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

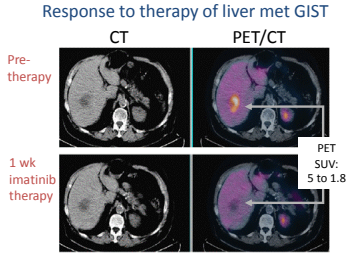
Quantitative imaging can characterize hallmarks of disease and response to therapy

short term drivers

- Clinical research, Clinical trials, and Drug discovery
- New molecular diagnostic agents
- Assessing individual response to therapy
- New clinical imaging procedures

increasing volume

Response to therapy of liver met GIST



Castell and Cook, British J Cancer 2008

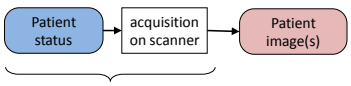
Quantitative Imaging Biomarkers

- 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.

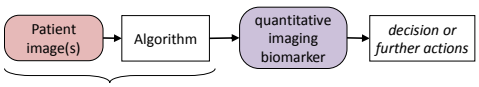
Imaging biomarker examples

Biomarker	Assay
Tumor volume	CT, MRI
β-amyloid	PET
Tumor proliferation	PET
Bone mineral density	DXA, CT
Receptor occupancy	PET
Plaque composition	US, IR, MRI, PET

Jeff Evelhoch, Merck Research Laboratories



Determines bias and variance of image values



Determines bias and variance of quantitative imaging biomarker

Obuchowski, Stat Methods Med Res, 2014

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, Radilogy 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

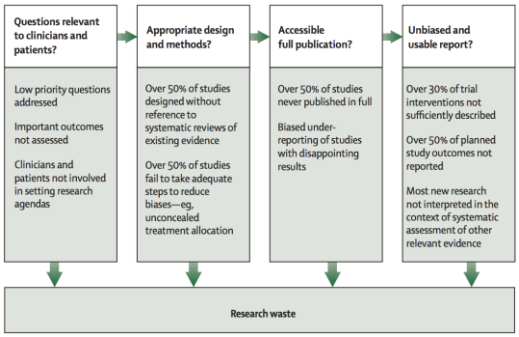
Problems with reproducibility

- 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

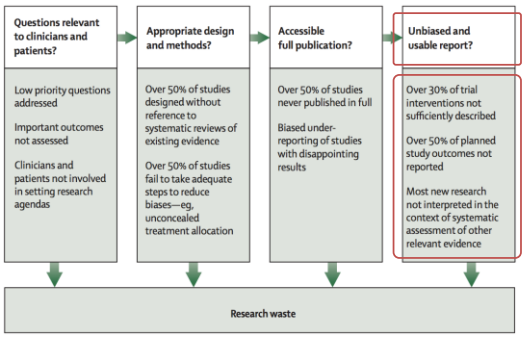
Waste in research

Chalmers, Lancet 2009



Waste in research

Chalmers, Lancet 2009



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"

Abstract
Trials: missing effect size and confidence interval (38%); no mention of adverse effects (49%)[7]

Methods
Trials: 40–89% inadequate treatment descriptions[11, 13]
fMRI studies: 33% missing number of trials and durations[3]
Survey questions: 65% missing survey or core questions[15]
Figures: 31% graphs ambiguous[45]

Results
Clinical trials: outcomes missing: 50% efficacy and 65% harm outcomes per trial incompletely reported[6]
Animal studies: number of animals and raw data missing[17] (54%, 92%); age and weight missing (24%)
Diagnostic studies: missing age and sex (40%)[35]

Discussion
Trials: no systematic attempt to set new results in context of previous trials (50%)[69]

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

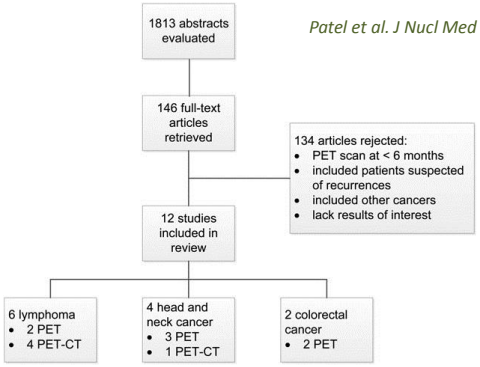
Cooper et al. Ann Emerg Med 2002

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[1], Nira Hadar[1], Jounghee Lee[1], Barry A. Siegel[2], Bruce E. Hillner[3], and Joseph Lau[1]

J Nucl Med 2013; 54:1518–1527

"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."



