Opinion Paper

Kenneth Sikaris*

Performance criteria of the post-analytical phase

DOI 10.1515/cclm-2015-0016
Received January 7, 2015; accepted March 24, 2015; previously published online April 18, 2015

Abstract: Quality in healthcare is ideally at an optimal benchmark, but must be at least above the minimal standards for care. While laboratory quality is ideally judged in clinical terms, laboratory medicine has also used biological variations and state-of-the-art criteria when, as is often the case, clinical outcome studies or clinical consensus are not available. The post-analytical phase involves taking quality technical results and providing the means for clinical interpretation in the report. Reference intervals are commonly used as a basis for data interpretation; however, laboratories vary in the reference intervals they use, even when analysis is similar. Reference intervals may have greater clinical value if they are both optimised to account for physiological individuality, as well as if they are harmonised through professional consensus. Clinical decision limits are generally superior to reference intervals as a basis for interpretation because they address the specific clinical concern in any patient. As well as providing quality data and interpretation, the knowledge of laboratory experts can be used to provide targeted procedural knowledge in a patient report. Most profoundly critically abnormal results should to be acted upon to minimise the risk of mortality. The three steps in quality report interpretation, (i) describing the abnormal data, (ii) interpreting the clinical information within that data and (iii) providing knowledge for clinical follow-up, highlight that the quality of all laboratory testing is reflected in its impact on clinical management and improving patient outcomes.

Keywords: clinical decision support; clinical outcome; post-analytical; quality; reference interval.

***Corresponding author: Kenneth Sikaris,** Sonic Healthcare, Melbourne Pathology, Collingwood, Victoria, Australia, E-mail: ken.sikaris@mps.com.au

Introduction

Quality can be defined as a standard of excellence that can vary from being unacceptably poor to exceeding expectations. In healthcare, we aim to maintain quality above the minimum standards of care, while more optimal standards or benchmarks are aspired to. Figure 1A shows a continuum of quality where various terminologies can be represented in relation to minimal or optimal standards. Even when optimal performance is not achieved, performance may still be acceptable as long as it is above the minimum standard. It is undesirable, however, to be too close to the minimum standard because a small drop in performance could become unacceptable.

These general quality standards can be transferred to a clinical framework (Figure 1B). By using the Hippocratic principle of 'primum non-nocere', harm is undesirable, if not unacceptable, and the ideal aim is perfect health (albeit not achievable for many patients).

The general quality standard framework is also similar to the levels of quality defined by biological variability theory (Figure 1C) [1]. Analytical variations, also referred to as measurement uncertainty, are usually defined as a the dispersion of results obtained for a single sample compared to the average of those measurements and summarised as the coefficient of analytical variation (CV_a). It is generally accepted that CV_a should never be so broad as to blur the true state of the patient. However, an individual patient also has day to day intraindividual biological variations (CV_i). When CV_a exceeds CV_i, it is impossible to tell if deviation in result is due to measurement errors or real changes in the patient's status. According to the biological variability theory, CV_a must be kept below CV_i, and the fraction usually found acceptable is $CV_a < 0.5\ CV_i$ [2].

When clinical quality standards have not been defined, or biological variation targets are not achievable, typical performance (or state of the art) may be used as the framework for quality (Figure 1D). Whether it is the best 25% or best 10% of laboratories that define a state of the art benchmark varies with the generally arbitrary nature of this framework. Similarly, whether the minimum standard is used to penalise a small number of laboratories in a regulatory framework or used to encourage a larger

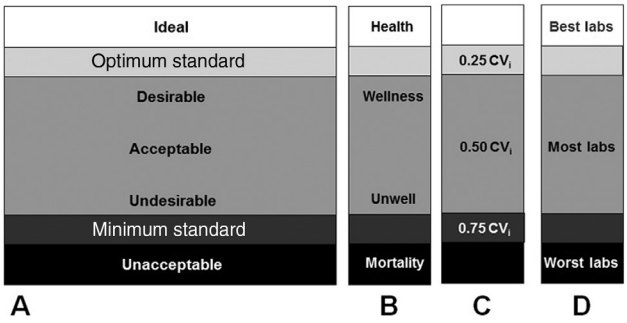

Figure 1: (A) General framework for quality using minimal and optimal standards to separate unnacceptable performance from acceptable and ideal performance levels. (B) Clinical outcome performance criteria where adverse outcomes are undesirable or unnacceptable, while health is the ideal state. (C) Biological variability performance criteria using 0.25/0.50/0.75 CV_i as optimal, desirable and minimal standards [1, 2]. (D) State-of-the-art performance criteria where the performance of most labs form the basis of acceptable quality, while the performance of the best labs is the optimal benchmark, and the worst labs define unacceptable performance.

number of laboratories to improve in a quality assurance framework is up to whoever seeks to define arbitrary state-of-the-art quality standards.

The post-analytical phase

The ISO15189 standard for medical laboratory quality [3] defines the post-analytical phase as the processes following the examination (which include review of results). The following processes include retention and storage of clinical material as well as disposal of the sample (and waste). In terms of the quality of pathology reports, however, the post-analytical phase includes the formatting, releasing, reporting and retention of the examination results for future access.

The results of technical analysis in clinical pathology are usually thought to be quantitative however, some measurements, e.g., serology and drug screening, are converted to an ordinal scale, e.g., negative, equivocal and positive. There are many quantitative tests that are interpreted on an ordinal scale including pregnancy tests (not pregnant/possibly pregnant/pregnant) and HbA_{1c} (healthy/pre-diabetic/diabetic/poorly controlled diabetic). Ordinal results are qualitative terms that have some sequential logic.

