Standardization and Harmonization in Laboratory Medicine: A Matter of Patient Safety

Mauro Panteghini
University of Milan Medical School
Research Centre for Metrological Traceability in Laboratory Medicine (CIRME)

Drivers for global harmonization in laboratory medicine

- Clinical governance and guidelines (EBM)
- Laboratory accreditation, consolidation and networking
- Advances in IT and electronic health records

Lab-related causes of diagnostic mistakes

The triad of elements of value in laboratory information

Terminology

- HARMONIZE → to make something free from disagreement
- STANDARDIZE → to make something conform to a fixed standard, shape, quality, etc.

“Wrongs” anywhere compromise test result quality and patient safety!
Key preanalytical steps identified as the most critical and in need of immediate harmonization

Test ordering
Transport and storage
Patient Preparation
Sampling
Management of unsuitable specimens
Quality Indicators
Patient Identification
Paediatric and neonatal sampling

Rates of inappropriate testing

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Error rate (95% CI)</th>
<th>Difference (95% CI)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subgroup</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overutilization</td>
<td>26.1 (16.2, 24.0)</td>
<td>(reference)</td>
<td>114</td>
</tr>
<tr>
<td>Normalization</td>
<td>40.6 (33.8, 47.4)</td>
<td>24.2 (12.5, 36.0)</td>
<td>18</td>
</tr>
<tr>
<td>Initial testing</td>
<td>43.9 (35.4, 52.5)</td>
<td>(reference)</td>
<td>18</td>
</tr>
<tr>
<td>Repeat testing</td>
<td>7.4 (2.5, 12.0)</td>
<td>-36.5 (-46.6, -26.4)</td>
<td>55</td>
</tr>
<tr>
<td>BMP</td>
<td>26.0 (22.2, 33.6)</td>
<td>-15.9 (-23.8, -26.0)</td>
<td>41</td>
</tr>
</tbody>
</table>

NHS Atlas of Diagnostic Variation

- 72-fold difference in PSA requesting between most and least requesting groups of GPs
- 89-fold difference in BNP requesting
- 106-fold difference in annual rate of creatinine requesting
- 446-fold difference in calprotectin
- 4/5-fold difference in average even when outliers are removed.

Why we need to reduce variation in test requests?

“...the existence of persistent unwarranted variations in health care directly impacts on equity of access to services, the health outcomes of populations and efficient use of resources.”

Implementing recommendations on correct use of tumor markers largely decrease the number of ordered tests, without any impact on clinical outcome
In general, we recommend that one TM is usually enough for disease management, and thus we established that no more than two different TMs should be ordered in the same transaction.

Orders asking for more than two TMs are always discussed with the requesting clinicians.

Each blocked request contained a median of three TMs (up to seven). After discussion with requesting clinicians, the agreed total number of TMs decreased of –43.3%.

Removal from the test menu of obsolete and useless tests

- Removing tests that offer no incremental information would save money, avoid additional investigations arising from incidental and clinically irrelevant abnormalities, and improve the risk to benefit ratio.

- For instance, deleting myoglobin, total creatine kinase (CK) and CK MB isoenzyme determinations from laboratory order forms in patients admitted to ED leads to significant cost saving and reduces possible confusion in data interpretation and patient management. Overall testing costs were reduced by €104,871 per annum.

Markers still used for the diagnosis of AMI in addition to cardiac troponin

The Cardiac Marker Guideline Uptake in Europe (CARMAGUE) Study of the EFLM WG Cardiac Markers

Diagnostic algorithm for a diagnosis of heart failure of non-acute onset

Avoid pathophysiologic duplications

- Serum creatinine and urea
- Erythrocyte sedimentation rate and C-reactive protein
- Aspartate aminotransferase and alanine aminotransferase
- Total and conjugated bilirubins
- Amylase and lipase
- Prothrombin time and activated partial thromboplastin time
- Free T4 and free T3.
- Total and free prostate specific antigen
- Fibrin and fibrinogen saturation
Testing of AST in addition to ALT should be limited because:
- ALT is the more liver-specific enzyme. Moreover, increases of ALT activity persist longer than AST.
- The incremental benefit of routine determination of AST, in addition to ALT, is limited.
- Laboratories reporting abnormal ALT results (i.e., >twice the URL) should offer AST as a reflex test and calculate the AST-to-ALT ratio because it provides useful diagnostic and prognostic information.

Vetting of restricted tests
Procalcitonin testing for diagnosing and monitoring sepsis [JADVIDIA Centaur BRAHMS PCT assay, BRAHMS PCT Sensitive Kryptor assay, Elecsys BRAHMS PCT assay, LIAISON BRAHMS PCT assay and VIDAS BRAHMS PCT assay]

Procalcitonin results can help doctors to diagnose bacterial infection and decide about starting or stopping antibiotic treatment.

