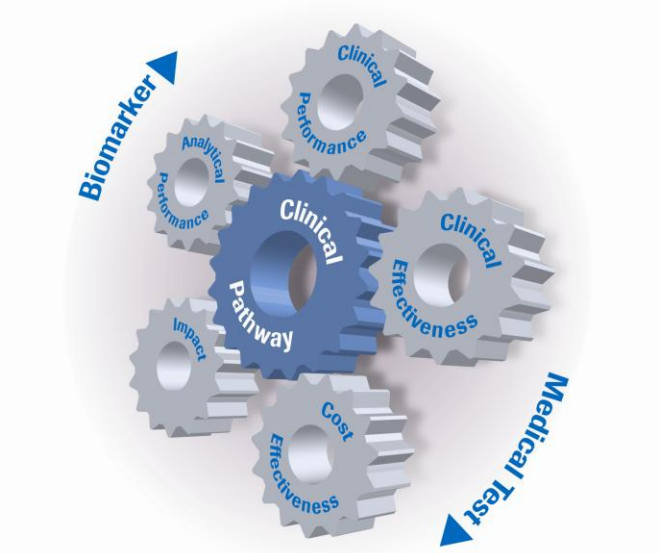
Analytical Performance Specifications based on Outcome Studies: Is it possible?

Andrea Rita Horvath
Test Evaluation Working Group of EFLM
www.efcclm.eu

SEALS Department of Clinical Chemistry,
Prince of Wales Hospital, Sydney
rita.horvath@sesiahs.health.nsw.gov.au
School of Public Health, University of Sydney

SOUTH EASTERN AREA LABORATORY SERVICES
NSW GOVERNMENT Health Pathology
EFLM EUROPEAN FEDERATION OF CLINICAL CHEMISTRY AND LABORATORY MEDICINE

Analytical Performance in the The Test Evaluation Cycle



EFLM WG-TE

Definitions

- **Analytical performance:** the ability of an ***assay*** to conform to predefined technical specifications and to correctly detect or measure a particular analyte/measurand.
- **Clinical performance:** the ability of a **biomarker** to conform to predefined clinical specifications in detecting patients with a particular clinical condition or in a physiological state
- **Clinical effectiveness:** the ability of a ***test*** to improve outcomes relevant to the individual patient or patient population.