There are other types of qualitative results that cannot be ordered because there is no underlying sequential logic to the variety of the results. Examples include the results of serum protein electrophoresis where a particular pattern may indicate health vs. inflammation vs. myeloma vs. nephrotic syndrome, but these results do not represent a pathological sequence within patients or clinical severity sequence between patients. Such qualitative data are categorical and typically involve the identification of a distinct and independent pattern. Other categorical results include the interpretation of several hormone levels, e.g., TSH+fT4±fT3. Histopathology reporting can be considered a categorical classification of image data performed by human experts. The technical quality of the slides given to the histopathologist is analogous to the technical quality of the numerical data given to someone categorizing the numerical results.

While the quality of quantitative analysis can be measured as imprecision, bias, total error or measurement uncertainty, the quality of qualitative data cannot be measured in these ways.

The International Standard for Proficiency Testing (ISO 17043:2010) [4] in its Appendix A defines interpretive tests as a separate class of test to categorical qualitative data. It recognises that in proficiency testing, the quality of interpretation depends more on a participant's competence in identifying a pattern rather than a technical assessment of the laboratory, in general. As all individuals are fallible, Appendix B of the international standard [4] suggests that performance standards for qualitative data should ideally be evaluated by expert consensus (B3.2.1a). The standard also suggests the use of a five-point scale (5-Very Good, 4-Good, 3-Satisfactory, 2-Unsatisfactory, 1-Poor). These agreement scales are effectively 'Likert' [5] scales, which are often used to measure agreement between observers (5-Strongly agree, 4-Agree, 3-Neither agree nor disagree, 2-Disagree, 1-Strongly disagree) and have been applied in many interpretive areas of clinical medicine, which compare against expert interpretations such as radiology [6] and prescribing [7].

When a participant's interpretation is identical as a recognised group of experts, then performance is ideal. If the interpretation is not identical, then expert consensus is required to determine if the result is acceptable because, despite the interpretation differing, it may still lead to a similarly optimal clinical response. If the interpretation is different and will also lead to a different suboptimal clinical outcome, that interpretation is incorrect or unacceptable (see Table 1).

Interpretive comments are integral to histopathology, they are increasingly being considered in haematology [8], microbiology [9], genetics [10] and clinical chemistry [11], where it is generally desired by clinicians [12]. There is some evidence that they lead to improved outcomes

Table 1: Interpretive agreement 'Likert' scale.

Level	Interpretation	Definition
5	Ideal	The identical interpretation as the experts leading to optimal diagnosis or treatment
4	Acceptable	A different interpretation but one which would lead to the same optimal diagnosis or treatment
3	Intermediate	A different interpretation that may not lead to the same diagnosis or treatment
2	Incorrect	A different interpretation that leads to a diagnosis or treatment error
1	Unacceptable	A different interpretation that will lead to a major diagnosis or treatment error

compared to reports without comments [13]. The quality of interpretive commenting in clinical chemistry has been assessed by proficiency testing schemes and has found that unacceptable interpretation can be made [14–17] and lead to the conclusion that formal training of pathologists and clinical scientists should be provided [18–20], concentrating on how to comment as much as what to comment [21, 22].

The ideal interpretive comment [23]:

(i) describes the abnormalities in the technical data,
(ii) interprets that information including the clinical implications such as for diagnosis and
(iii) provides knowledge for follow-up including further testing or specialist referral.

Defining appropriate follow-up testing certainly lies within the expertise of senior clinical laboratory professionals so much so that 'reflex testing' is the term used for follow-up tests that are performed automatically by the laboratory in order to avoid unnecessary clinical delays [24–26].

In Stockholm in 1999, HMJ Goldschmidt highlighted that in the post-analytical phase, raw data, such as the numbers in a laboratory result, are converted to information when meaning is given to that data [27]. That data and information can then be related to an expert's knowledge base and experience (laboratorian and/or clinician) and convert to new procedural knowledge for that specific patient assisting medical decisions including treatment [19] The application of laboratory data and information to conceptual, strategic and procedural knowledge [28] is at the core of creating clinical value through pathology testing. The impact of test misinterpretation on patient safety ultimately lies with the treating clinician and can, therefore, be considered in the: post-post-analytical phase [29].

Reference limits and flagging abnormal results

The provision of a system for interpreting numerical data against reference limits or clinical decision values is a mandatory consideration in a pathology report (ISO 15189; 5.8.5.j) [3]. These interpretive limits are present in most clinical pathology reports. It is easy to underestimate the routine importance of these limits and any abnormal flags they generate for a busy clinician scanning dozens of reports and trying to pick out the salient points.

Reference intervals are typically statistical confidence limits for the typical spread of results to be found in a healthy reference population. There are some special forms of reference limits for substances not normally found in healthy people such as therapeutic ranges for drug levels, detection limits for toxins (or drugs of abuse), legal limits such as for alcohol.

In contrast to *reference intervals*, which are designed to confirm health (*absence* of *any* disease) with high specificity (typically 95%), *clinical decision limits* are more clinically focussed and generally aim to confirm the *presence* of *a particular* disease or clinical risk with appropriately high sensitivity. Receiver operator curves (ROC) have also gained some popularity as a method to balance specificity and sensitivity to create '*optimal*' *cut-offs*. ROC optimal cut-offs have reduced specificity compared to reference intervals and reduced sensitivity compared to clinical decision limits.

Because individuals vary so much in health, and in disease, both reference intervals and clinical decision limits can be personalised to apply to a particular individual. For example, hormone reference intervals can vary depending on the patient's gender and age, or the clinical decision point for the presence of insulin resistance may vary in pregnancy compared to non-pregnant adults. Personalised medicine ideally aims to incorporate as many relevant patient characteristics as possible into an interpretation within that clinical setting.

