



Opinion Paper

Matthias Orth*

Are regulation-driven performance criteria still acceptable? – The German point of view

DOI 10.1515/cclm-2014-1144

Received November 21, 2014; accepted March 9, 2015; previously published online April 3, 2015

Abstract: Performance criteria should be a challenge for the laboratories to improve their quality. In countries with mandatory proficiency testing, the definition of performance criteria is a particular issue. If the definition of performance criteria is mandated from the regulatory bodies to medico-scientific institutions, scientific approaches (i.e., based on biological variation), the state-of-the-art approach (i.e., based on technical feasibility) as well as medical needs can be used to set up performance criteria such as the Richtlinie der Bundesärztekammer (RiliBÄK; Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations) in Germany. The experiences with RiliBÄK show that these performance criteria have to be revised on an ongoing basis.

Keywords: external quality assessment scheme (EQAS); Health Technology Assessment; internal quality; quality management system.

Introduction

Access to high quality laboratory medicine services might be guaranteed solely by the dedicated quality control-oriented people working in the medical laboratory with guidance from global and national regulations [e.g., ISO 15189, CLIA and Richtlinie der Bundesärztekammer (RiliBÄK; Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations)]. Patients cannot judge the quality management systems of the medical laboratory; they have to rely on compliance to the

quality standards. Standardized, comparable patient laboratory results are necessary to guide diagnosis and treatment decisions ensuring patient safety.

Health Technology Assessment (HTA)

Laboratory medicine is one of the most innovative disciplines in medicine and every year new tests are introduced to potentially benefit patient care. However, with limited resources, the introduction of innovative technological solutions must be thoroughly assessed for their potential to improve efficiency, productivity, and safety. Health Technology Assessment (HTA) is the preferred tool used to assess the value of innovative technologies and their effectiveness and cost impact on the healthcare system. It is expected that with HTA, only innovative technologies that optimize patient care and improve outcomes will be approved for reimbursement [1]. The concept behind HTA is a prohibition of use unless permission is granted (positive list) and HTA in laboratory medicine can be regarded as the highest level of regulation-driven performance criteria. In general, HTA is performed by independent institutions employing scientific methods covering medical, social, economic and ethical issues. HTA relies heavily on outcome studies. Outcomes related to laboratory testing can influence clinically, operationally and economically. Measurable outcomes include: 1) availability (or lack) of testing on mortality, morbidity, incidence of disease, admission, readmission or discharge rate, or on quality of life; 2) method or the turnaround time of testing on the length of hospital stay or on patient satisfaction; 3) accuracy of the test on diagnosis or prognosis; and 4) test availability on cost per life-year [2]. However, in most countries, the primary aim of HTA is at decision-making in health politics and healthcare budgeting.

HTA assessment of efficiency and effectiveness of new technological solutions faces three main methodological challenges in general. The first is to verify that the evidence of efficiency gains and improvements in health

*Corresponding author: Priv.-Doz. Dr. med. Matthias Orth, Institut für Laboratoriumsmedizin, Vinzenz von Paul Kliniken gGmbH, Adlerstrasse 7, 70199 Stuttgart, Germany, E-mail: matthias.orth@vinzenz.de
<http://orcid.org/0000-0003-2881-8384>

remains valid when different definitions of health outcomes are used. Much of the evidence focuses on crude measurements such as life expectancy, failing to consider the quality of the years of life gained. The second challenge is to disentangle the relative influence of health systems on health outcomes from the impact of other determinants of population health. The third is the time lag between the introduction of a new technological solution and its impact on health outcomes, a problem that may involve ‘false savings’ because they may lead to increased costs or other unintended consequences in the long-term (e.g., screening tests with high rates of false-positives followed by extensive diagnostic procedures or even invasive treatment measures) as well as to undervalue positive effects of new technologies when outcomes can be detected only after long periods of observation such as in screening programs for low grade types of cancer or for risk markers for slowly progressing diseases such as coronary heart disease.

HTA and Stockholm recommendations

Fraser has proposed a hierarchical approach to set quality specifications derived on factors explaining biological and analytical variation [3]. On top of this hierarchy are assessments of the effect of analytical performance on clinical decision-making (outcome studies), second professional recommendations, third quality specifications laid down by regulation or external quality assessment scheme (EQAS) organizers and last, published data on the state-of-the-art. It is not surprising, that most work for setting quality specifications even lower in hierarchy than HTA or the good use of laboratory medicine (GULM) has been done so far on ‘simple chemical tests’ such as electrolytes, glucose, or Hb_{A1c} [4] leaving wide areas of laboratory medicine without widely-accepted recommendations [5] and the practical use of these recommendations has not reached the whole area of laboratory medicine due to numerous reasons, most of them not under sufficient control of the people working in the laboratory.