EFLM WG-TE

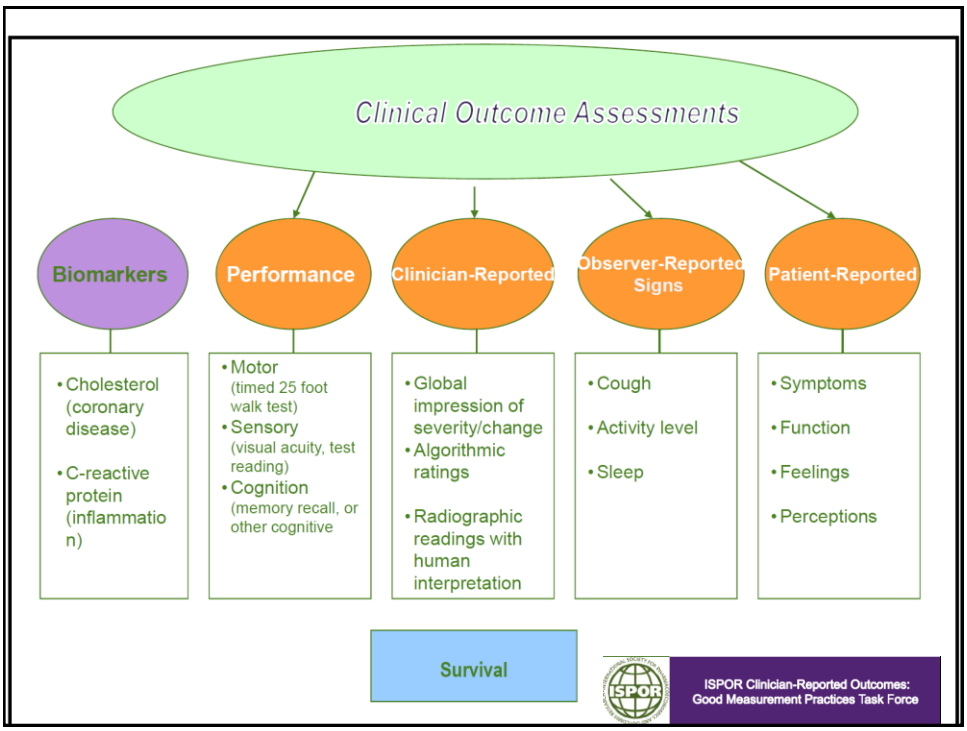
Definitions

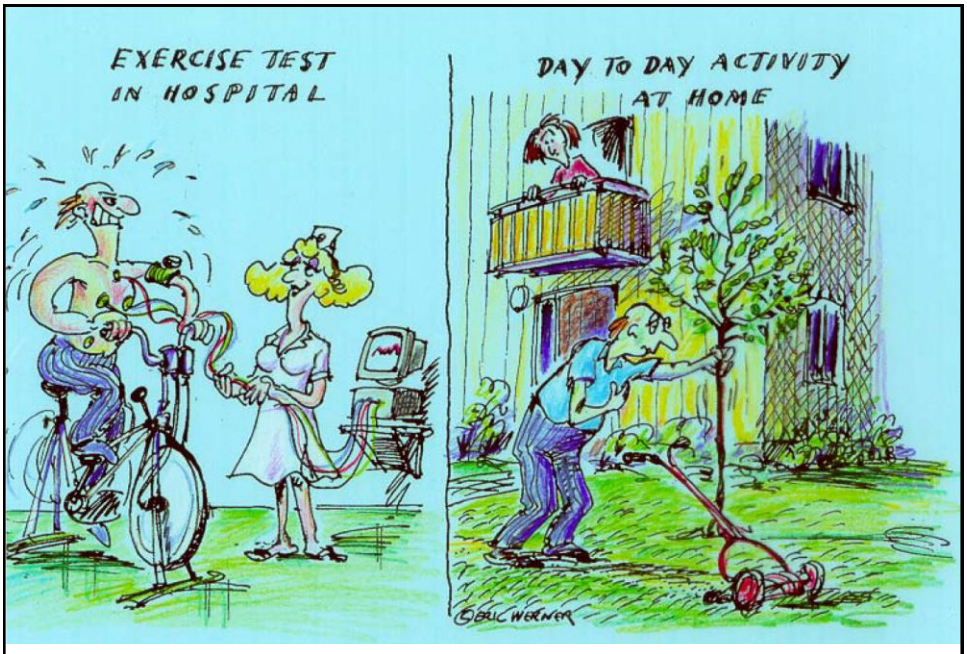
- **Analytical performance specifications:** Criteria that specify (in numerical terms) the quality required for analytical performance in order to deliver laboratory test information that would satisfy *clinical needs* for improving *health outcomes*.
- **Clinical needs:** refers to any desirable testing or treatment component of a clinical pathway where existing care could improve in order to achieve better *health outcomes* for patients.

EFLM WG-TE

Definitions

- **Health outcomes: are a change in the health status or well-being of an individual, group or population which is attributable to a (series of) planned intervention(s).**
  - **Whose perspective** – patient, population, health care staff, policy makers
  - **Type of outcome** – subjective (QoL), objective (all-cause mortality
  - **Timing** – short-term or long-term
  - **Composite endpoint**
  - **Surrogate or intermediate outcome** (HbA1c, cholesterol)
- **[Organisational/ economic outcomes]**





Indirect evidence of benefit
(easy to measure)

Direct evidence of benefit
(not easy to measure)

