From laboratory instrumentation to physician's brain calibration: the next frontier for improving diagnostic accuracy?

Giuseppe Lippi[1], Gianfranco Cervellin[2]

[1]Section of Clinical Biochemistry, University of Verona, Verona, Italy; [2]Emergency Department, University Hospital of Parma, Parma, Italy
Correspondence to: Prof. Giuseppe Lippi. Section of Clinical Biochemistry, University Hospital of Verona, Piazzale LA Scuro, 37100, Verona, Italy. Email: giuseppe.lippi@univr.it.

Abstract: Instrument calibration is a common practice in laboratory medicine aimed to translating electronic or optical signals into clinically usable concentrations for many analytes. Although much technological advancements have contributed to make the analytical process a much safer enterprise, the risk of analytical errors, including the use of unacceptable analytical calibrations, has not been completely voided. There are many weapons that we can use to lower the risk of overlooking calibration errors in laboratory medicine, such as the use of standard operating procedures (SOPs), internal quality controls (IQCs) and automatic inactivation of analyzer functioning in case of IQC failure. Unlike instrument calibration, the clinical decision making strongly relies on cognitive skills, and involves innate ability, expertise and even luck. As for many other human activities, training of healthcare staff and optimization of clinical interpretation through a process of "physicians' calibration" are effective to lower the risk of medical and diagnostic errors. Physicians' calibration can hence be defined as a set of activities aimed to define, under specified clinical conditions, the correlation between specific signals (signs and symptoms) present in a given patient and the likelihood of a certain pathology, so aligning the clinical decision making with the most accurate diagnosis. However, the possible solutions for improving physicians' calibration are challenging, since humans are not machines, and the learning process is much more difficult than using neural networks. Moreover, there are no universal "standard materials" that can be used in clinics, since there is no patient completely similar to another. Therefore, achieving a substantial improvement of diagnostic accuracy needs a multifaceted strategy, entailing recovered enthusiasm on traditional clinical skills teaching, systematic exploration of new educational approaches, a facilitating process for more effective teamwork and, finally, substantial focus of investments in basic science of clinical diagnosis.

Keywords: Errors; calibration; diagnosis; Gestalt

The testing process

Laboratory medicine is a clinical discipline centered on generating diagnostic information from values of measurable analytes, which are supposed to assist clinicians in diagnosis, prognostication, therapeutic monitoring and follow up of a kaleidoscope of human pathologies. Since the early 1980s, the total testing process has been conventionally divided in five main domains, thus entailing the pre-preanalytical (i.e., test ordering), preanalytical, analytical, postanalytical and post-postanalytical (i.e., test interpretation) phases (*Figure 1*) (1). After more than a century of experience in diagnostic testing, it is now rather clear that most problems, and thereby the risks of error, are more likely to involve some manually-intensive activities of the preanalytical phase (between 60–70% of total), followed by the postanalytical (around 15–20% of total) and analytical (between 10–15%) phases (2).
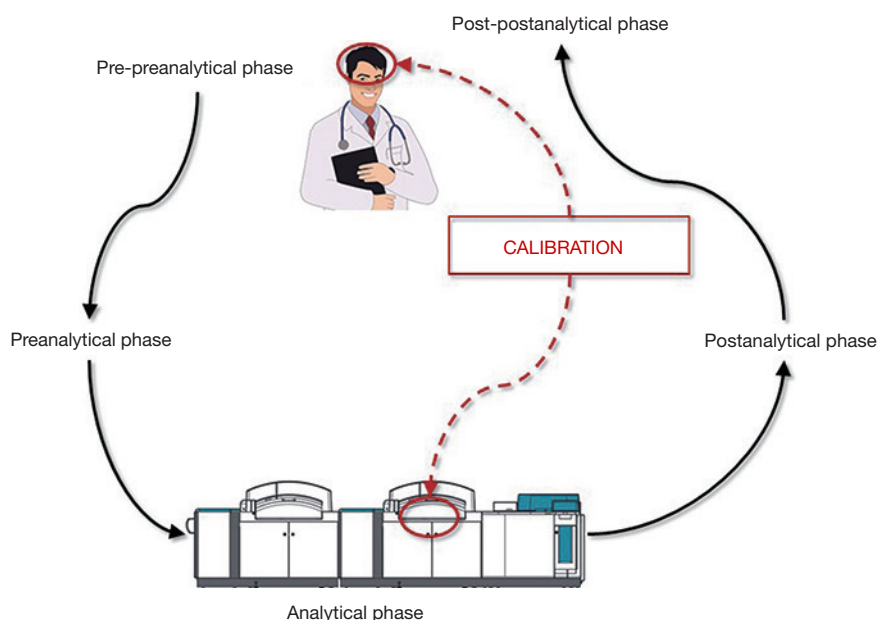
Figure 1 Development of the total testing process.