There was not enough evidence to recommend that this test is used in the NHS.
NICE has recommended further research and data collection to show the impact of adding procalcitonin testing to standard clinical practice.

Restricted Policy
- PCT can be ordered by ICU as an aid in decision for continuing or stopping antibiotics
- PCT can be ordered in neonates with suspected late-onset sepsis (>72 h of life)
  - Blood sampling at presentation; PCT cut-off, 0.5 µg/L
Prevention is better than cure: stopping inappropriate requests before they reach the laboratory

The potential of electronic requesting acting as ‘enabling factor’

Table 4: Reasons for 4852 Identification Errors

<table>
<thead>
<tr>
<th>Reason for Identification Error</th>
<th>No. of Errors</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary specimen label error</td>
<td>2691</td>
<td>55.5</td>
</tr>
<tr>
<td>Initial registration form error</td>
<td>1088</td>
<td>22.4</td>
</tr>
<tr>
<td>Other clerical error</td>
<td>604</td>
<td>12.4</td>
</tr>
<tr>
<td>Other reason for error</td>
<td>205</td>
<td>4.2</td>
</tr>
<tr>
<td>Unadjusted label error</td>
<td>184</td>
<td>3.8</td>
</tr>
<tr>
<td>Result entry error</td>
<td>80</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Table 5: Frequency of Adverse Event Types (N = 345)

<table>
<thead>
<tr>
<th>Type of Event</th>
<th>No. of Adverse Events</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significant patient inconvenience</td>
<td>251</td>
<td>72.8</td>
</tr>
<tr>
<td>No known change in treatment or patient outcome</td>
<td></td>
<td>16.4</td>
</tr>
<tr>
<td>Patient impacted by identification error; nature of impact unknown</td>
<td>78</td>
<td>22.6</td>
</tr>
<tr>
<td>Change in patient treatment; no known change in patient outcome</td>
<td></td>
<td>16.4</td>
</tr>
</tbody>
</table>

Characteristics of Tests Performed in 10 Calgary Area Hospitals That Were Improperly Requested

- Test Type
  - Annual Test Count at the Study Laboratory
  - Percentage Required in Appropriation
  - Revenues at the Study Laboratory

- Test Type
  - Annual Test Count at the Study Laboratory
  - Percentage Required in Appropriation
  - Revenues at the Study Laboratory
Reasons for Rejecting Chemistry and Hematology Specimens

- CAP Q-Probe 95-02, Chemistry Specimen Acceptability (n=461 labs)
  - 60% Hemolyzed
  - 11% Insufficient quantity
  - 7% Inadequately labeled
  - 3.5% Improper collection tube
  - 2% Clotted

- CAP Q-Probe 92-05, Hematology Specimen Acceptability (n=604 labs)
  - 65% Clotted
  - 10% Insufficient quantity
  - 5% Unacceptable variance (delta check)
  - 5% Inadequately labeled
  - 2% Platelet clumps
  - 2% Hemolyzed

Evidence Review
Straight Needle Venipuncture vs. IV Catheter Starts
Conclusions and Recommendations

- Straight Needle Venipuncture must replace the use of intravenous catheters as the primary method of collecting blood samples in the Emergency Department in order to significantly reduce rates of hemolysis.

LMBP Recommendation Statement

- Clinical Biochemistry 2012; 45:1012-32

SAMPLE REJECTION FOR HEMOLYSIS

- Practices for identifying and rejecting hemolyzed specimens are highly variable in clinical laboratories.
- Hemolysis is still identified using a visual scale by >20% of clinical laboratories in the UK and 48% in the USA.

Harmonize management of unsuitable samples: the most challenging issue?

- Provide comment and no result: 16%
- Adjust sample result using algorithms: 3%
- Do not process specimen and contact the ward: 27%
- Perform analysis and associate comment: 54%
Visual handling of hemolyzed samples increases the risk of reporting inaccurate results for cTnT, K and bilirubin, possibly affecting the clinical decision and patient outcome.
The choice of anticoagulant

PLASMA GLUCOSE DETERMINATION: GOLD STANDARD FOR SAMPLE COLLECTION

NATIONAL ACADEMY OF CLINICAL BIOCHEMISTRY (NACB) GUIDELINES FOR LABORATORY ANALYSIS IN DIABETES MELLITUS

- Tubes with only enolase inhibitors, such as NaF, should not be relied on to prevent glycolysis
- A tube containing a rapidly effective glycolysis inhibitor, such as citrate buffer, should be used for collecting the sample