How can we judge the relative quality of these various approaches to cut-offs? Are clinical decision limits more useful than traditional reference intervals? Reference intervals were more commonly used 20 years ago [30, 31], but increasingly, laboratories no longer quote 'healthy' reference intervals for analytes such as cholesterol [32] because they each have clinical decision limits; it seems that the latter have priority. The principle that clinical decision limits – associated with risk and clinical outcome – are superior to reference intervals has a similarity to the Stockholm Consensus for defining analytical quality

[33]. This similarity has been reviewed in the context of developing a similar hierarchy for the quality of clinical decision limits and reference intervals [34]. It is logical that the quality of analytical measurement does not, by itself, define the quality of any laboratory report, when a poor quality reference interval can undermine the clinical value of a high-quality measurement.

A hierarchy for post-analytical quality criteria

The Stockholm hierarchy can be simplified to three quality criteria: (i) quality based on clinical outcome, (ii) quality based on biological variability and (iii) quality based on state of the art.

Post-analytical quality and state of the art

Using 'state of the art' as the basis for defining reference limits sounds ideal; however, it depends on what we mean by 'state of the art'. If we mean what is commonly done, then as most laboratories get their reference intervals from the manufacturer's kit insert, is that might what be the best thing to do? Well it might be, but it might not....

The Clinical Laboratory Standards Institute (CLSI) C28-A3 standard for reference intervals [35] was developed with the International Federation of Clinical Chemistry (IFCC). The standard states that the laboratory director can *transfer* a reference interval from, e.g., a kit insert, as long as they are confident of two things; first, that the analytical system is comparable and, second, that the test subject population is comparable. Unfortunately, this confidence is often not realised as, First, because analytical system may have changed platform, performance, calibration and/or traceability in the period since the test was established, and generally, the analytical system may be performing differently in the hands of the testing laboratory compared to the original reference laboratory. Second, confidence in the insert reference intervals is often not realised, as the test subjects (reference population) during the original study may be different to the reference population expected by the testing laboratory. Potential variances in populations, which are often not provided as details in kit inserts include age, gender, ethnicity, and often, measures are not taken to exclude disease, particularly obesity. It is not surprising, therefore, that most kit inserts often avoid calling these limits 'reference intervals' preferring to call them 'expected values' also adding statements such as, "Each laboratory should investigate the transferability of the expected values to its own population and if necessary determine its own reference ranges". It is for these particular reasons that the CLSI C28-A3 standard provides extensive guidance on how to validate these reference intervals using small or large numbers of reference intervals as an alternative to a formal reference interval study.

Despite using similar analytical traceability procedures, the reference intervals provided by manufacturers vary significantly and, therefore, probably highlight limitations of their reference interval studies [36]. The reference intervals used by laboratories vary more widely compared to analytical differences [37–41]. The unacceptable variation in reference limits between laboratories has led to professional initiatives for developing harmonisation of reference intervals [42–45], but this is certainly not a simple task [46, 47]. If laboratory testing methods are standardised (or can be harmonized), laboratories could potentially share reference interval data to make the reports more reliable [48]. It could be argued that the development of harmonised reference intervals endorsed by professional societies will require laboratories that use different intervals, to review them as Hyltoft Petersen explains [49], "Arguments for establishing common reference intervals are not needed. On the contrary, lack of such common reference intervals should be explained".

Post-analytical quality and biological variation

The use of biological variation as a basis for defining the quality reference intervals used in the post-analytical phase may seem a new idea but is an integral part of defining reference intervals. Reference intervals are, in fact a combination of three sources of variation, most obviously, interindividual biological variation (group variation or CV_g), but also including intraindividual biological variation (CV_i) as well as the analytical measurement uncertainty at the time of the study (CV_a).

The study of reference intervals is, therefore, the study of all these variations. However, it also encompasses fundamental philosophies that may not be immediately appreciated. You do not have to restrict your thinking to laboratory tests to appreciate that humans vary in their normal characteristics from one to another. The Gaussian distribution has for centuries been a

framework for describing the variation in human faculties [50].

The first time a patient has any measurement performed, we do not actually know if their result is 'normal' for them. Therefore, we use the spread of results in other apparently healthy individuals to judge if their result is unlikely to be normal for them.

The subsequent times a patient has measurements performed, as we already have a previous result, we should be less concerned if the new result is normal or not compared to others and more concerned with whether that new result has changed more than expected allowing for the usual biological variations expected in an individual from day to day. CV_i is the basis for reference change values.

When patients vary much more from one to another than they do individually from day to day, in other words $CV_i << CV_g$, reference intervals will lose their usefulness because a patient may have drifted too far from their own usual range of values, before they have moved out of the larger range of values probable for all individuals. The ratio of CV_i to CV_g is called the 'index of individuality', and reference intervals lose their usefulness if this index is below 0.6 [51–53]. The variation between individuals can be reduced if we group them into similar groups, such as men vs. women, young vs. old, pregnant vs. non-pregnant. This highlights the importance of creating physiologically specific reference intervals with similar groups being partitioned into their own specific reference interval. The usefulness of reference intervals depends crucially on understanding the physiological differences between groups and appropriately partitioning reference intervals in order to maximise the index of individuality [54].

In summary, the true study of reference intervals and their usefulness is inseparable from an understanding of biological variation.

