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@sesi.ahs.health.nsw.gov.au
School of Public Health, University of Sydney

Analytical Performance in the The Test Evaluation Cycle

Andrea Rita Horvath
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.

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.
**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]

**Clinical Outcome Assessments**

- **Biomarkers**
  - Cholesterol (coronary disease)
  - C-reactive protein (inflammation)

- **Performance**
  - Motor (timed 25 foot walk test)
  - Sensory (visual acuity, test reading)
  - Cognition (memory recall, or other cognitive)

- **Clinician-Reported**
  - Global impression of severity/change
  - Algorithmic ratings
  - Radiographic readings with human interpretation

- **Observer-Reported Signs**
  - Cough
  - Activity level
  - Sleep

- **Patient-Reported**
  - Symptoms
  - Function
  - Feelings
  - Perceptions

![ISPOR Clinician-Reported Outcomes: Good Measurement Practices Task Force](image-url)
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*

- Medical test → Test result → Clinical action
  - No direct effect, Side effect possible
  - Treat: Patient outcome
  - No treat: Patient outcome


Andrea Rita Horvath
Indirect linkage

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

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

Lukas P Staub1, Sarah J Lord1, R John Simes1, Suzanne Dyer1, Nehmat Houssami2, Robert YM Chen3 and Les Inglis3

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>
<thead>
<tr>
<th>Quality level</th>
<th>Imprecision goal (as CV)</th>
<th>Bias goalb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>&lt;13%&lt;sup&gt;a&lt;/sup&gt;</td>
<td>&lt;20%</td>
</tr>
<tr>
<td>Desirable</td>
<td>&lt;10%</td>
<td>&lt;10%</td>
</tr>
<tr>
<td>Optimum</td>
<td>&lt;6%&lt;sup&gt;a&lt;/sup&gt;</td>
<td>&lt;2.4%</td>
</tr>
</tbody>
</table>

<sup>a</sup>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.

<sup>b</sup>Assuming a diagnostic misclassification of 1.8%, 1.0%, and 0.5%.

Although the definition of analytical performance goals for cTnl 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 hypoglycaemia | Clin Chem 2010;56(7):1091-7 Clin Chem 2014;60(4):644-50 |
| HbA1c       | Intralaboratory CV <2%                  | Biological variation                     | NACB 2012                                      |
|             | Interlaboratory CV <3.5%                |                                          |                                                 |
| Cholesterol | CVa ≤ 3.0%, Bias≤ 3.0%, TE≤ 8.9%       | Diagnostic accuracy                      | NCEP/CRMLN 2004                                |

Andrea Rita Horvath
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?

![Diagram showing RCT design](image)

**What to do when RCTs are not available or possible?**

- 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

---

Andrea Rita Horvath
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

Evidence Linkage 1 - Core
- Evidence of comparative test accuracy
- Comparative assessment of test invasiveness & safety considerations

Optimization
- Test as accurate
  - Not as safe? ➔ NET HARM
  - As safe? = potential alternative test
  - Safer? = potential replacement test

Disease Spectrum Change
- Test more accurate
  - Not as safe? ➔ TRADE-OFF
  - As safe? = potential replacement or additional test
  - Safer? = replacement test

Trade-Off
- Test is less accurate or accuracy unknown
  - Not as safe?
  - As safe but no other advantages? ➔ NET HARM
  - As safe? Plus a pragmatic reason for use?
  - Safer? ➔ TRADE-OFF

Evidence Linkage 2 - Patient Management
- No change in management ➔ NO ADDED BENEFIT
- Impact on diagnostic and treatment strategy or impact uncertain

Evidence Linkage 3 - Treatment Effectiveness
- Implications of treatment on test positives (TP/PF)
- Implications of non-treatment for test negatives (TN/FN)
- Prognostic or further clinical evidence (if required)

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
Andrea Rita Horvath

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

- 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

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.

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 acknowledges the support of Roche Diagnostics and EFLM
Figure 2. Functional relationship between the percentage of the population at high risk and analytical bias. From Hyldtoft Petersen P, Horder M. Scand J Clin Lab Invest 1992;58 (suppl 208):79.