Claeys K et al. Clin Chem 2011;57:e1–47

PLASMA GLUCOSE DETERMINATION: GOLD STANDARD FOR SAMPLE COLLECTION

The choice of anticoagulant

Challenges reported by US primary care physicians when using lab test results

<table>
<thead>
<tr>
<th>% of respondents regarding factor as very or extremely problematic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results not received in a timely manner</td>
</tr>
<tr>
<td>Previous results not easily available</td>
</tr>
<tr>
<td>Errors in results are suspected</td>
</tr>
<tr>
<td>Results are inconsistent with patient’s symptoms</td>
</tr>
<tr>
<td>Report format</td>
</tr>
<tr>
<td>Lab-number variation in normal range</td>
</tr>
<tr>
<td>Lab-number variation in report format</td>
</tr>
<tr>
<td>Lab-interpretation is difficult to understand</td>
</tr>
<tr>
<td>Not enough information in lab report</td>
</tr>
</tbody>
</table>

Potential factors affecting 13 million pts/yr, raising significant concerns about the safety and efficient use of lab tests

Terminology in Laboratory Medicine: IFCC Survey 2015

Which nationally agreed system of laboratory terminology is in use in your country?

Not specified 42%
NPU 21%
LOINC 16%
Others (e.g. national system) 21%

CLINICAL CLASSIFICATION OF SUBJECTS UNDERGOING FASTING PLASMA GLUCOSE TEST

<table>
<thead>
<tr>
<th>% samples</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>≤5.55 mmol/L</td>
<td>normal, no diabetes</td>
</tr>
<tr>
<td>&gt;5.55 mmol/L</td>
<td>abnormal, impaired glucose tolerance</td>
</tr>
<tr>
<td>&gt;7.00 mmol/L</td>
<td>diabetes</td>
</tr>
</tbody>
</table>

Terminology in Laboratory Medicine: LOINC vs. NPU vs. others

National Survey 2015: post-analytical phase

How do you define the analyte in the report?

<table>
<thead>
<tr>
<th>Analyte</th>
<th>% of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microalbumin</td>
<td>28%</td>
</tr>
<tr>
<td>Albuminuria</td>
<td>42%</td>
</tr>
<tr>
<td>U-Albumin</td>
<td>19%</td>
</tr>
<tr>
<td>Urine Albumin</td>
<td>11%</td>
</tr>
</tbody>
</table>

Current use of SI units in Europe

Declared % of SI adoption:

- 100%
- 75–80%
- 50–59%
- 25–29%
- 10–24%
- <10%
Avoiding the unnecessary use of decimals is matter of patient safety. Units for cTn should be harmonized regardless of the analytical sensitivity of the assay used.

The Cardiac Marker Guideline Uptake in Europe (CARMAGUE) Study of the EFLM WG Cardiac Markers

Cardiac troponin Units

Variations in reference intervals: a risk for lab result interpretation

- The used R.I. are often significantly different among laboratories
- The origin of these R.I. may be quite different (manufacturers, literature, internal studies, other laboratories, undefined)
- In the majority of cases, differences among laboratories have no justification

For some analytes, variation in reference intervals is greater than the analytical inaccuracy in their measurement

Lack of proper reference intervals may hamper the implementation of standardization

- The implementation of standardization can modify the analyte results
- Without adequate R.I. this situation can impair the interpretation of the results and, paradoxically, worsen the patient’s outcome
- The absence of reliable R.I. for the newly standardized commercial methods hampers their adoption
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 recorded.

Case study: Serum 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 recorded.

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 was expected if the recalibrated Roche assay will be used together the manufacturer’s newly recommended LRL.


2.5 th percentile traceable to NIBSC 03/178 IS home-made calibration

1.3 μL Traceable reference intervals: how a problem becomes a solution

Historically
Method-dependent results ➔ Method-dependent reference intervals

Traceability era
Standardized methods that provide traceable results ➔ Traceable reference intervals

Premise
To be used together with commercial assays that provide traceable results to established IFCC reference measurement systems

Common reference intervals for enzymes in adults

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>European</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>AST</td>
<td>11–34</td>
<td>14–32</td>
</tr>
<tr>
<td>ALT</td>
<td>8–41</td>
<td>9–59</td>
</tr>
<tr>
<td>GGT</td>
<td>6–40</td>
<td>12–68</td>
</tr>
<tr>
<td>CK</td>
<td>34–145</td>
<td>46–171</td>
</tr>
</tbody>
</table>

Traceable reference intervals: how a problem becomes a solution

Premise
To be used together with commercial assays that provide traceable results to established IFCC reference measurement systems

Histologically
Method-dependent reference intervals

Traceability era
Standardized methods that provide traceable results ➔ Traceable reference intervals