Post-analytical quality and clinical outcome

The highest criterion for defining quality of analysis, or reference limits, is related to whether differences can be linked to adverse clinical outcomes for the patient. For example, we must have HbA_{1c} methods that can distinguish between HbA_{1c} of 53 mmol/mol (7.0%) and 64 mmol/mmol (8.0%) because the DCCT studies showed that these two values represent significantly different clinical outcome risks [55]. Similarly when a HbA_{1c} value of 48 mmol/mol (6.5%) was defined as a diagnostic threshold for diabetes [56] because of its association with, e.g., retinopathy, this value is also far more important than any attempt at defining the reference interval for HbA_{1c}. Finally, rather than defining a reference interval for HbA_{1c}, the next limit of interest is around 39 mmol/mol (5.6%) because that defines a prediabetic state that is associated with increased cardiovascular risk [57]. This example of HbA_{1c} illustrates both the strength and weakness of clinical decision limits. First, clinical decision limits make reference intervals redundant because we can focus on the diseases we are worried about rather than hypothetically trying to confirm 'health'. However, every clinical decision limit relates only to one particular clinical concern, and different limits may be needed for alternative clinical concerns. The difficulty in distinguishing prostate-specific antigen (PSA) clinical decision limits for benign prostatic hypertrophy vs. prostate cancer vs. prostatitis is probably why we have stuck to trying to improve PSA reference intervals through age-related partitioning and various PSA ratios.

Although clinical decision limits are ideally derived from formal clinical outcome studies, these are less common than those defined by consensus of clinicians. Recently, the International Association of Diabetes in Pregnancy Society Groups (IADPSG) established clinical decision limits for gestational diabetes [58] using the high-quality outcome data for the HAPO study [59]. However, even when good outcome studies are available, the selected arbitrary cut-offs selected are based on pragmatic considerations including what the consensus group negotiates to constitute a significant clinical risk along the continuum of risk [60].

Reference distributions derived from apparently healthy individuals are, nevertheless, indirectly associated with clinical risk. If a patient has a result outside the reference limits, they generally have an increased risk of morbidity and mortality. In fact, for PSA, the higher the PSA level is above the median value of the reference distribution, the risk of disease rises exponentially, while below that age-related median the risk is negligible [61]. Similar increases in clinical risk that start below the upper reference limit have been shown for many analytes including vitally important tests like cardiac troponin [62].

Critical risk limits and critical changes

The term critical limit is often poorly defined and may refer to either limits defining immediate high risk requiring

immediate attention (critical risk limit) or limits defining high risk that does not require immediate medical attention, but would benefit from a shorter reporting timeframe than routine results (significant risk limit) [63]. The most extreme measure of clinical outcome is mortality, and laboratories usually try to define critical risk limits to trigger the immediate notification of such results to clinicians. The methods laboratories use to establish their critical limits vary [64–67]. State of the art approaches are common including borrowing critical limits from other laboratories, critical limit surveys [68, 69] or the literature, in general [70].

Biological variation has not been formally used in this context. The ambiguous term 'critical difference' in biological variation discussions is not related to mortality considerations and has been superseded by the term 'reference change value' (or RCV). These statistically significant differences according to biological variability are not necessarily of 'critical' concern.

Ideally, critical risk limits should be based on clinical outcome studies that show that patients with results above that limit have an intolerable risk of mortality if left untreated. Here, we run into the same problem defining what an intolerable risk of mortality is. Each clinician may have a different opinion and it may vary according to each patient. Ideally, we are best defining critical risk limits in collaboration between laboratory and expert clinicians. It is quite interesting that the typical critical risk limits for sodium of <120 or >150 mmol/L [71] as well as potassium levels of <2.6 or >6.0 mmol/L, all represent approximately a 30% inpatient mortality risk [72].

The issue of critical risk limits demonstrates the quality required in reliable analytical data, the interpretation of that result against an agreed critical risk limit and, most importantly, the expected clinical responses required to improve clinical outcome. Without a clinical response, the data and interpretation are potentially of little value. It has been shown that some critical notifications, such as low albumin, rarely lead to clinical action, whereas others, such as high calcium, usually lead to immediate action [54]. The necessary clinical action may not eventuate if the result is analytically unreliable, and for calcium, we know that calcium is one of the few analytes that cannot meet biological variability (0.5 CV_i) goals, and its interpretation also suffers from increased uncertainty due to a variety of albumin adjustment formulae and albumin methods. Nevertheless, when clinicians are notified and acknowledge a critical calcium abnormality, the clinical actions are significant including treatment, further testing and a change in the diagnosis for 25% of these patients [73].

Post-analytical quality indicators

Surveys continue to show that most laboratory errors occur in the pre-analytical phase [74, 75]. When the International Federation of Clinical Chemistry (IFCC) working group on laboratory errors and patient safety defined a set of 25 laboratory quality indicators [76], the majority (16) were related to the pre-analytical phase, while only four were analytical but another five were post-analytical. This initial set was subsequently reviewed for clinical importance and applicability and four of the post-analytical indicators that remained as first priorities including transcription errors, turnaround time (TAT), incorrect reports and delay in critical result notification [77]. Defining performance criteria for these post-analytical indicators is problematic as an acceptable negative clinical impact of post-analytical errors may be difficult to define. As biological variation theory is also not relevant to these errors, the predominant performance criteria for these post-analytical indicators are based on state-of-the-art criteria such as the typical error rate in peer laboratories.

Post-analytical quality is the ultimate check on the coherence of the pre-analytical, analytical and post-analytical quality and the usefulness of the answer obtained in the context of the clinician patient interaction [69]. Many post-analytical errors such as dilution errors, calculations, QC failures, improper validation and incorrect units [78] could be argued as the final phase of analytical quality control, and a major function for validation systems is to identify pre-analytical and analytical errors [79]. TAT is similarly usually included as post-analytical quality issue [80]; however, the analytical TAT, including validation, is usually only a fraction of the complete diagnostic TAT [81], which includes pre-analytical collection and transport and the time to clinical review following report release. Clinician delays in reviewing results are a quality issue [82], but this falls under the laboratory's responsibility mainly in the context of defining critical risk limits or significant risk limits [83, 84].