Key principles

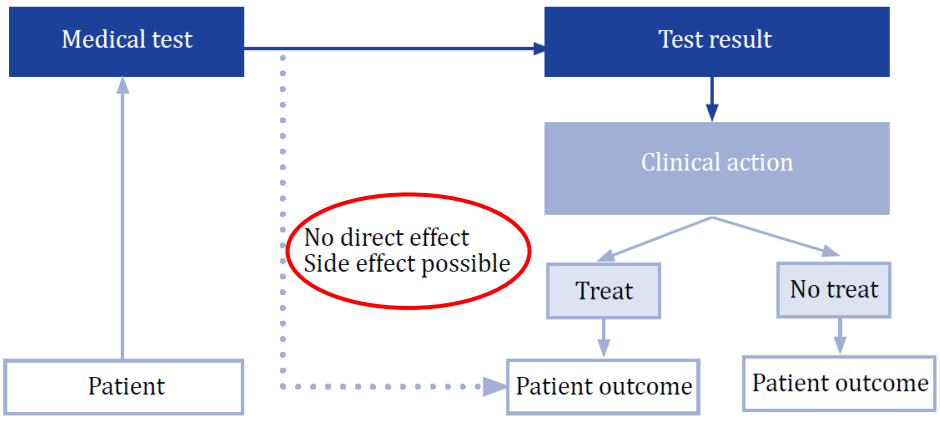
- **Few tests have definitive role in managing a condition – thus their impact on health outcomes is varied**
- **Most laboratory tests are used for multiple purposes and in combination with other laboratory or other tests**
- **The link between testing and health outcomes is indirect and is dictated by the clinical pathway, and**
- **the *purpose and role of the test* in the clinical pathway.**

Definitions

- **Test purpose:** **describes the intended use of the test and how the test information will be used to improve clinical outcomes**
  - hs-Troponin for diagnosing ACS
  - hs-Troponin as a prognostic marker of cardiovascular disease
  - HbA1c for diagnosing diabetes mellitus
  - HbA1c for monitoring test to assess diabetes control
- **Test role:** **how the test will be positioned to alter the existing clinical pathways in a specific condition or target population**
  - **Triage:** hs-Troponin to triage patients with ACS
  - **Replacement:** Troponin to replace CK-MB in diagnosing ACS
  - **Add-on:** BNP added to hs-Troponin testing to assess prognosis of CVD

From testing to outcomes

Example of non-therapeutic devices:
relation between tests and patient outcomes



KNAW (2014). Evaluation of new technology in health care. In need of guidance for relevant evidence. Amsterdam, KNAW (www.knaw.nl).

Indirect linkage

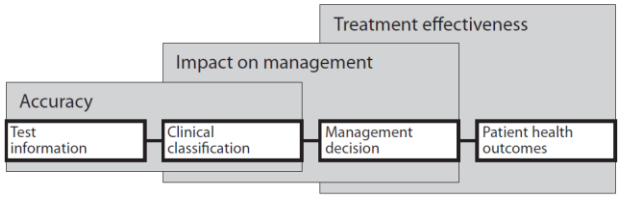


Figure 1 Test-treatment pathway showing Accuracy, Impact on management and Treatment effectiveness as determinants of health outcomes. Adapted from Staub et al. [9]

Diagnostic or prognostic accuracy and classification of the condition are not 'true' health outcomes.

Staub *et al. BMC Medical Research Methodology* 2012, **12**:12
http://www.biomedcentral.com/1471-2288/12/12

Staub *et al. BMC Medical Research Methodology* 2012, **12**:12
http://www.biomedcentral.com/1471-2288/12/12

BMC Medical Research Methodology

CORRESPONDENCE Open Access

Using patient management as a surrogate for patient health outcomes in diagnostic test evaluation

Lukas P Staub[1*], Sarah J Lord[1], R John Simes[1], Suzanne Dyer[1], Nehmat Houssami[2], Robert YM Chen[3] and Les Irwig[2]

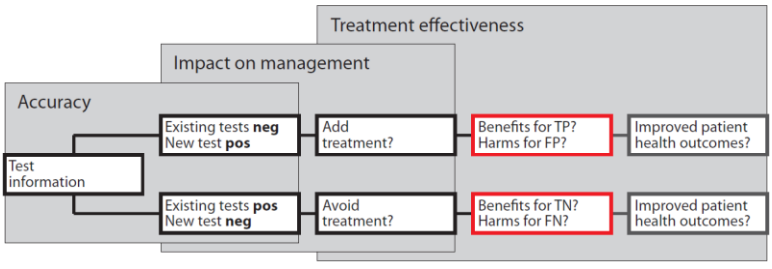


Figure 2 Identifying critical assumptions that changes in patient management improve patient health outcomes. Abbreviations: pos = positive, neg = negative, TP = true positive, FP = false positive, TN = true negative, FN = false negative

APS based on clinical needs are often defined in terms of allowable misclassification rates

Table. Recommended analytical performance goals for cardiac troponin measurement for definition of the limit of quantitation of assays.