European
Asian

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>European</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>AST</td>
<td>11–34</td>
<td>14–32</td>
</tr>
<tr>
<td>ALT</td>
<td>8–41</td>
<td>9–59</td>
</tr>
<tr>
<td>GGT</td>
<td>6–40</td>
<td>12–68</td>
</tr>
<tr>
<td>CK</td>
<td>34–145</td>
<td>46–171</td>
</tr>
</tbody>
</table>

Premise
To be used together with commercial assays that provide traceable results to established IFCC reference measurement systems
**The comparator is as important as the result**


**Opinion**

American Liver Guidelines and Cutoffs for “Normal” ALT: A Potential for Overdiagnosis

Mauro Panteghini,1 Khosrow Adeli,1 Francesco Curri,7,8 Sivare Sandberg,5 and Andrea Rita Horwath7

Improvements required:

– Awareness of differences in measuring systems (lack of standardization)
– Using only assays with proved traceability (standardized)
– Laboratory professional involvement with clinical guidelines

---

**ALP pediatric reference intervals: A chaos**

Higgins V et al. 69th AACC Annual Scientific Meeting Abstract A-259

Canadian Reference Intervals Survey


Traceable reference intervals for alkaline phosphatase in serum of pediatrics

Pinto G, 12 years: 145–490
13-14 years: 165–470
15-16 years: 195–530
17-18 years: 215–550
19-20 years: 235–575

The result standardization issue: an absolute priority for public health

Good results are our contribution to healthcare

→ Our customers (i.e., doctors and patients) expect laboratory results to be equivalent for being interpreted in a consistent manner

---

**Improved Cholesterol Measurement Trueness Saves Health Care Resources**


Improvement in trueness since 1968 has been estimated to save ranging from $338 million to $7.6 billion in treatment costs and lives saved

---

**Adverse clinical effect of analytical bias in creatinine on the distribution of estimated GFR values**

Klee GG et al., Clin Chem Lab Med 2007;45:737

Assuming a 10% analytical bias in creatinine, the cumulative distribution of eGFR values for GFRs ranges from 0 to 90 ml/min/1.73 m².

---

---
Biased results → Patient harm:
• Wrong diagnosis
• Wrong management
• Incorrect monitoring

In short: the lack of standardization may become an ethical issue

“Standardization of laboratory tests has an ethical dimension as it aims to affect the way diagnostic tests are used in order to guarantee optimal care for patients in a global world.”

Bossuyt X et al., Ann Rheum Dis 2008;67:1061

Biased results
Patient harm:
• Wrong diagnosis
• Wrong management
• Incorrect monitoring

→ To become equivalent for being interpreted in a consistent manner, results must be traceable to higher order references.

Both require IVD manufacturers to ensure traceability of their measuring systems to recognized higher order references

Role of IVD manufacturers

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


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.

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
The quality of glucose measurement may be dependent on the type of traceability chain selected for trueness transferring, sometimes making difficult (e.g., chain C) to achieve the suitable limits for measurement uncertainty on clinical samples.

Limitations of CE mark

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

Limitations of CE mark

(Stating compliance with legislation, mainly by means of European standards)

Analytical systems measuring serum ALP marketed by four IVD companies

Limitations of CE mark

(Stating compliance with legislation, mainly by means of European standards)
Case study: Creatinine

Abbott Diagnostics in a document released on August 2014 informed customers that the internal release specification for CAL was 45% from the target value of NIST SRM 967a L1, which is more than two times higher than the SRM expanded uncertainty.

IVD Manufacturer Validation Criterion for Calibrator Traceability

<table>
<thead>
<tr>
<th>Insert Range</th>
<th>Lot 30517/900 (Mean)</th>
<th>Lot 40627/900 (Mean)</th>
<th>Lot 40627/900 (Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target: 0.05%</td>
<td>0.02</td>
<td>0.08</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Manufacturers’ release specification is 45% from the target

= 3.53%

Role of Professional Organization:
- Unequivocal definition of the *measurand* as the quantity subject to measurement
- Definition of the *reference measurement system*
- Definition of *analytical performance specifications*

The Essential Question...

“What amount of medical harm due to analytical error is it ok to let go undetected?”

The Temple of Laboratory Standardization
There are as many limits as there are EQA providers

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 EQA
2. inform users about the APS they use
3. plan harmonization (common EQA APS would support uniform analytical performance and true quality improvement)
Key Messages

- As leaders in our profession we have a key role to play in facilitating better patient’s outcome.
- One barrier to improved outcomes is excessive variability.
- Only a relatively small percentage of procedures within the total examination process (TEP) have been standardized or harmonized.
- Where TEP procedures have been standardized or harmonized, evidence of improved clinical outcome is emerging.
- As a profession we should:
  - Facilitate the standardization or harmonization of many procedures as possible.
  - Work with clinical colleagues to demonstrate improved outcomes.