Challenges of HTA for diagnostic procedures

The current focus on HTA for introducing new technologies in healthcare might pose a severe threat to

innovations in laboratory medicine as well as for using established laboratory tests. The challenges in the practical use of HTA or assessments of the effect of analytical performance on clinical decision-making when applied to diagnostic procedures such as laboratory medicine testing are manifold.

Even more than in HTA of new treatment options such as drugs or medical devices, qualifying performance testing in the medical laboratory by HTA is a yet unresolved challenge [6] given the general concept of laboratory medicine which only delivers data to the attending physicians such as the presence or absence of a certain disease. Most meta-analyses for diagnostic test studies still pool diagnostic sensitivity and specificity values only [7]. The potential medical value is always indirect by changing the pathway of care such as by triggering or stopping a certain drug treatment, a surgical intervention or an additional diagnostic procedure such as a CT-scan. In this context, true randomized prospective trials with hard end points – the hallmark of HTA – are nearly impossible to perform [8] and the number of studies fulfilling the criteria for HTA are marginal [9]. As has been pointed out by Sandberg, diagnostic and analytical performance goals of a certain laboratory test might even have to be defined for different clinical situations and have to be revised in specified intervals thereafter [10]. HTA adds a further level of complexity to the concept of quality indicators and performance goals in the medical laboratory that not only analytical quality indicators have to be agreed on for tests but also for testing intervals.

Furthermore, many laboratory tests are used for the exclusion of certain diseases and the benefit of any of these tests used for this purpose is highly dependent on the prevalence in the respective population [11] as well as the availability of other diagnostic methods and the cost structure of the health system in this population. Noteworthy, the acceptance of a monetary gain of a certain medical procedure is not equally accepted, e.g., the concept of costs per QALY is accepted in some countries (with wide differences among countries) [12], but is not accepted and even considered to be unethical in Germany.

It is questionable whether in HTA schemes the value of a single laboratory test can be assessed on the background of a wide array of confounders in regard to other laboratory tests used in this condition and the heterogeneous treatment strategies applied to patients with a certain disease. Overall, the direct patient relevant value of a laboratory test or of tightened performance criteria on patient outcome can only very rarely be observed [13] and surrogate parameters will be used routinely [14].

Further challenges arise from companion diagnostics and from direct to consumer testing (DTC). In companion diagnostics, a certain test result of a (new) laboratory test is the prerequisite for the prescription of a drug [15]. For the regulation of the drug, the approval of the laboratory test is a sine qua non. There is substantial concern that the HTA of new laboratory tests will be shifted from laboratory medicine to the drug companies. One might even argue that setting performance goals for a blood count – a test frequently used in patients with chemotherapy – should therefore also be done by the drug companies. Another critical issue jeopardizing patients' safety is in DTC. Some believe that no quality criteria at all have to be followed if laboratory tests are performed by non-healthcare professionals allowing a free movement of services under the consumer rights directive 2011/83/EU [16].

For the assessment of the medical value of a certain laboratory test, other factors affecting the uncertainty of the test result have also to be addressed. Biological variation can be a massive confounder. However, for some tests in fact extensive preanalytical precautions (such as in regard of timing, food intake, body posture, physical activity and drug intake) can minimize these variation. In other tests, standardization, traceability, and improvement of the reagents used should be the focus that can diminish the bias between different tests of different vendors [5, 17]. The discussion is still open as to whether different analytical performance standards might be acceptable between central laboratory tests and point of care tests such as in the determination of cardiac peptides or creatinine [18, 19].

Reference method values are only available for 80–90 tests [20]. Another challenge with reference method values vs. specific method consensus values for EQAS might arise from the difficulties in obtaining quality controls which behave similar to patient samples. In fact, despite having reference method values available for glucose, the quality control materials for whole blood testing available cannot be used on all testing platforms and method specific consensus values must be used for those platforms where a certain quality control materials is not suited [21]. It is in the responsibility of the EQAS organizer to switch the matrix and composition of external quality control samples so that all testing platforms, irrespective of their manufacturer, can be compared to the reference method values. Commutable sample materials, if used in EQAS, are helpful to identify measurands in need of harmonization and for surveillance of the success of harmonization program tests [20].

