Steps of the process and different responsibilities in implementing traceability of patient results and defining their uncertainty

**Profession (e.g., JCTLM, IFCC, EFLM):**
- Define analytical objectives: reference measurement systems (traceability chain) and associated clinically acceptable uncertainty (fit for purpose)

**Diagnostic manufacturers:**
- Implement suitable measuring systems (platform, reagents, calibrators, controls) fulfilling the above-established goals

**End users (clinical laboratories):**
- Survey assay and laboratory performance through IQC and EQA redesigned to meet metrological criteria

---

**Expected consequences**

1. Experts define reference measurement systems
2. Industry implements traceability to them
3. Users (and industry) abandon non-specific methods
4. EQAs provide commutable materials and reference to track trueness-based grading
5. Professionals establish clinically allowable errors
6. Individual laboratories monitor their performance by participating in EQA and applying allowable limits
Expected consequences
1. Experts defines reference measurement systems
2. Industry implements traceability to them
3. Users (and industry) abandon non-specific methods
4. EQAs provide commutable materials and trueness-based grading
5. Professionals establish clinically allowable errors
6. Individual laboratories monitor their performance by participating to EQA and applying allowable limits

Fulfillment of the Requirements of the EU IVD Directive by Manufacturers
- Preparation of the necessary technical documentation
- All data that characterize the product
- Testing protocols
- Labels and instruction for use
- Assigned values and metrological traceability
  - Traceability chain and calibration hierarchy
  - Transfer protocols
  - Commutability testing
  - Determination of uncertainty (fit for purpose)
- Stability testing

Role of IVD manufacturers
IVD manufacturers should define a calibration hierarchy to assign traceable values to their system calibrators and to fulfill during this process uncertainty limits, which represent a proportion of the uncertainty budget allowed for clinical laboratory results.
Paradigm shift in the thinking

- If the manufacturer has to assume total responsibility for supplying products, in terms of traceability and uncertainty of the system ("CE marked"), it is no longer possible to consider separately the components of each analytical system (i.e., platform, reagents, calibrators and control materials), which in terms of performance can only be guaranteed and certified by the manufacturer as a whole.

- Any change introduced by users or third parties (e.g., the use of reagents, calibrators or control materials from other suppliers) may significantly alter the quality of the analytical system performance, removing any responsibility from the manufacturer and depriving the system (and, consequently, the produced results) of the certification originally provided through CE marking.

Limitations of CE mark

- Does not mean that manufacturer has transferred trueness successively
- Does not mean that uncertainty of calibrator meets clinical needs

Assessment of enzyme measurements in 70 European laboratories

![Enzyme measurements graph]

Assessment of enzyme measurements in 4 European countries

![Enzyme measurements graph]

CK is nicely standardized and a substantial improvement in analytical performance of marketed GGT assays was demonstrated. Conversely, aminotransferases, LDH and AMY still showed major disagreement suggesting the need for improvement in implementing traceability to higher order references.

Analytical systems measuring serum ALT marketed by four IVD companies

<table>
<thead>
<tr>
<th>Company</th>
<th>Platform</th>
<th>Principle of method</th>
<th>Calibrator</th>
<th>TE scores per analytical platform: TE scores vary considerably within users of instruments from the same manufacturer!!</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott</td>
<td>Architect</td>
<td>Calibration factor</td>
<td>M1</td>
<td>TE scores vary considerably within users of instruments from the same manufacturer!!</td>
</tr>
<tr>
<td>Beckman</td>
<td>Alinity</td>
<td>System calibrator</td>
<td>B1</td>
<td>TE scores vary considerably within users of instruments from the same manufacturer!!</td>
</tr>
<tr>
<td>Nichols</td>
<td>Nichols</td>
<td>Test System</td>
<td>B2</td>
<td>TE scores vary considerably within users of instruments from the same manufacturer!!</td>
</tr>
<tr>
<td>Roche</td>
<td>Cobas</td>
<td>Test System 2</td>
<td>B3</td>
<td>TE scores vary considerably within users of instruments from the same manufacturer!!</td>
</tr>
<tr>
<td>Siemens</td>
<td>Dimension</td>
<td>Take-Home Test</td>
<td>B4</td>
<td>TE scores vary considerably within users of instruments from the same manufacturer!!</td>
</tr>
<tr>
<td>Thermo</td>
<td>Chiron</td>
<td>Test System 3</td>
<td>B5</td>
<td>TE scores vary considerably within users of instruments from the same manufacturer!!</td>
</tr>
</tbody>
</table>
American Liver Guidelines and Cutoffs for "Normal" ALT: A Potential for Overdiagnosis