Interpretive commenting, as a post-analytical quality indicator, has been given a lower priority largely because standardised methods to assess the quality of interpretation generally are not available and most existing assessment is educational [85]. There is little evaluation or audit of the post-analytical interpretive service [86], and this remains a grey area of responsibility between clinician and laboratory [87]. Although many laboratory accreditation standards include interpretability of reports in their checklists, this is often limited and narrow [88]. However, when medically qualified staff are employed within the laboratory to ensure the clinical quality of results, they

are ethically obliged, if not medico-legally responsible, to assist their clinical colleagues' care for patients. Performance criteria for interpretive commenting are, therefore, clinically focussed and generally rely on the opinions of experts in clinical interpretation, rather than accepting the commonest interpretation in a state-of-the-art approach.

Ensuring clinical value

While the quality of analysis is undoubtedly important, so too is the quality of the final report including its reference intervals, clinical interpretations and notifications. These contain the information and knowledge from laboratory specialists that should support clinical decision-making. Meaningful use criteria require the use of clinical decision support systems (CDSS) on high-priority health conditions to improve clinical quality measures, and simple CDSS tools may be associated with improved adherence to guidelines [89]. These include laboratory results and notifications and may lead to improved clinical outcomes [90].

Harms do arise from laboratory testing and include all phases from pre-analytical issues (such as inappropriate test ordering), analytical issues (such as inaccurate results), but also include the post-analytical issues such as misapplication of appropriate and accurate test results through cognitive failure [91]. A recent review showed that the quality gaps in laboratory medicine, as perceived in primary care, include not only delays but communication gaps, errors in judgement and cognition and a lack of patient centeredness [92].

Incorrect interpretation of diagnostic tests has been estimated as accounting for 37% of malpractice claims in primary care [93] and emergency departments [94]. The most common cognitive problems leading to fatal misdiagnosis involve faulty synthesis, particularly premature closure, i.e., the failure to continue considering reasonable alternatives after an initial diagnosis was reached [95]. There were typically six factors contributing to each case where harm occurred, and the breakdown in multiple barriers fit with Reason's 'Swiss Cheese' model of errors [96]. Laboratory tests and their misinterpretation are an important contributor to misdiagnosis because of the emphasis put on laboratory testing for diagnosis and monitoring decisions.

The impact of laboratory tests on clinical outcome can be summarised in a sequence of three questions [97]:

Does a laboratory test change the way a clinician thinks about a patient? Then if so:

Does that change in thinking alter the way the clinician manages the patient? Then if so:

Does that change in management affect clinical outcome (i.e., mortality/morbidity)?

There are some specific areas where laboratory interpretation has been of particular concern. The quality and quantity of post-analytical advice for therapeutic drug monitoring may be deficient with possible impacts on clinical decision-making [98]. Similarly, warfarin monitoring is of low quality when variation in post-analytical interpretations could have substantial effects of clinical action [99]. Variations in interpretation of what constitutes a significant change in diabetes monitoring with HbA_{1c} may also impact on treatment [100]. The expansion of genetic testing highlights that the reporting of nucleotide data is insufficient because this data must be interpreted to clearly answer the clinical question [101]. The focus on the clinical implications of a result for each particular patient and the increasing use of shared electronic clinical repositories will facilitate the practice of personalised medicine.

Conclusions

Ideally, the quality of laboratory report should be judged on its ability to answer the question(s) in the clinician's mind when requesting the test on that patient. Both quality analytical data and the interpretation of that data against the clinical context of that patient are crucial to quality in post-analytical interpretation. The quality of the post-analytical phase also reminds us that clinical laboratories should primarily aim to be clinically effective, by supporting clinical decision-making and ensuring improved outcomes for patients [102, 103]. Whenever clinical outcome criteria cannot be applied to post-analytical quality, other criteria including biological variability and state-of-the-art performance criteria can be considered.