'Richtlinie der Bundesärztekammer' (RiliBÄK) as example of regulation driven-performance criteria

Ideal analytical performance goals (Table 1) should be universally applicable, modifiable (e.g., reflecting newer technologies or recent scientific discoveries) and should be accompanied by a legal framework which allows a continuous, self-learning improvement of the system. Analytical performance criteria should be primarily driven by the patient's need leaving economical issues aside. Two opposite concepts might be employed to accomplish this goal. One concept focuses on science and altruistic motifs, the other concept stresses the role of the regulatory bodies. The major advantage of the first concept is that the whole process is kept within the scientific community and can be modified if deemed necessary. In the second concept, the regulatory bodies establish the quality standards, a procedure easily hampered by inflexibility when modification is necessary. In the following, the German experience, the so called RiliBÄK [22], are used as a blueprint for a concept with quality standards set up by a regulatory body.

The legal background behind the RiliBÄK is the EU IVD directive which was transformed into national legislation regulation by the German Medical Devices Act ('Medizinproduktegesetz') and the German Medical Devices Operator Ordinance ('Medizinproduktebetriebsverordnung'). The RiliBÄK are compiled by the German Medical Association ('Bundesärztekammer') as the designated body and every professional employing laboratory tests in human healthcare is obliged to comply with all regulations specified in the RiliBÄK.

The RiliBÄK consists of a part A (the description of a quality management system closely resembling DIN EN ISO norm 15189 as a framework for structural quality) and part B with extensive appendices covering analytical performance goals in internal as well as in external quality programs. Part B has several tables that contain

Table 1: Quality criteria to be covered by regulation.

-
- Performance criteria for daily routine quality controls
 - Performance criteria for EQAS
 - Performance criteria for tests with numeric as well as for alphanumeric results
 - Use of reference method values and/or method specific values for EQAS
 - Optional: minimum time interval/maximum frequency for ordering a specific test
-

the requirements for 84 selected quantitative and 50 semi-quantitative tests in hematology, coagulation, clinical chemistry, TDM, endocrinology, and serology in different matrices (such as serum, plasma, whole blood, urine, cerebrospinal fluid) as well as for genetic and microbiological tests and sperm analysis [22]. Tables for quantitative tests contain detailed information about the root mean square of measurement deviation (RMSMD) including the respective analytical range for this RMSMD. In some tests, different RMSMDs are given dependent on the concentration range. The maximum allowable deviation in EQAS is provided as well as whether the target value is obtained by a reference method or as a method-specific consensus value. RiliBÄK goals are primarily based on traceability to recognized primary standards and by the definition of uncertainty of measurement by determining the various components of total analytical error. A detailed comparison of RiliBÄK with the limits of other EQAS as well as with biological variation is given by Haeckel et al. [23]. In brief, limits set in RiliBÄK are based on the 90th, in some case of the 95th percentile of EQAS participants. Most tests demonstrate acceptable or even optimal coefficient of variation (CV) leaving only very few tests (i.e., tests with very low biological variation such as sodium) with an analytical quality unacceptable for clinical decision-making. The mathematical concept and the advantages of RMSMD have been discussed in detail previously [24–26]. In short, the major advantage in current practice is a rapid (i.e., instant, on-line) assessment of analytical control samples and the detection of critical deviations by the operator of the test system as soon as possible. The automatic calculation of RMSMD is integrated into all major laboratory information systems in Germany. A drawback of the use of the RMSMD might be that a violation of the minimum performance goals will not give an information on whether systematic (i.e., bias) or random (i.e., imprecision) issues have caused the violation. However, using the integrated plots and the automatic calculation of bias and imprecision (which has been mandatory in the previous RiliBÄK for the retrospective analysis and which are still available in the QM-modules of the laboratory information systems), a quick decision about the probable source of error can be done and corrections to the assay/system quickly implemented.

The key points of the RiliBÄK are the selection of tests for the appendices in part B and the specifications for RMSMD for internal quality control and the maximum allowable deviation in EQAS. Revised appendices have been published in 2014 [22] to update the originals from 2007. For serum testing few changes have been introduced such as CA 19-9 and lipase were removed from the appendix

and CA 15-3 and FSH added. In some immunochemical assays, the maximum allowable deviation and/or the RMSMD ranges were tightened (e.g., in free triiodthyronine from 24.0% to 20.0% maximum allowable deviation and from 14.5% to 13.0% for the RMSMD).

