients with Treated Lymphoma, Colorectal Cancer, and Head and Neck Cancer: A Systematic Review

Kamal Patel, Nina Hadar, Joavaher Lee, Barry A. Siegel, Bruce E. Hilker, and Joseph Lau


"Studies were generally of poor quality, with more than half being retrospective ... we were unable to use all available data because test accuracy was not consistently defined and reporting was incomplete."
Accompanying commentary by Eary and Krohn

- "The most disturbing message from this analysis is that the data available to address significant questions about surveillance imaging using 18F-FDG PET were inadequate for their analysis."
- "An enormous amount of effort was put into conducting, analyzing, and reporting the 1,813 published studies that were reviewed for inclusion in the analysis of Patel et al., but nearly all of them were judged inadequate for this analysis."
- "Why is it that our literature includes so many publications of weak quality?"
Some potential solutions for the general problem

Unbiased and usable report
Poor awareness and use by authors and editors of reporting guidelines
• Increase author and journal awareness of and training in reporting guidelines, such as CONSORT and STARD statements (http://www.equator-network.org)
Many journal reviewers focus on expert judgments about contribution to knowledge, rather than methods and usability
• Supplement peer review of studies with review by methodologists and end users
Space restrictions in journals prevent publication of details of interventions and tests
• Support free access repositories—separate from any publications—so that clinicians and researchers have details of the treatments, test, or instruments studied

Chalmers, Lancet 2009

Some Reporting Standards
STARD - STAndards for the Reporting of Diagnostic accuracy studies

REMARK - REporting recommendations for tumour MARKer prognostic studies

CONSORT - Reporting guideline for Parallel group randomised trials

STROBE – Strengthening the Reporting of Observational Studies in Epidemiology

PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses

BRISQ - Biospecimen Reporting for Improved Study Quality

“At the end of the talk, there were one or two questions about the role of medical writers, and on the tricky subject of authorship, but then an almost angry outburst from one participant who said ‘Why haven’t we been told about these guidelines before?’. I couldn’t answer that.”
# STARD Statement

**STAsndards for the Reporting of Diagnostic accuracy studies**

**Objective of the STARD initiative**

The objective of the STARD initiative is to improve the accuracy and completeness of reporting of studies of diagnostic accuracy, to allow readers to assess the potential for bias in the study (internal validity) and to evaluate its generalizability (external validity).

The STARD statement consists of a checklist of 25 items and recommends the use of a flow diagram which describe the design of the study and the flow of patients.

**News**

April 2006

More than 200 biomedical journals encourage the use of the STARD statement in their instructions for authors.

---

## STARD checklist for reporting of studies of diagnostic accuracy

*(version January 2003)*

<table>
<thead>
<tr>
<th>Section and Topic</th>
<th>Item</th>
<th>Reference</th>
</tr>
</thead>
</table>
| **Introduction** | 1. Identify the study as a study of diagnostic accuracy (sensibility, specificity, or likelihood ratios). | [STARD Statement](https://www.equator-network.org/reporting-guidelines/stard/)
| **Participants** | 2. Specify the sample that was enrolled in the study before data collection, if done. | [STARD Statement](https://www.equator-network.org/reporting-guidelines/stard/)
| | 3. The study population: The inclusion and exclusion criteria, setting and locations where data were collected. | [STARD Statement](https://www.equator-network.org/reporting-guidelines/stard/)
| | 4. Participant recruitment: If non-randomized based on presenting complain, results from previous tests, or the fact that the participants had received the index test or the reference standard? | [STARD Statement](https://www.equator-network.org/reporting-guidelines/stard/)
| | 5. Participant eligibility: Describe the proportion of the sample who were included and the reasons for exclusion. | [STARD Statement](https://www.equator-network.org/reporting-guidelines/stard/)
| | 6. Data collection: The data collection protocol, including the test and reference standard data, were performed (perspective study) or after a prospective design? | [STARD Statement](https://www.equator-network.org/reporting-guidelines/stard/)
| | 10. The number, testing, and reporting of the patients including all patients. | [STARD Statement](https://www.equator-network.org/reporting-guidelines/stard/)
| | 11. Whether or not the results of the index test and reference standard were blinded, if done. | [STARD Statement](https://www.equator-network.org/reporting-guidelines/stard/)
| **Discussion** | 25 Discuss the clinical applicability of the study findings. | [STARD Statement](https://www.equator-network.org/reporting-guidelines/stard/)