Author contributions: The author has 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. Fraser CG. Biological variation: from principles to practice. Washington: AACC Press, 2001.
2. Cotlove E, Harris EK, Williams GZ. Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. III. Physiological and medical implications. Clin Chem 1970;16:1028–32.
3. ISO 15189:2012. Medical laboratories – requirements for quality and competence. Geneva, Switzerland: International Organization for Standardization, 2012.
4. ISO 17043:2010 Conformity assessment – general requirements for proficiency testing. Geneva, Switzerland: International Organization for Standardization, 2010.
5. Likert R. A technique for the measurement of attitudes. Arch Psychol 1932;140:1–55.
6. Rosenkrantz AB, Lim RP, Haghighi M, Somberg MB, Babb JS, Taneja SS. Comparison of inter-reader reproducibility of the prostate imaging reporting and data system and Likert scales for evaluation of multiparametric prostate MRI. AJR Am J Roentgenol 2013;201:W612–8.
7. Cooper JA, Ryan C, Smith SM, Wallace E, Bennett K, Cahir C, et al. The development of the PROMPT (PRescribing Optimally in Middle-aged People's Treatments) criteria. BMC Health Serv Res 2014;14:484.
8. Plebani M. Total quality in laboratory medicine: the case of haematology and coagulation testing. Int Jnl Lab Hem 2007;29(Suppl 1):36.
9. Cunney R, Aziz HA, Schubert D, McNamara E, Smyth E. Interpretative reporting and selective antimicrobial susceptibility release in non-critical microbiology results. J Antimicrob Chemother 2000;45:705–8.
10. Walley T. Evaluating laboratory diagnostic tests. BMJ 2008;336:569–70.
11. Kilpatrick ES, Freedman DB. National clinical biochemistry audit group. A national survey of interpretative reporting in the UK. Ann Clin Biochem 2011;48:317–20.
12. Barlow IM. Are biochemistry interpretative comments helpful? Results of a general practitioner and nurse practitioner survey. Ann Clin Biochem 2008;45:88–90.
13. Kilpatrick ES. Can the addition of interpretative comments to laboratory reports influence outcome? An example involving patients taking thyroxine. Ann Clin Biochem 2004;41:227–9.
14. Vasikaran SD, Penberthy L, Gill J, Scott S, Sikaris KA. Review of a pilot quality-assessment program for interpretative comments. Ann Clin Biochem 2002;39:250–60.
15. Lim EM, Vasikaran SD, Gill J, Calleja J, Hickman PE, Beilby J, et al. A discussion of cases in the 2001 RCPA-AQAP chemical pathology case report comments program. Pathology 2003;35:145–50.
16. Lim EM, Sikaris KA, Gill J, Calleja J, Hickman PE, Beilby J, et al. Quality assessment of interpretative commenting in clinical chemistry. Clin Chem 2004;50:632–7.
17. Vasikaran SD. Anatomy and history of an external quality assessment program for interpretative comments in clinical biochemistry. Clin Biochem 2014 Dec 24.doi: 10.1016/j.clinbiochem.2014.12.014. [Epub ahead of print].
18. Li P, Challand GS. Experience with assessing the quality of comments on clinical biochemistry reports. Ann Clin Biochem 1999;36:759–65.
19. Laposata M. Patient-specific narrative interpretations of complex clinical laboratory evaluations: who is competent to provide them? Clin Chem 2004;50:471–2.
20. Vasikaran SD, Lai LC, Sethi S, Lopez JB, Sikaris KA. Quality of interpretative commenting on common clinical chemistry results in the Asia-Pacific region and Africa. Clin Chem Lab Med 2009;47:963–70.
21. Marshall WJ, Challand GS. Provision of interpretative comments on biochemical report forms. Ann Clin Biochem 2000;37:758–63.
22. Challand GS, Vasikaran SD. The assessment of interpretation in clinical biochemistry: a personal view. Ann Clin Biochem 2007;44:101–5.
23. Vasikaran S. Interpretative commenting. Clin Biochem Rev 2008;29(Suppl 1):S99–103.
24. Paterson JR, Paterson R. Reflective testing: how useful is the practice of adding on tests by laboratory physicians. J Clin Pathol 2004;57:272–5.
25. Jones BJ, Twomey PJ. Comparison of reflective and reflex testing for hypomagnesaemia in severe hypokalaemia. J Clin Pathol 2009;62:816–9.
26. Srivastava R, Bartlett WA, Kennedy IM, Hiney A, Fletcher C, Murphy MJ. Reflex and reflective testing: efficiency and effectiveness of adding on laboratory tests. Ann Clin Biochem 2010;47:223–7.
27. Goldschmidt HM. Post-analytical factors and their influence on analytical quality specifications. Scand J Clin Lab Invest 1999;59:551–4.
28. Payne PR, Mendonca EA, Johnson SB, Starren JB. Conceptual knowledge acquisition in biomedicine: a methodological review. J Biomed Inform 2007;40:582–602.
29. Laposata M, Dighe A. "Pre-pre" and "post-post" analytical error: high-incidence patient safety hazards involving the clinical laboratory. Clin Chem Lab Med 2007;45:712–9.
30. Laker MF, Reckless JP, Betteridge DJ, Durrington PN, Miller JP, Nicholls DP, et al. Laboratory facilities for investigating lipid disorders in the United Kingdom: results of the British hyperlipidaemia association survey. J Clin Pathol 1992;45:102–5.
31. Engel JA, Petersen EC, Wilson JE, McManus BM. Progress in blood lipid reporting practices by clinical laboratories in North America. Changes from 1985 to 1990. Arch Pathol Lab Med 1992;116:229–34.
32. Hutchesson AC, O'Kane MJ, Neely RD, Oleesky DA. Provision of laboratory services for lipid analysis in the United Kingdom. Ann Clin Biochem 2007;44:273–80.
33. Kenny D, Fraser CG, Hyltoft Petersen P, Kallner H. Consensus agreement (strategies to set global analytical quality specifications in laboratory medicine – Stockholm). Scand J Clin Lab Invest 1999;59:585.
34. Sikaris K. Application of the Stockholm hierarchy to defining the quality of reference intervals and clinical decision limits. Clin Biochem Rev 2012;33:141–8.
35. Clinical and Laboratory Standards Institute (CLSI). Defining, establishing, and verifying reference intervals in the clinical laboratory: approved guideline, 3rd ed. CLSI Document C28-A3 (ISBN 1-56238-682-4). Clinical and Laboratory Standards Institute, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2008.
36. Koumantakis G, Jones G, Tate J. AACB pathology harmonisation organising committee, AACB quality initiatives in pathology – manufacturers reference intervals for common chemistry analytes. Clin Biochem Rev 2011;32:S31.