Of importance is the deliberate non-cross referencing of quality standards in medicine to a DIN EN ISO norm despite having similar content in RiliBÄK (part A) and in DIN EN ISO 15189:2014-11. First, the development of these norms is primarily driven by economic interests to allow free movement of goods and services between countries. In addition to other interested parties, medical scientific societies and medical associations can participate in the development of these norms but have no right to veto even when crucial issues are concerned. The technical committees in charge of these norms must not obey professional rules or medicolegal rules unlike state medical boards. Second, the use of DIN EN ISO norms in medicine has been widely refused since healthcare – being regulated by state law – is not within the regulatory scope of the European Union, i.e., norms, e.g., covering qualification issues of the personnel working in the laboratory such as in DIN EN ISO 15189:2014-11 might be in conflict with current legislation if the obedience of these norms is mandatory in medical laboratories.

Legal consequences of violating RiliBÄK criteria

Low-performing assay systems might pose a threat to patients' safety and therefore should be improved or, if improvement is not feasible, removed from market. Low performing assays can be detected by EQAS especially when target values are based on reference methods using commutable control samples. In RiliBÄK, assay systems with repeated failures in EQAS have to be reported to the IVD-regulatory bodies so that complaints to the manufacturer can be filed. This process has previously primarily relied on voluntary complaints from the users of the tests only and was therefore not sufficient for a qualified, reproducible detection of low-performing test systems [27].

This process has to be litigable because of the economic impact of erroneous removal of tests from the market. In this context, it might be difficult to use scientifically derived quality standards for this purpose since the scientific organization (without being a legal body) might be charged by certain manufactures for setting too tight standards. Therefore, the evaluation scheme of EQAS included in RiliBÄK with an optimized percentage

of true positives does not only minimize the number of test withdrawals (employing both a theoretical model and a practical evaluation of field data for setting the quality standards) but also has only minimum legal risks for the scientists involved in defining the quality standards.

Another advantage of the RiliBÄK concept is the universal application of these quality standards. Analytes not included in the appendix lists can be treated similar to listed tests. In case of accreditation (in Germany, about 95% of laboratories serving outpatients are accredited according to DIN EN ISO 15189) [28], the compliance of internal and external quality control schemes with the quality criteria will be scrutinized.

The discrimination between ‘professional recommendations’ and ‘quality specifications laid down by EQAS or by regulation’ might be misleading. In our understanding, EQAS can and should be organized by scientific societies. In particular in Germany, the limits set by RiliBÄK are as a whole more scientifically- and medically based than some recommendation of experts and are consistent with a compromise between the state-of-the-art and biological variation approaches [23]. Of particular importance is the independence, in particular from in vitro diagnostics manufacturers and from standard setting bodies, of the legal body in charge for setting the specifications both for the internal and the external quality control. The Medical Association as the legal body in charge to set off the specification does have the option to tighten quality standards of tests deemed to be medically insufficient. This can be done, e.g., by switching from method-specific values for EQAS to reference method values or by tightening the target value ranges or even setting performance standards to unreachable limits if a certain assay is clinically obsolete.

Conclusions

If a clinical utility of a new laboratory test is expected from the medical community, regulation-driven performance criteria for medical laboratory testing – even when based on analytical performance goals low in the hierarchy – might be a promising alternative to HTA if they are widely accepted both by medical professionals, laboratory specialists and from the health-economic network. The current focus on HTA by healthcare policy makers may pose a severe threat to the introduction of new laboratory tests for patient use. Regulation driven performance criteria have to be developed together with medical professionals. If referrals to DIN EN ISO norms are made, the

federal organization of the healthcare system has to be respected. Performance criteria should be established for a wide array of laboratory tests and updated on a regular basis. Results from EQAS testing can be used in a formalized process to revise performance goals.

Acknowledgments: I would like to thank Jessie Shih and Lieselotte Lennartz for many fruitful discussions.

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. EC-EPC. The 2012 ageing report: economic and budgetary projections for the 27 EU Member States (2010–2060). Brussels: European Union, 2012.
2. Oosterhuis WP, Bruns DE, Watine J, Sandberg S, Horvath AR. Evidence-based guidelines in laboratory medicine: principles and methods. *Clin Chem* 2004;50:806–18.
3. Fraser CG. General strategies to set quality specifications for reliability performance characteristics. *Scand J Clin Lab Invest* 1999;59:487–90.
4. Driskell OJ, Holland D, Hanna FW, Jones PW, Pemberton RJ, Tran M, et al. Inappropriate requesting of glycosylated hemoglobin (Hb A1c) is widespread: assessment of prevalence, impact of national guidance, and practice-to-practice variability. *Clin Chem* 2012;58:906–15.
5. Orth M, Aufenanger J, Hoffmann G, Hofmann W, Klosson R, Lichthagen R, et al. Recommendations for the frequency of ordering laboratory testing. *J Lab Med* 2014;38:231–8.
6. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *J Am Med Assoc* 1995;274:645–51.
7. Willis BH, Quigley M. Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review. *BMC Med Res Method* 2011;11:27.
8. Byron SK, Crabb N, George E, Marlow M, Newland A. The health technology assessment of companion diagnostics: experience of NICE. *Clin Canc Res* 2014;20:1469–76.
9. Czoski-Murray C, Lloyd Jones M, McCabe C, Claxton K, Oluboyede Y, Roberts J, et al. What is the value of routinely testing full blood count, electrolytes and urea, and pulmonary function tests before elective surgery in patients with no apparent clinical indication and in subgroups of patients with common comorbidities: a systematic review of the clinical and cost-effective literature. *Health Tech Ass* 2012;16:i–xvi,1–159.