Quality level	Imprecision goal (as CV)			Bias goal[a]
	Outcome-based	Biological variability[a]	Expert opinion	
Minimum	<13%[b]	<7.3%	<20%	±21.6 %
Desirable	<10%[c]	<4.9%	<10%	±14.4 %
Optimum	<6%[d]	<2.4%	–	±7.2 %

[a] Calculated according to Fraser CG, Hyltoft Petersen P, Libeer JC, Ricos C. Proposal for setting generally applicable quality goals solely based on biology. Ann Clin Biochem 1997;34:8-12.
[b] Assuming a diagnostic misclassification of 1.8%, [c] 1.0%, and [d] 0.5%.

Although the definition of analytical performance goals for cTnI and cTnT measurements is still under discussion, a total CV <10% together with an assay bias within ±15% may reasonably represent a good compromise for minimum requirements. This is consistent with the minimum total error goal for serum cTn measurement estimated at ~33%

Panteghini M, Troponin monograph, Chapter 8, CBR 2012

Examples for diagnostic or prognostic misclassification driven APS

Test	APS	Origin of APS	Reference
Hs-Troponin	CVa≤10% at the 99th percentile and able to detect Tn in at least 50% of the reference population	Diagnostic and prognostic accuracy	NICE2014
Glucose (plasma)	CVa≤2.9%, Bias≤2.2%, TE ≤6.9%	Biological variation	NACB 2012
Glucose (POCT)	TE for 95% of samples ≤15% at glucose conc.≥5.6 mmol/l (100 mg/dl) and to ≤0.8 mmol/l (15mg/dl) at glucose concentrations <5.6 mmol/l (100 mg/dl). Lower desirable TE in tight glucose-control protocols to avoid hypoglycemia	Outcome simulation - impact on insulin dosing errors and hypo- , hyperglycaemia	Clin Chem 2010;56(7):1091-7 Clin Chem 2014;60(4):644-50
HbA1c	Intralaboratory CV <2% Interlaboratory CV <3.5%	Biological variation	NACB 2012
Cholesterol	CVa ≤ 3.0%, Bias≤ 3.0%, TE≤ 8.9%	Diagnostic accuracy	NCEP/CRMLN 2004

How analytical performance specifications can be developed using outcomes data ?

Outcome studies

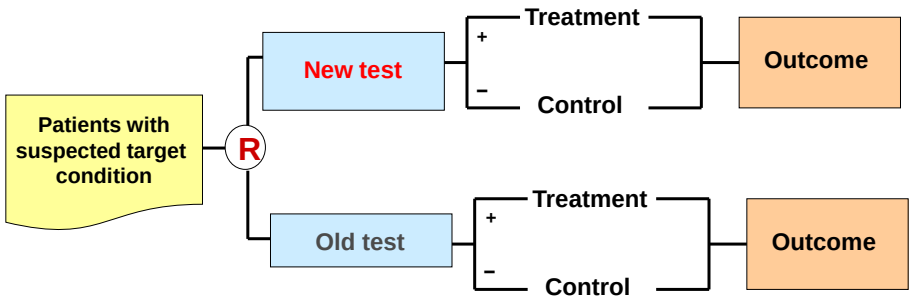
1. Assess the impact of analytical performance of the test on

- **clinical outcomes** *(direct)*
- **the probability of clinical outcomes - simulation studies** *(indirect)*

2. Survey of clinicians´ and/or experts' opinion – investigating the impact of the analytical performance of the test on medical decisions and subsequent patient management as intermediate to patient health outcomes *(indirect)*

RCT design to assess impact of analytical performance on outcome

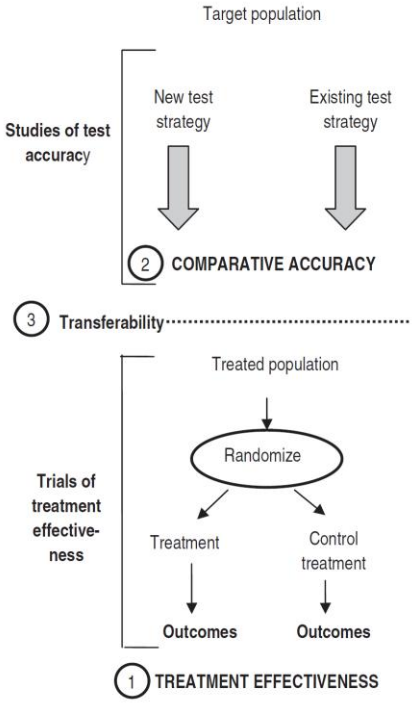
Do patients who undergo the new test with more advanced analytical performance fare better (in terms of health outcomes) than those who have the old test?



