Editorial

Sandra Secchiero and Mario Plebani

A new integrated tool for assessing and monitoring test comparability and stability

DOI 10.1515/cclm-2015-0062

Performing accurate measurements, comparable over time and location and across assays, is essential to ensure appropriate clinical practice [1]. One step towards achieving this goal is using assays that are metrologically traceable to a higher order reference measurement system or harmonized using internationally recognized procedures [2, 3]. In Europe, diagnostic manufacturers are required by the European Union Directive on in vitro diagnostic medical devices to demonstrate the metrological traceability of their analytical systems [4]. Thus, in principle, laboratories using CE-marked homogeneous systems (calibrator, reagent and instrument from the same manufacturer) can assume accuracy and interchangeability of their measurements results. However, it is the task of clinical laboratory profession to verify that the alignment process has been correctly implemented and that the performances of marketed systems are really appropriate for their clinical use. In particular, the laboratory has to verify the consistency of the manufacturer's declared performance during routine operations performed strictly in accordance with their instruction, and to participate in external quality assessment (EQA)/proficiency testing (PT) schemes structured so that they provide objective information on the analytical quality of measurements performed by clinical laboratories and on the performances of the peer group assays [5–7]. EQA programs using commutable materials and reference method target values are becoming more widespread, but until now they are applied only to few constituents (i.e., hemoglobin A_{1c}) [8, 9]. Generally, an EQA program for clinical biochemistry on serum, performed by an independent national or regional center, with a sufficient number of participants so that the peer groups can be well defined, consists in a panel of about 25–35 measurands. Even if controls materials are made ad hoc, namely to evaluate clinically important concentrations, the commutability for all measurands and diagnostic systems is difficult to be achieved. Moreover, being target values by reference methods available for a few measurands and extremely expensive, it is very difficult to get a reference target values for all measurands. Finally, just for the intrinsic design of an EQA program, the data are not available in real time. In this issue of *Clinical Chemistry and Laboratory Medicine*, the team of Thienpont and colleagues, whose remarkable dedication to the improvement of quality, standardization and harmonization of clinical laboratory testing is well appreciated at an international level and by the readers of the journal [10–14], describes a new project aimed to establish a bottom-up cooperation between laboratories and manufacturers, so that they can pursue the common objective of assessing test comparability and stability of laboratory results [15]. The "Empower" project, proposed as an independently operated "online" tool that should monitor comparability and long-term stability between peer groups and laboratories, comprises four pillars: 1) Master comparison with panels of frozen single-donation samples; 2) Monitoring of patients percentiles; 3) Monitoring of IQC data, both across laboratories and manufacturers; 4) Conceptual and statistical education about analytical quality in the medical laboratory and elaboration of actionable experiments for analytical quality management and assurance. Laboratories are free to participate in all pillars of the project or to select the most appropriate one(s) for their purpose and the paper reports the status of the project with respect to the first two points. The focus of the master comparisons, which are conducted across assays and laboratories, is on how well the intrinsic analytical quality of assays released by the respective manufacturers is reproduced by the end users under "field" conditions all over the world. Results about master comparison with a panel of 20 frozen single donation samples for calcium, magnesium, albumin, total protein [16, 17], sodium [18], creatinine, glucose, phosphate, uric acid, total cholesterol, HDL cholesterol, LDL cholesterol and triglycerides [19], have been previously described. In a recent survey AST, ALT, LDH, GGT and ALP, sodium, potassium and chloride have been added. A limitation of these studies was that for the most measurands the target values were the "all method trimmed mean"