37. Sonntag O. Is this normal? – This is normal! Implication and interpretation of the so-called normal value. J Lab Med 2003;27:302–10.
38. Jones GR, Barker A, Tate J, Lim CF, Robertson K. The case for common reference intervals. Clin Biochem Rev 2004;25:99–104.
39. Cembrowski GS. Beyond analytical quality: the importance of post-analytical quality in assuring value. Proceedings of the Canadian Society of Clinical Chemistry (CSCC), Saskatoon, June 2010.
40. Berg J, Lane V. Pathology harmony; a pragmatic and scientific approach to unfounded variation in the clinical laboratory. Ann Clin Biochem 2011;48:195–7.
41. Jones GR, Koetsier SD. RCPAQAP first combined measurement and reference interval survey. Clin Biochem Rev 2014;35:243–9.
42. Rustad P, Felding P, Lahti A. Nordic reference interval project 2000. Proposal for guidelines to establish common biological reference intervals in large geographical areas for biochemical quantities measured frequently in serum and plasma. Clin Chem Lab Med 2004;42:783–91.
43. Reed M. The New Zealand approach to harmonised reference intervals. Clin Biochem Rev 2012;33:115–8.
44. Berg J. The UK pathology harmony initiative; the foundation of a global model. Clin Chim Acta 2014;432:22–6.
45. Koerbin G, Sikaris KA, Jones GR, Ryan J, Reed M, Tate J. AACB committee for common reference intervals. Evidence-based approach to harmonised reference intervals. Clin Chim Acta 2014;432:99–107.
46. Ceriotti F. Prerequisites for use of common reference intervals. Clin Biochem Rev 2007;28:115–21.
47. Tate JR, Sikaris KA, Jones GR, Yen T, Koerbin G, Ryan J, et al. Harmonising adult and paediatric reference intervals in Australia and New Zealand: an evidence-based approach for establishing a first panel of chemistry analytes. Clin Biochem Rev 2014;35:213–35.
48. Klee GG. Clinical interpretation and reference intervals and reference limits. A Plea for assay harmonisation. Clin Chem Lab Med 2005;42:752–7.
49. Hyltoft Petersen P, Rustad P. Prerequisites for establishing common reference intervals. Scand J Clin Lab Invest 2004;64:285–92.
50. Sikaris KA. Biochemistry on the human scale. Clin Biochem Rev 2010;31:121–8.
51. Harris EK. Effects of intra- and inter-individual variation on the appropriate use of normal ranges. Clin Chem 1974;20:1535–42.
52. Petersen PH, Sandberg S, Fraser CG, Goldschmidt H. Influence of index of individuality on false positives in repeated sampling from healthy individuals. Clin Chem Lab Med 2001;39:160–5.
53. Fraser CG. Inherent biological variation and reference values. Clin Chem Lab Med 2004;42:758–64.
54. Sikaris KA. Physiology and its importance for reference intervals. Clin Biochem Rev 2014;35:3–14.
55. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993;329:977–86.
56. International Expert Committee. International expert committee report on the role of the A1C assay in the diagnosis of diabetes. Diabetes Care 2009;32:1327–34.
57. Selvin E, Steffes MW, Zhu H, Matsushita K, Wagenknecht L, Pankow J, et al. Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. N Engl J Med 2010;362:800–11.
58. International Association of Diabetes and Pregnancy Study Groups Consensus Panel, Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010;33:676–82.
59. HAPO Study Cooperative Research Group, Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, et al. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008;358:1991–2002.
60. McIntyre HD, Metzger BE, Coustan DR, Dyer AR, Hadden DR, Hod M, et al. Counterpoint: establishing consensus in the diagnosis of GDM following the HAPO study. Curr Diab Rep 2014;14:497.
61. Vickers AJ, Cronin AM, Björk T, Manjer J, Nilsson PM, Dahlin A, et al. Prostate specific antigen concentration at age 60 and death or metastasis from prostate cancer: case-control study. BMJ 2010;341:c4521.
62. Omland T, de Lemos JA, Sabatine MS, Christophi CA, Rice MM, Jablonski KA, et al. Prevention of Events with Angiotensin Converting Enzyme inhibition (PEACE) trial investigators. A sensitive cardiac troponin T assay in stable coronary artery disease. N Engl J Med 2009;361:2538–47.
63. White GH, Campbell CA, Horvath AR. Is this a critical, panic, alarm, urgent or markedly abnormal result? Clin Chem 2014;60:1589–70.
64. Howanitz PJ, Steindel SJ, Heard NV. Laboratory critical values policies and procedures: a college of American pathologists Q-Probes study in 623 institutions. Arch Pathol Lab Med 2002;126:663–9.
65. Tillman J, Barth JH. ACB national audit group. A survey of laboratory 'critical (alert) limits' in the UK. Ann Clin Biochem 2003;40:181–4.
66. Zeng R, Wang W, Wang Z. National survey on critical values notification of 599 institutions in China. Clin Chem Lab Med 2013;51:2099–107.
67. Campbell CA, Horvath AR. Harmonization of critical result management in laboratory medicine. Clin Chim Acta 2014;432:135–47.
68. Wagar EA, Friedberg RC, Souers R, Stankovic AK. Critical values comparison: a college of American pathologists Q-Probes survey of 163 clinical laboratories. Arch Pathol Lab Med 2007;131:1769–75.
69. Piva E, Sciacovelli L, Laposata M, Plebani M. Assessment of critical values policies in Italian institutions: comparison with the US situation. Clin Chem Lab Med 2010;48:461–8.
70. Table of critical limits. Medical Laboratory Observer. Accessed 6–7 August, 2009.
71. Guerin MD, Martin AL, Sikaris KA. Change in plasma sodium associated with mortality. Clin Chem 1992;38:317.
72. Sikaris KA, Martin A, Guerin MD. Relationship between reference intervals and mortality-based reference ranges. Clin Biochem Rev 1991;12:81.
73. Howanitz PJ, Cembrowski GS. Post-analytical quality improvement: a college of American pathologists Q- Probes study of elevated calcium results in 525 institutions. Arch Pathol Lab Med 2000;124:504–10.
74. Plebani M. The detection and prevention of errors in laboratory medicine Ann Clin Biochem 2010;47:101–10.
75. Kumar SA, Jayanna P, Prabhudesai S, Kumar A. Evaluation of quality indicators in a laboratory supporting tertiary cancer care facilities in India. Lab Med 2014;45:272–7.