10. Sandberg S, Thue G. Quality specifications derived from objective analyses based upon clinical needs. *Scand J Clin Lab Invest* 1999;59:531–4.
11. Houben PH, Winkens RA, van der Weijden T, Vossen RC, Naus AJ, Grof RP. Reasons for ordering laboratory tests and relationship with frequency of abnormal results. *Scand J Prim Health Care* 2010;28:18–23.
12. Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality-adjusted life year: in search of a standard. *Med Dec Making* 2000;20:332–42.
13. Mills NL, Churchhouse AM, Lee KK, Anand A, Gamble D, Shah AS, et al. Implementation of a sensitive troponin I assay and risk of recurrent myocardial infarction and death in patients with suspected acute coronary syndrome. *J Am Med Assoc* 2011;305:1210–6.
14. Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne JA, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *Br Med J* 2013;346:f457.
15. Khoury JD, Catenacci DV. Next-generation companion diagnostics: promises, challenges, and solutions. *Arch Pathol Lab Med* 2015;139:11–3.
16. Orth M, Lupp P. Direct-to-consumer-testing: Fluch oder Segen für die Patienten? *Dtsch Arztebl* 2015;112:A174–7.
17. Plebani M. Evaluating laboratory diagnostic tests and translational research. *Clin Chem Lab Med* 2010;48:983–8.
18. Mueller C, Scholer A, Laule-Kilian K, Martina B, Schindler C, Buser P, et al. Use of B-type natriuretic peptide in the evaluation and management of acute dyspnea. *New Engl J Med* 2004;350:647–54.
19. Straseski JA, Lyon ME, Clarke W, Dubois JA, Phelan LA, Lyon AW. Investigating interferences of a whole-blood point-of-care creatinine analyzer: comparison to plasma enzymatic and definitive creatinine methods in an acute-care setting. *Clin Chem* 2011;57:1566–73.
20. Miller WG, Tate JR, Barth JH, Jones GR. Harmonization: the sample, the measurement, and the report. *Ann Lab Med* 2014;34:187–97.
21. Wood WG. Problems and practical solutions in the external quality control of point of care devices with respect to the measurement of blood glucose. *J Diab Sci Tech* 2007;1:158–63.
22. Neufassung der Richtlinie der Bundesärztekammer zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen; Rili-BÄK; – Richtlinie der Bundesärztekammer zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen. *Dtsch Arztebl Int* 2014;111:1583–618. English version: Revision of the “Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK”. *J Lab Med* 2015;39:26–69.
23. Haeckel R, Wosniok W, Kratochvila J, Carobene A. A pragmatic proposal for permissible limits in external quality assessment schemes with a compromise between biological variation and the state of the art. *Clin Chem Lab Med* 2012;50:833–9.
24. Haeckel R, Wosniok W. Benefits of combining bias and imprecision in quality assurance of clinical chemical procedures. *J Lab Med* 2007;31:87–9.
25. Macdonald R. Quality assessment of quantitative analytical results in laboratory medicine by root mean square of measurement deviation. *J Lab Med* 2006;30:111–7.
26. Macdonald R. Reply to Haeckel and Wosniok: Root mean square of measurement deviation for quality assessment of quantitative analytical results in laboratory medicine compared to total error concepts. *J Lab Med* 2007;31:90–1.
27. Siekmeier R, Lutz J. Experience with post-market surveillance of in vitro diagnostic medical devices for lay use in Germany. *Clin Chem Lab Med* 2007;45:396–401.
28. Oosterhuis WP, Zerah S. Laboratory medicine in the European Union. *Clin Chem Lab Med* 2015;53:5–14.

Copyright of Clinical Chemistry & Laboratory Medicine is the property of De Gruyter and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.