Mauro Panteghini,1,2 Rhoonoe A. Adiel,1 Ferruccio Cametti,2 Svene Sandberg,2 and Andrea Rita Horvath2

Expected consequences

1. Experts define reference measurement systems
2. Industry implements traceability to them
3. Users (and industry) abandon non-specific methods
4. EQAs provide commutable materials and trueness-based grading
5. Professionals establish clinically allowable errors
6. Individual laboratories monitor their performance by participating to EQA and applying allowable limits

Despite the availability of a reference measurement system (RMS) for standardizing ALT results in clinical samples, the current evidence is, however, that ALT is still measured by methods that give quite differing values (3). Assay performance also varies considerably within users of instruments from the same manufacturer (4). This is mainly due to the use on the same platform of various reagents with different analytical selectivity for ALT.

Percentage of Italian laboratories declaring to use methods for measuring enzyme employing the IFCC analytical principles

But, those who said to report enzyme results traceable to the IFCC RMPs, did they accurately recover the targets set by the reference laboratory?
Analytical systems measuring serum ALP marketed by four IVD companies


What is a “peer group”?

- Same model instrument/reagents/calibrator from one manufacturer?
- Same instrument family from one manufacturer?
- Instruments from different manufacturers that use the same reagent and calibrator?
- Methods with the same measurement principle with different reagents and calibrators?

Case study #1: Alkaline Phosphatase @

Platform [Architect c16000]

Reagents [Architect Alkaline Phosphatase cod. 7D55]

Calibrators [Calibration factor]

Control material(s) [Technopath Multichem-S plus]

Architect = 0.992 RMP + 0.9 U/L

R² = 0.9999

Architect ALP combined measurement uncertainty (uc) based on:

- uncertainty of values assigned by RMP (uref) = 1.25% bias (ubias) = 0.4% precision (uimp) = 1.5% (average CV Jan-Aug 2017)

uc = 2.0%
In collaboration with the EQA provider, a survey was issued to assess among participating laboratories using the Architect system which calibration factor was used. Among 39 interviewed laboratories: 

- 87% used theoretical CF [2150]
- 13% used experimental CF [2290]

The ‘peer-group’ consensus value used in the EQA was therefore expected to be strongly influenced by the type of calibration adopted by the majority of laboratories, i.e. the 'theoretical' CF.

We assume that this significantly lowers the EQA median value used as reference for evaluating the performance of individual participating laboratories and may explain our [apparent] positive total error. We expect that Abbott does indicate only one CF, i.e. that obtained by correlation results using clinical samples with RMP-assigned values.

Case study #2: Transferrin @EQAS results 2017

Technopath Multichem - S plus [Architect Transferrin cod. 1E04] [for the verification of system alignment]

Option #1
Plasmaproteins Cal (PC) cod. 11200D manufactured by Sentinel for Abbott

Option #2
Specific Proteins Multiconstituent Calibrator (SPMC) cod. 1E78 Abbott

uc = 2.1%
cuc = 1.7%
uc = 0.92%
uc = 1.7%
uc = 2.1%

Given the availability of two options warranting the CE mark provided by Abbott to calibrate the Architect Transferrin method, a correlation study (CLSI-EP 09A3) was performed to investigate the effect of different manufacturer’s calibrators on the same measuring system. In average, results were 7.6% biased.

Therefore, we must improve

- Post-market surveillance of IVD medical devices
Currently, the full information about calibration is usually not available.

Manufacturers only provide the name of higher order reference material or procedure to which the assay calibration is traceable, without any description of implementation steps and their corresponding uncertainty.