What to do when RCTs are not available or possible?

As a start...

- **Define the 'evidentiary reference' for analytical performance: i.e. the capabilities of the existing analytical test which was used to report estimates of test accuracy, decision thresholds and treatment effectiveness** (NB: only 'state of the art' at this stage)
- **Specify analytical performance at the relevant clinical decision threshold**
- **Consider the impact of variations in analytical performance on health outcomes and define :**
  - the relevant intended and unintended outcomes
  - the mechanisms and time frames in which outcomes may occur
  - existing test-treatment pathway for that indication,
  - proposed purpose, position and role of the test in the pathway
  - the key clinical decisions and actions the test will inform
  - and their potential linkage to health outcomes



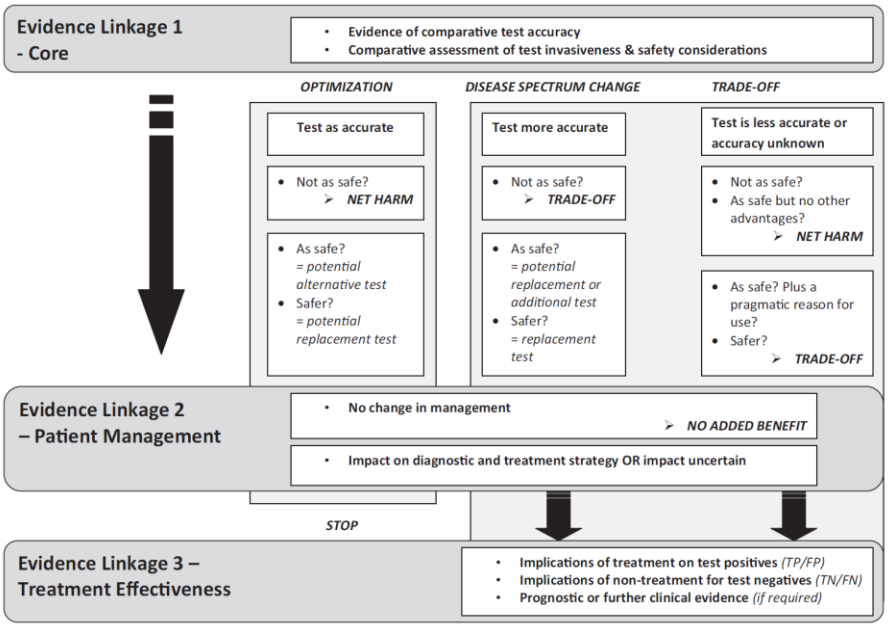
Linked evidence transferability assumption

Linking comparative diagnostic accuracy to treatment effectiveness data would be a sufficient proxy if

- **the population receiving the new test is to all intents and purposes the same that would receive treatment for the condition**
- **there is good evidence that treatment positively impacts on the health outcomes in this population.**

Lord S et al. 2005

Decision framework for the linked evidence approach



International Journal of Technology Assessment in Health Care, 2013;29(3):343–350.

Indirect or linked evidence approach

- an alternative when direct trial evidence of the clinical effectiveness of a test is not available, or is inadequate for decision making
- valuable specifically for tests that are modifications of an existing test
- validity depends on how well the 'intermediate' outcomes were proven to be linked and able to predict the relevant long-term health outcomes
- insufficient if the patient spectrum identified by the new version of the test is very different
- sequential linkages of evidence will increase the uncertainty of transferability between each linkage
- analytic frameworks or decision trees and flow charts enhance transparency when reviewing medical test performance

Modelling

- To model the clinical outcomes of misclassification requires clinical evidence about the consequences for patients.
- Where clinical evidence about these consequences is not available, the model estimates will be based on *assumptions* drawn from what evidence there is about disease prognosis, treatment benefits, harms etc.
- These assumptions will need to be tested.
- The model can only be as good as your assumptions are

Clinical Chemistry 56:7
1091–1097 (2010)

Glucose Meter Performance Criteria for Tight Glycemic Control Estimated by Simulation Modeling

Brad S. Karon,[1] James C. Boyd,[2] and George G. Klee[1*]

Table 1. Frequency of insulin dosing errors as a function of error condition for 29 920 000 simulated glucose values using the gaussian error model.