rather than obtained using reference methods (with the exception of total cholesterol, creatinine and uric acid). However, it should be highlighted that the primary focus of the study was to evaluate the differences between methods on real patient materials and avoiding any commutability issue, rather than the accuracy of the assays. Another limitation of these studies was the narrow concentrations range covered by the single donation samples, leaving open the question about the analytical performances at other concentrations that, in fact, should be observed in real patient samples. However, these specially designed surveys (which distribute a relatively small number of single donor samples to a relatively small number of laboratories representing the major homogeneous systems), showed that, even for "simple" clinical chemistry measurands, the standardization status of certain assays is still a matter of concern and there is much room for improvement. The data further showed that assays, for some tightly regulated measurands, do not meet the optimal bias limits necessary for clinical use and this also applied for cholesterol and creatinine, in spite of dedicated standardization programs, such as the National Cholesterol Education Program and the National Kidney Disease Education Program [20, 21]. It is very important to emphasize that these surveys must be complementary to those of independent national or regional EQA schemes, which distribute commutable controls materials with a range of concentration for each measurand covering the main clinical situations, among a large number of participants. The paper published in this issue focuses on the second pillar of the "Empower" project: the monitoring of patients percentiles to which currently 124 laboratories participate with 250 devices. The laboratories calculate instrument-specific daily medians from outpatient results and send the data (by e-mail or electronic transfer) to the author's database. Via a user interface the participating laboratory can plot for each measurand the course of the moving median. Each plot also shows the long-term median of the concerned individual laboratory, as well as the peer group or all devices median [22]. Additional numerical information are provided on the long-term imprecision (robust CV, %) and bias calculated in comparison to a "desirable" target, such as the medians of the reference intervals determined in the trueness-based "Nordic Reference Interval Project (NORIP)" [23]. Laboratories can use their medians as a tool to monitor the mid- to long-term stability of their own calibration status, in comparison to their peer, and/or to uncover shifts/drifts and the sources thereof. Another asset of percentile monitoring design is that it shows the instrument-specific stability in one plot allowing laboratories to monitor the interchangeability of results among different instruments, and detect the occurrence of instrument-specific special events. Results from patient percentile monitoring show that laboratories with high daily throughput and/or low variation in patient population typically perform with low variation and mostly good concordance between the different instruments while laboratories with a lower throughput or higher variation in patient population (typical for laboratories operating in a medium-size hospital) have a higher long-term variation in performance. Observations about drifts or shifts, or transient to long-term bias of the laboratory compared to its peer, or between different instruments used in a laboratory are very interesting. For example, shifts applied to several laboratories belonging to the same peer are an index that a major manufacturer event is occurred, i.e., a reagent or calibrator lot change. The major contribution of the "Empower" project is that it works with data generated from real patients samples and can be linked to observations in daily IQC practice. Obviously, it will not solve all of the problems that laboratory medicine face in trying to achieve interchangeability and accuracy of clinical laboratory test results, but the Empower database can become a source for "big data mining" and contribute to the definition of common reference intervals and reliable clinical decision limits. It is very important to emphasize that monitoring the patients medians is not a substitute for daily IQC but is a complementary observation tool from patient data that can cover much longer observations time. In a pilot study [12], for measurands with season-independent concentrations (i.e., calcium, phosphate, FT4, and TSH), it was demonstrated that laboratories can improve quality assurance by mid- to long-term quality management from patient percentiles. This additional IQC practice, in fact, may help to early alert the laboratory to potential problems with a direct impact and return on clinical decision making. The online access to the same data might promote the dialogue between laboratorians and manufacturers as the authors advocate the "Empower" project as a new integrated tool for quality management intended to establish cooperation between the two main stakeholders. This additional quality tool may help in the identification of root causes for analytical instability so that adequate corrective actions can be taken at the side of both the laboratories and/or the manufacturers.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Financial support: None declared.
Employment or leadership: None declared.

Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Plebani M. Harmonization in laboratory medicine: the complete picture. Clin Chem Lab Med 2013;51:741–51.
2. Miller WG, Myers GL, Gantwer ML, Kahan SE, Schönbrunner ER, Thienpont LM, et al. Roadmap for harmonization of clinical laboratory measurement procedures. Clin Chem 2011;57:1108–17.
3. Siekmann L. Metrological traceability – a concept for standardization in laboratory medicine. Clin Chem Lab Med 2013;51:953–7.
4. Directive 98/79/EC of the European Parliaments and of the Council of 27 October 1998 on in vitro diagnostic medical devices. L331. Off J Eur Communities 1998;41:1–37.
5. Braga F, Panteghini M. Verification of in vitro medical diagnostics (IVD) metrological traceability: responsibilities and strategies. Clin Chim Acta 2014;432:55–61.
6. Miller WG. The role of proficiency testing in achieving standardization and harmonization between laboratories. Clin Biochem 2009;42:232–5.
7. Sciacovelli L, Secchiero S, Zardo L, Zaninotto M, Plebani M. External quality assessment: an effective tool for Clinical Governance in laboratory medicine. Clin Chem Lab Med 2006;44:740–9.
8. Little RR, Rohlfing CL, Sacks DB for the National Glycohemoglobin Standardization Program (NGSP) Steering Committee. Status of hemoglobin A_{1c} measurements and goals for improvement: from chaos to order for improving diabetes care. Clin Chem 2011;57:205–14.
9. Mosca A, Weykamp C. Feasibility of an EQAS for HbA1c in Italy using fresh blood samples. Clin Chem Lab Med 2014;52:151–3.
10. Van Houcke SK, Van Aelst S, Van Uytfanghe K, Thienpont LM. Harmonization of immunoassays to the all-procedure trimmed mean – proof of concept by use of data from the insulin standardization project. Clin Chem Lab Med 2013;51:103–5.
11. Van Houcke SK, Thienpont LM. "Good samples make good assays" – the problem of sourcing clinical samples for a standardization project. Clin Chem Lab Med 2013;51:967–72.
12. Van Houcke SK, Stepman HC, Thienpont LM, Fiers T, Stove V, Couck P, et al. Long-term stability of laboratory tests and practical implications for quality management. Clin Chem Lab Med 2013;51:1227–31.
13. Stepman HC, Stöckl D, De Grande LA, Thienpont LM. Standardize the serum albumin assay now: calibrate it to 60% of the serum total protein assay. Clin Chem Lab Med 2013;51:151–2.
14. Stöckl D, Van Uytfanghe K, Van Aelst S, Thienpont LM. A statistical basis for harmonization of thyroid stimulating hormone immunoassays using a robust factor analysis model. Clin Chem Lab Med 2014;52:965–72.
15. De Grande LA, Goossens K, Van Uytfanghe K, Stöckl D, Thienpont LM. The Empower project – A new way of assessing and monitoring test comparability and stability. Clin Chem Lab Med 2015;53:1197–204.
16. Stepman HC, Stöckl D, Acheme R, Sesini S, Mazziotta D, Thienpont LM. Status of serum calcium and albumin measurement in Argentina assessed in 300 representative laboratories with 20 fresh frozen single donation sera. Clin Chem Lab Med 2011;49:1829–36.
17. Van Houcke SK, Rustad P, Stepman HC, Kristensen GB, Stockl D, Roraas TH, et al. Calcium, magnesium, albumin, and total protein measurement in serum as assessed with 20 fresh-frozen single-donation sera. Clin Chem 2012;58:1597–9.
18. Stepman HC, Stöckl D, Stove V, Fiers T, Couck P, Gorus F, et al. Long-term stability of clinical laboratory data—Sodium as benchmark. Clin Chem 2011;57:1616–7.
19. Stepman HC, Tiikkainen U, Stöckl D, Vesper HW, Edwards SH, Laitinen H, et al. Measurements for 8 common analytes in native sera identifies inadequate standardization among 6 routine laboratory assays. Clin Chem 2014;60:855–63.
20. Expert panel on detection, evaluation, and treatment of high blood cholesterol in adults. Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). JAMA 2001;285:2486–97.
21. Myers GL, Miller WG, Coresh J, Fleming J, Grrendberg N, Greene T, et al. for the National Kidney Disease Education Program Laboratory Working Group: recommendations for improving serum creatinine measurement – a report from the laboratory working group of the National Kidney Disease Education Program. Clin Chem 2006;52:5–18.
22. STT-Consulting – Empower – Master Comparison 2014. Available from: http://www.stt387consulting.com/news.php?rubriek=8/. Accessed on December 2014.
23. Nordic Reference Interval Project (NORIP). Available from: http://pweb.furst.no/norip/. Accessed on December 2014.

Corresponding author: Mario Plebani, CCLM Editor-in-Chief, Department of Laboratory Medicine, University-Hospital of Padua, Via Giustiniani 2, 35128, Padua, Italy, Phone: +39 0498212792, Fax: +39 049663240, E-mail: mario.plebani@unipd.it; and Centro di Ricerca Biomedica, University-Hospital of Padua, Padua, Italy
Sandra Secchiero: Department of Laboratory Medicine, University-Hospital of Padua, Padua, Italy; and Centro di Ricerca Biomedica, University-Hospital of Padua, Padua, Italy