Profession (e.g., JCTLM, IFCC, EFLM):
Define analytical objectives: reference measurement systems (traceability chain) and associated clinically acceptable uncertainty (fit for purpose)

Diagnostic manufacturers:
Implement suitable measuring systems (platform, reagents, calibrators, controls) fulfilling the above established goals

End users (clinical laboratories):
Survey assay and laboratory performance through IQC and EQA redesigned to meet metrological criteria

Analytical Quality Control in the Traceability Era

External Quality Assessment

Analytical quality of measurement

Check alignment

Reliability of the analytical system

Internal Quality Control

Expected consequences

1. Experts define reference measurement systems
2. Industry implements traceability to them
3. Users (and industry) abandon non-specific methods
4. EQAs provide commutable materials and trueness-based grading
5. Professionals establish clinically allowable errors
6. Individual laboratories monitor their performance by participating to EQA and applying allowable limits

Steps of the process and different responsibilities in implementing traceability of patient results and defining their uncertainty

Steps: Post-marketing surveillance of IVD metrological traceability
Quality of EQA target – Concepts

- True value assignment to EQA materials allows objective evaluation of the performance of laboratory measurements through an trueness-based (instead of inferior consensus-based) grading of the competency of participating clinical laboratories.

- Analytically valid reference measurement procedure (ISO 15193)

- Competent reference laboratory (ISO 17025/ISO 15195 accreditation)

Joint Committee for Traceability in Laboratory Medicine listed

ACTLM Database Status – June 2015
Reference measurement services:

- 130 reference measurement services listed
- 12 Reference Laboratories accredited for compliance with ISO 15195/ISO 17025 and 2 NMIs

<table>
<thead>
<tr>
<th>Analyte Categories</th>
<th>Number of Analytes</th>
<th>Analytes</th>
<th>Number of Reference Laboratories</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzymes</td>
<td>5</td>
<td>ALT, AST, ALP, GGT, LDH</td>
<td>9</td>
<td>Germany</td>
</tr>
<tr>
<td>Metabolites and Seleniums</td>
<td>9</td>
<td>creatinine, uric acid, urate (total), phosphate, calcium, amylase, BUN, Cr (direct), cholesterol, HDL-C, LDL-C, triglycerides</td>
<td>10</td>
<td>Belgium, France, Germany, Italy, Spain, United Kingdom, China</td>
</tr>
<tr>
<td>Non-peptide Amines</td>
<td>10</td>
<td>17 Aminobenzimidazoles, adenine, adenosine, guanine, cytosine, thymine, ribose, deoxyribose</td>
<td>4</td>
<td>Belgium, Germany, United Kingdom</td>
</tr>
<tr>
<td>Proteins</td>
<td>6</td>
<td>4 protein, total protein</td>
<td>6</td>
<td>France, Germany, Italy, Spain, China</td>
</tr>
<tr>
<td>Valinines</td>
<td>2</td>
<td>Valinines</td>
<td>2</td>
<td>Belgium</td>
</tr>
<tr>
<td>Total</td>
<td>130</td>
<td>130</td>
<td>39</td>
<td></td>
</tr>
</tbody>
</table>

List of reference measurement services

Linda Thienpont @ CIRME 2015

Trueness verification in EQA: time to care about the quality of the samples!


Autem censeo Carthaginem delendam esse

[However, I think that Carthage should be destroyed]
EQA and patient results have the same relationship between measurement procedures.
EQA results reflect performance for patient results.

- EQA samples are frequently not validated to be commutable.
- Commutability is assumed based on how the samples were prepared:
  - May be reasonable for single donation
  - Potential limitation for spiked pools or supplemented

**Expected consequences**

1. Experts define reference measurement systems
2. Industry implements traceability to them
3. Users (and industry) abandon non-specific methods
4. EQAs provide commutable materials and trueness-based grading
5. Professionals establish clinically allowable errors
6. Individual laboratories monitor their performance by participating in EQA and applying allowable limits

---

**Working Group on Commutability**

Upcoming recommendations for assessing commutability

- **Part 1: General experimental design**
- **Part 2:** Based on the difference in bias between a reference material and clinical samples
- **Part 3:** Based on the calibration effectiveness of a reference material

---

**Table 3. Evaluation capabilities of IFCC related to scheme design.**

<table>
<thead>
<tr>
<th>Evaluation capabilty</th>
<th>Individual laboratory</th>
<th>Relative to participant results</th>
<th>Repeatability</th>
<th>Standardization or calibration</th>
<th>Measurement procedure calibration linearity</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQA samples</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>EQA and patient results</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Measurement Procedure 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Measurement Procedure 2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

---

**Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine**

Model 1: Based on the effect of analytical performance on clinical outcomes

a. Done by direct outcome studies – Investigating the impact of analytical performance on the clinical performance of the test on clinical classifications or decisions and hence on the probability of patient outcome, e.g., by simulation or decision analysis.

b. Done by indirect outcome studies – Investigating the impact of analytical performance of the test on clinical classifications or decisions and hence on the probability of patient outcome, e.g., by simulation or decision analysis.