Error condition	10% error, %	15% error, %	20% error, %
No change	71.4	58.7	48.8
1-category	28.4	39.3	44.8
2-category	0.2	2.0	6.1
≥3-category	0.0	0.02	0.3

- **Glucose meters with TEa=15% are unlikely to produce large (3-category) insulin dosing errors**
- **Increasing performance to 10% TEa should reduce the frequency of 2-category insulin dosing errors**
- **Additional studies are necessary to determine the clinical impact of such errors**

Relationship between the frequency of hypoglycemia and the imprecision of glucose measurements

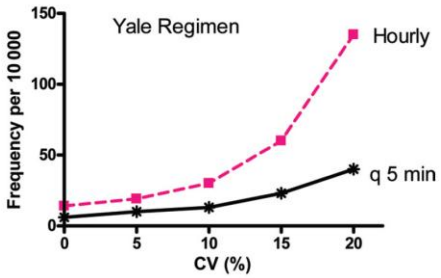


Fig. 1. Effects of imprecision, in the absence of bias, on the frequency of hypoglycemia in modeled patients.

The frequency of true glucose concentrations <60 mg/dL is expressed as the number of hypoglycemic results at the 1-h time points divided by the number of hourly measurements (10 000) in the 100 patients modeled for each CV. q 5 min, 5-min intervals.

- **higher measurement imprecision increased the rates of hypoglycemia and hyperglycemia**
- **The adverse effects of measurement imprecision were lower at the higher measurement frequency.**
- **Quality specifications for glucose meters are not transferable to continuous glucose monitoring**

Clin Chem 2014;60(4):644–650

Key messages

- Setting APS based on outcome data is complex but not impossible
- The link between testing and health outcomes is indirect and is dictated by the clinical pathway
- Mapping the pathway and clear definition of outcomes is essential
- Diagnostic or prognostic accuracy is an insufficient proxy outcome measure

Key messages

- Direct evidence for APS would be ideal but under specific circumstances a linked evidence approach can be used and often is sufficient for regulatory approval of a new biomarker
- APS could be different for different test applications, but if a test is used for multiple purposes the strictest APS should take precedence
- APS should be commensurate with the impact of the laboratory test on subsequent medical decisions and actions

Test Evaluation Working Group

Patrick Bossuyt	University of Amsterdam
Christa Cobbaert	University of Leiden
Christopher Ebert	Roche Diagnostics
Sally Lord	University of Sydney
Lieselotte Lennartz	Abbott Diagnostics
Phil Monaghan	Birmingham, UK
Sverre Sandberg	University of Bergen
Andrew StJohn	AACB, Australia
Wilma Verhagen-Kamerbeek	Roche Diagnostics

The working group acknowledeges the support of
Roche Diagnostics and EFLM

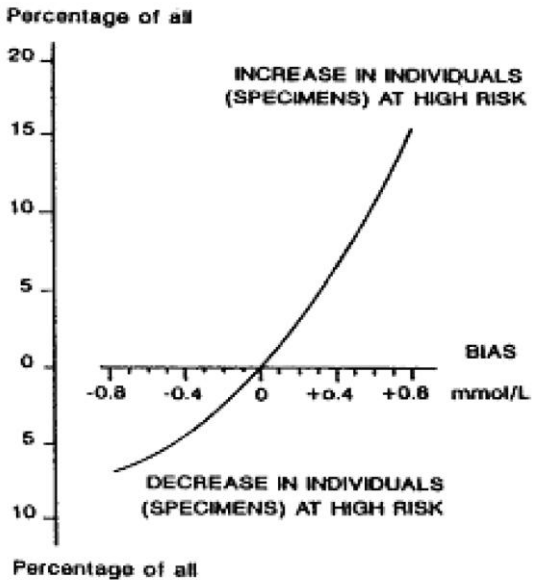


Figure2. Functional relationship between the percentage of the population at high risk and analytical bias. From Hyltoft Petersen P, Horder M. Scand J Clin Lab Invest 1992;58 [suppl 208]:79.