Patel et al. J Nucl Med 2013

Patel JNM 2013

Study	Prospective study design?	Clear eligibility criteria?	Selection bias likely?	Index and reference tests adequately described?	Masked outcome assessment?	Clear reporting with no discrepancies?	Overall grade
PET/CT: lymphoma							
Crocchiolo (22)	No	Yes	No	Yes	No	Yes	B
El-Galaly (23)	No	Yes	No	Yes	No	Yes	B
Lee (21)	No	Yes	Yes	No	No	Yes	C
Rhodes (24)	No	Yes	No	Yes	No	Yes	B
PET/CT: head and neck cancer							
Abgral (25)	Yes	Yes	No	Yes	Yes	Yes	A
PET: lymphoma							
Hosein (27)	No	Yes	Yes	Yes	No	Yes	C
Zinzani (26)	Yes	Yes	No	Yes	No	No	B
PET: head and neck cancer							
Lowe (29)	Yes	Yes	Yes	Yes	Yes	No	B
Péré (30)	Yes	Yes	Yes	Yes	Yes	Yes	B
Salaun (31)	No	Yes	No	Yes	Yes	Yes	B
PET: colorectal cancer							
Selvaggi (28)	No	Yes	Yes	Yes	Yes	Yes	C
Sobhani (20)	Yes	Yes	No	Yes	Yes	No	B

Grade	Criteria
A	Adhered to recognized standards for diagnostic test studies Clear descriptions of design, population, test, reference standard, outcomes No major reporting omissions and no obvious source of bias
B	Some deficiencies, but considered unlikely to result in a major bias
C	Serious design or reporting deficiencies

146 papers
⇩
12 useable
⇩
1 grade A data

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 reviews 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

CONSORT=Consolidated Standards of Reporting Trials. STARD=Standards for the Reporting of Diagnostic Accuracy Studies.

Chalmers, Lancet 2009

Some Reporting Standards

STARD - STAndards for the Reporting of Diagnostic accuracy studies
- Bossuyt et al. Clin Chem 2003;49:1–6

REMARK - REporting recommendations for tumour MARKer prognostic studies
- McShane et al. Br J Cancer 93: 387–391.

CONSORT - Reporting guideline for Parallel group randomised trials
- Moher et al. JAMA 2010:285: 1987–1991

STROBE – Strengthening the Reporting of Observational Studies in Epidemiology
- Vandenbroucke et al. . PLoS Med 4: e297. 2007

PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- Moher et al. PLoS Med 6: e1000100, 2009

BRISQ - Biospecimen Reporting for Improved Study Quality
- Moore et al. Cancer Cytopathol. 2011;119(2):92-101.

Latest guest blogger

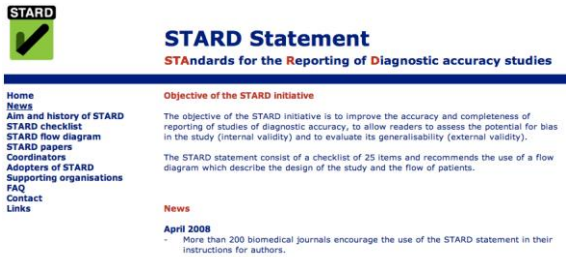
Why aren't researchers told about reporting guidelines?

July 3, 2014

Liz Wager reflects on a recent talk that she gave to postgraduate students at a careers day for biomedical researchers considering working in medical communications **Read More**

"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."

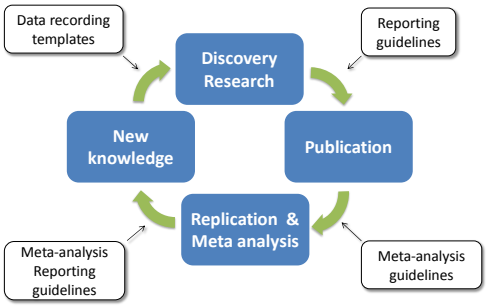


STARD checklist for reporting of studies of diagnostic accuracy
(version January 2003)

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	
METHODS			
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	
Test methods	7	The reference standard and its rationale.	
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	

Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	
	13	Methods for calculating test reproducibility, if done.	
RESULTS			
Participants	14	When study was performed, including beginning and end dates of recruitment.	
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	
Test results	17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	
	20	Any adverse events from performing the index tests or the reference standard.	
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	
	22	How indeterminate results, missing data and outliers of the index tests were handled.	
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	
	24	Estimates of test reproducibility, if done.	
DISCUSSION	25	Discuss the clinical applicability of the study findings.	

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

Draft List of FDG-PET Study Characteristics and Results to Report

1. Information needed for repeating a study and/or to determine inclusion/exclusion in the meta-analysis:

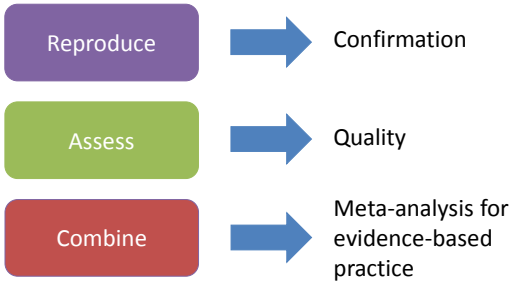
- Objective of the study
- Study design
- Whether study was retrospective versus prospective as per QUADAS
- Clinical or experimental setting
- Target clinical population
- Where the study was conducted and whether it was single or multi-site
- Key aspects of the image acquisition and generation: e.g. 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

Draft List of FDG-PET Study Characteristics and Results to Report

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



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

Biomarkers Definitions Working Group. Clin Pharmacol Ther 2001;69(3):89–95.