Model 2: Based on components of biological variation of the measured

Model 3: Based on data on the art of the measurement (i.e., the highest level of analytical performance technically feasible).
Opinion Paper

Ferruccio Ceretti*, Piliz Fernandez-Calle, George S. Kleo, Gunnar Nordin, Scrie Sandberg, Thomas Striethet, Jean-Louis Vives-Corrons and Mauro Poggesi, on behalf of the EFLM Task and Finish Group on Allocation of laboratory tests to different models for performance specifications (TFG-DM)

Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference

Analytical performance specification (APS) derivation should be added to the Miller’s EQA categorization

Category 1/2A → Milan model 1 or 2 as basis for APS
Category 1/2B → Other models

Grading different quality levels

The utility to elaborate specifications at different levels of quality to move, in case, from desirable to minimum quality goals and, in the meantime, ask IVD manufacturers to work for improving the quality of assay performance

IDEAL

OPTIMUM STANDARD (no need to improve)

DESIRABLE STANDARD (satisfactory)

MINIMUM STANDARD (just satisfactory)

UNACCEPTABLE

There are as many limits as there are EQA providers

Comparison of evaluation procedures used by European external quality assessment scheme organizers for haemoglobin and leucocyte concentration

Table 1: Criteria and/or acceptable performance for haemoglobin and leucocyte concentration

Table 2: Prevalence of analytically correct results reported by the participating EFLM organizers for 2 levels of 25% of the participants for haemoglobin and leucocyte concentration in the 1st EFLM Strategic Conference

EQA Performance Specifications

Why?

(tighter) → Quality improvement
A portion of labs fail

Adopt tighter guidelines to improve "TQM" Quality

(looser) → Regulatory
All labs pass

Adopt more lenient guidelines to avoid excessive negative impact on "TQM"
Basic elements that need to be considered:

a) nature of the EQA material, including commutability, which may affect the result interpretation;
b) procedure used to assign the target value;
c) data set to which APS are applied;
d) analytical property being assessed (i.e., TE, bias, imprecision);
e) rationale for the selection of the APS;
f) type(s) of model used to set APS

We need these to:
1. compare APS from EQAs
2. inform users about the APS they use
3. plan harmonization (common EQA APS would support uniform analytical performance and true quality improvement)

### Requirements for the applicability of EQA results in the evaluation of the performance of participating laboratories in terms of traceability of their measurements

<table>
<thead>
<tr>
<th>Feature</th>
<th>Aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQA materials value-assigned with reference procedures by an accredited laboratory</td>
<td>To check traceability of commercial system to reference measurement systems</td>
</tr>
<tr>
<td>Proved commutability of EQA materials</td>
<td>To allow transferability of participating laboratory performance to the measurement of patient samples</td>
</tr>
<tr>
<td>Definition and use of the clinically permissible measurement error</td>
<td>To verify the suitability of laboratory measurements in clinical setting</td>
</tr>
</tbody>
</table>

### POST-MARKET SURVEILLANCE

Serum Albumin: Norwegian Survey 2011

The results postulate an urgent need for improving traceability implementation of albumin assays by IVD manufacturers.

Biological limits for bias

Current State of Harmonization of Serum Albumin Measurements

Expected consequences
1. Experts define reference measurement systems
2. Industry implements traceability to them
3. Users (and industry) abandon non-specific methods
4. EQAs provide commutable materials and trueness-based grading
5. Professionals establish clinically allowable errors
6. Individual laboratories monitor their performance by participating in EQA and applying allowable limits

Case study #3: Creatinine @

Metrological traceability chain and measurement uncertainty of Abbott Architect enzymatic creatinine assay
Abbott Diagnostics in a document released on August 2014 informed customers that the internal release specification for CAL was ±5% from the target value of SRM 967a L1, which is more than two times higher than the SRM expanded uncertainty.