---

7/21/2014
A virtuous circle

Quantitative Imaging Biomarker Reporting Working Group

- Outgrowth/merging of several activities
  - Development of QIBA Profiles
  - QIBA metrology series for QIBs
  - QIN Clinical trial design group
  - UK meta-analysis of PET-FLT v. KI-67 correlations (Chalkidou, E J Ca 2012)
- Decision to focus on PET imaging first
  - Address reporting
  - Does not address study quality
Draft List of FDG-PET Study Characteristics and Results to Report

Purpose: Summarize study characteristics to be reported for the study to be repeated and/or included as part of a meta-analysis

1. Information needed for repeating a study and/or to determine inclusion/exclusion in the meta-analysis:
   - Objective of the study
   - Study design
   - Whether study was retrospective versus prospective as per QUADAS
   - Clinical or experimental setting
   - Target clinical population
   - Where the study was conducted and whether it was single or multi-site
   - Key aspects of the image acquisition and generation: e.g. uptake time
   - What performance metric or summary statistic was computed during study
   - Amount of information persons performing the analysis are provided with.
   - If a reference test is also reported, it should carry equal weight to the quantitative imaging biomarker

2. Information needed for the meta-analysis and meta-regression:
   - Number of subjects in the study and total number of lesions
   - How multiple lesions were handled
   - The summary statistic of interest, with confidence intervals or standard errors.
   - For a test-re-test study, both sets of data should be reported.
   - Study characteristics that may be associated with the value of the summary statistic or performance metric of interest so that in the future, the study may be a part of a meta-regression.
   - Individual patient data should be supplied as supplemental material
**Properties of adequate reporting**

- Reproduce → Confirmation
- Assess → Quality
- Combine → Meta-analysis for evidence-based practice

**How to implement?**
*(beyond guidelines)*

- Could be proposed as modules to, e.g., STARD
- Many, or most, imaging journals have limited resources for checking compliance
- Reviewers may choose to ignore guidelines included as review criteria
- Authors could be asked to submit a checklist of compliance with guidelines
  - Not possible for all studies to meet all guidelines
  - 100% self-reported compliance cannot be mandatory

**Summary**

- Quantitative imaging biomarkers have a key role in realizing the potential of precision medicine
- Evidence-based medicine and cost containment mandate use of pooled data in meta-analysis
- A large fraction of scientific research is not useable or reproducible, often due to poor reporting
- There are reporting standards (e.g. STARD) that can be adopted for quantitative imaging biomarkers
- Efforts are underway for draft PET guidelines
- Means of adoption and sustainability require careful thought
Acknowledgements

• Members of the QIB Reporting Standards Working Group:
  Sally F. Barrington, Anastasia Chalkidou, Dominique Delbeke, Constantine Gatsonis, Otto Hoekstra, Erich Huang, Paul Kinahan, PhD (co-chair), Paul Marsden (co-chair), Lisa McShane, Dan Sullivan, Richard L. Wahl
• Members of QIBA and RSNA staff
• Members of QIN
• U01-CA148131

Quantitative Imaging Definitions

• A biomarker is an objectively measured indicator of biological/pathobiological process or pharmacologic response to treatment
• Qualified biomarker: A disease-related biomarker linked by graded evidence to biological and clinical endpoints and dependent upon the intended use
• Imaging biomarker: a number, set of numbers, or classification derived from an image (in general imaging biomarkers are not surrogate endpoints)
• Validated assay: An assay (i.e. quantitative imaging) that has documented performance characteristics showing suitability for the intended applications
  – needed for a qualified biomarker