76. Sciacovelli L, Plebani M. The IFCC working group on laboratory errors and patient safety. Clin Chim Acta 2009;404:79–85.
77. Plebani M, Astion ML, Barth JH, Chen W, de Oliveira Galoro CA, Escuer MI, et al. Harmonization of quality indicators in laboratory medicine. A preliminary consensus. Clin Chem Lab Med. 2014;52:951–8.
78. Abdollahi A, Saffar H, Saffar H. Types and frequency of errors during different phases of testing at a clinical medical laboratory of a teaching hospital in Tehran, Iran. N Am J Med Sci 2014;6:224–8.
79. Guidi GC, Poli G, Bassi A, Giobelli L, Benetollo PP, Lippi G. Development and implementation of an automatic system for verification, validation and delivery of laboratory test results. Clin Chem Lab Med 2009;47:1355–60.
80. Salinas M, López-Garrigós M, Santo-Quiles A, Gutierrez M, Lugo J, Lillo R, et al. Customising turnaround time indicators to requesting clinician: a 10-year study through balanced scorecard indicators. J Clin Pathol 2014;67:797–801.
81. Ervasti M, Penttilä K, Siltari S, Delezuch W, Punnonen K. Diagnostic, clinical and laboratory turnaround times in troponin T testing. Clin Chem Lab Med 2008;46:1030–2.
82. Rodríguez-Borja E, Villalba-Martínez C, Carratalá-Calvo A. Enquiry time as part of turnaround time: when do our clinicians really consult our results? J Clin Pathol 2014;67:642–4.
83. Rizk MM, Zaki A, Hossam N, Aboul-Ela Y. Evaluating laboratory key performance using quality indicators in Alexandria university hospital clinical chemistry Laboratories. J Egypt Public Health Assoc 2014;89:105–13.
84. Kirchner MJ, Funes VA, Adzet CB, Clar MV, Escuer MI, Girona JM, et al. Quality indicators and specifications for key processes in clinical laboratories: a preliminary experience. Clin Chem Lab Med 2007;45:672–7.
85. Sciacovelli L, Zardo L, Secchiero S, Zaninotto M, Plebani M. Interpretative comments and reference ranges in EQA programs as a tool for improving laboratory appropriateness and effectiveness. Clin Chim Acta 2003;333:209–19.
86. Barth JH. Clinical quality indicators in laboratory medicine: a survey of current practice in the UK. Ann Clin Biochem 2011;48:238–40.
87. Goldschmidt HM. The NEXUS vision: an alternative to the reference value concept. Clin Chem Lab Med 2004;42:868–73.
88. Hawkins R. Managing the pre- and post-analytical phases of the total testing process. Ann Lab Med 2012;3:5–16.
89. Lau B, Overby CL, Wirtz HS, Devine EB. The association between use of a clinical decision support tool and adherence to monitoring for medication-laboratory guidelines in the ambulatory setting. Appl Clin Inform 2013;4:476–98.
90. Mishuris RG, Linder JA, Bates DW, Bitton A. Using electronic health record clinical decision support is associated with improved quality of care. Am J Manag Care 2014;20:e445–52.
91. Epner PL, Gans JE, Graber ML. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. BMJ Qual Saf 2013;22(Suppl 2):ii6–10.
92. Smith ML, Raab SS, Fernald DH, James KA, Lebin JA, Grzybicki DM, et al. Evaluating the connections between primary care practice and clinical laboratory testing: a review of the literature and call for laboratory involvement in the solutions. Arch Pathol Lab Med 2013;137:120–5.
93. Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. Ann Intern Med 2006;145:488–96.
94. Kachalia A, Gandhi TK, Puopolo AL, Yoon C, Thomas EJ, Griffey R, et al. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. Ann Emerg Med 2007;49:196–205.
95. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med. 2005;165:1493–9.
96. Reason J. Human error. New York, NY: Cambridge University Press, 1990.
97. Berte LM, Nevalainen DE. The laboratory's role in assessing patient outcomes. Lab Med 1998;29:114–9.
98. Norris RL, Martin JH, Thompson E, Ray JE, Fullinfaw RO, Joyce D, et al. Current status of therapeutic drug monitoring in Australia and New Zealand: a need for improved assay evaluation, best practice guidelines, and professional development. Ther Drug Monit 2010;32:615–23.
99. Kristoffersen AH, Thue G, Sandberg S. Post-analytical external quality assessment of warfarin monitoring in primary healthcare. Clin Chem 2006;52:1871–8.
100. Skeie S, Perich C, Ricos C, Araczki A, Horvath AR, Oosterhuis WP, et al. Post-analytical external quality assessment of blood glucose and hemoglobin A1c: an international survey. Clin Chem 2005;51:1145–53.
101. Claustres M, Kožich V, Dequeker E, Fowler B, Hehir-Kwa JY, Miller K, et al. Recommendations for reporting results of diagnostic genetic testing (biochemical, cytogenetic and molecular genetic). Eur J Hum Genet 2014;22:160–70.
102. Piva E, Plebani M. Interpretative reports and critical values. Clin Chim Acta 2009;404:52–8.
103. Page EP, Woodcock SM. The clinical laboratory of the future: re-engineering laboratory services. Can J Med Technol 1994;56:155–60.