Post-market surveillance of IVD medical devices: further issues

- Possibility to select different types of traceability chains by IVD manufacturers
- Uncertainty (including imprecision) of the measuring systems for certain analytes may be too large
- Commercial assay may not be selective for the measurand

TRACEABILITY CHAINS AVAILABLE FOR IVD MANUFACTURERS FOR PLASMA GlUCOSE

IVD manufacturers may spend different amounts of the total uncertainty budget in implementing traceability of their measuring systems.
Post-market surveillance of IVD medical devices: further issues

- Possibility to select different types of traceability chains by IVD manufacturers
- Uncertainty (including imprecision) of the measuring systems for certain analytes may be too large
- Commercial assay may not be selective for the measurand

Further advances are needed to:
1. reduce uncertainty associated with higher-order metrological references (reference materials and procedures)
2. increase the precision of commercial HbA1c assays
Selectivity definition

“Property of a measuring system used with a measurement procedure, whereby it provides measured quantity value for one or more such that the values of each measurand are independent of other measurands or other quantities in the phenomenon, body, or substance being investigated.”

- The alkaline picrate method is unable to measure solely creatinine
- Endogenous and exogenous substances may significantly interfere: particularly, proteins in serum can lead to significant overestimation with various alkaline picrate methods
- The analytical non selectivity issue: the case of serum creatinine
- Unselective color reaction
- Selective enzymatic reaction

EQA for quantities where no high-order reference is available

System-dependent target values should be used to evaluate the performance of participating laboratories

HOWEVER

in this case the values assigned to the EQA materials should be determined by reference institutions (possibly including the manufacturer releasing that specific measuring system), working under strictly controlled conditions in order to maintain measurement uncertainty as low as possible, and not as a peer group mean.

Case study #4: Folate

• To improve assay harmonization, in 2016 Roche folate method has undergone recalibration to the WHO NIBSC 03/178 International Standard
• After recalibration, a significant change in the average folate measured values was internally recorded
Case study #4: Folate @

At a folate concentration around the lower reference limit of the old Roche assay, a positive bias of ~50% vs. the recalibrated Roche assay can be observed.

Roche Folate III assay code 04476433190 (home-made calibration)
Roche Folate III assay code 04476433190 (traceable to NIBSC 03/178 IS)

EQA exercise no. 4/2016

Do not forget the post-analytical EQA

Taking into account the ~50% difference experimentally found at the lower reference limit (LRL) level, the shift from 4.6 μg/L (Roche recommended LRL for old calibration) to 3.9 μg/L (Roche recommended LRL for recalibrated assay) appeared to be inconsistent.

Consequently, a misleading overestimate of the prevalence of folate deficiency is expected if the recalibrated Roche assay will be used together the manufacturer’s newly recommended LRL.


Conventional External Quality Assessment

Constraints limiting the introduction of EQA that meet metrological criteria
What COPERNICUS did was take the existing 'a priori' concept of the world and pose an alternative 'a priori' concept.

What TRACEABILITY does is take the existing 'a priori' concept of Quality Control and pose an alternative 'a priori' concept.

The earth is flat and fixed in space.
The earth is spherical and moves around the sun.

Equivalency-based grading
Accuracy-based grading

Table 1: Unique benefits of External Quality Assessment Schemes meeting metrological criteria.

- Giving objective information about quality of individual laboratory performance
- Creating evidence about intrinsic standardisation status/equivalence of the examined assays
- Serving as management tool for the clinical laboratory and IVD manufacturers, forcing them to investigate and eventually fix the identified problem
- Helping those manufacturers that produce superior products and systems to demonstrate the superiority of those products
- Identifying analytes that need improved harmonisation and stimulating and sustaining standardisation initiatives that are needed to support clinical practice guidelines
- Abandonment by users (and consequently by industry) of nonspecific methods and/or of assays with demonstrated insufficient quality

“It was the acceptance of the Copernican revolution that distinguishes modern man from his medieval predecessors.”

Robert M Pirsig – “Zen and the art of motorcycle maintenance”, 1974

Is it possible to create a perfect external control system?

P. H. PETERSEN
Department of Clinical Biochemistry, Odense University Hospital, Odense, Denmark