Calibration in laboratory medicine

Although the many technological advancements occurred over the past decades have contributed to make the analytical process a much safer enterprise, the risk of analytical errors has not been completely voided (1). Analytical errors can still occur as a result of analytical interference (e.g., for presence of interfering substances in the test sample, so including high values of bilirubin, turbidity, hemolysis, heterophilic antibodies, etc.), cross-reaction of antibodies between structurally similar molecules, deterioration of reagents or consumables, technical failures, instrument drift or calibration errors (3). As regards the last issue, calibration is a virtually unavoidable process in laboratory medicine, and is conventionally defined as a set of activities aimed to define, under specified operating conditions, the correlation between specific signals (electric or optical) generated by a given measuring system and analyte concentration, according to a process of translation based on comparison with standard materials characterized by predefined (and thereby "true") values (4). Therefore, the results of a calibration process should allow assigning clinically useful values to the measured analyte, with the obvious expectation that these values should be as close as possible to the true value of that measurable analyte.

The estimation of the real frequency of calibration errors, and of their potential impact on patient safety, is not an easy task. This is mainly due to the fact that each calibration process is then followed by measurement of internal quality controls (IQCs), which help identifying wrong or inaccurate calibrations and should lead to aborting the ensuing analytical session when there is clear evidence that calibration has failed. Yet, an interesting study showed that analytical inaccuracy not recognized, thus including calibration failures, may contribute to over 10% of all laboratory errors (5). In line with these findings, the National Institute of Standards and Technology (NIST) claims that calibration errors associated with the risk of generating analytic bias may have a negative impact on the overall number of patients overcoming decision thresholds in practice guidelines (4). These considerations would drive us to conclude that calibration errors still exist in laboratory medicine, and their impact on patient management is not meaningless. Nevertheless, there are many weapons that we can use to lower the risk of overlooking calibration errors in laboratory medicine. First and foremost, development, implementation and strict monitoring of adherence to standard operating procedures (SOPs) are invaluable means to standardize activities and help lowering the risk of trivial errors such as those due to non-compliance with IQCs data. Automatic inactivation of analyzer functioning in case of IQC failure is another important perspective, which may be seen as a viable means for by-passing human errors.

Laboratory testing has reached a pervasive significance in

clinical practice. The "testing imperative" can often become addictive for clinicians, although it is now clearly recognized that diagnostic testing is not devoid of limitations and unfavorable consequences. One among these, and probably the less recognized, is the illusory reassurance that normal test results, especially occurring when clinicians exclusively rely on lab test results, may almost entirely replace clinical examination (6).

Calibration of clinical decision making

Regardless of these important conclusions, the clinical decision making based on results of diagnostic tests still requires human (i.e., physician's) interpretation, which is hence another and potentially less controllable source of variability, typically belonging to the domain of the so-called post-postanalytical phase (7) (*Figure 1*). In an interesting editorial, just recently published by the *Journal of the American Medical Association* (*JAMA*), Adam S. Cifu has debated the intriguing issue that findings collected from history and physical examination may be both imprecise and inaccurate, so leading to a high risk of diagnostic inaccuracy (8). This may be due to either patient's or physician's issues. There is very little we can do for improving the former. As regards the latter circumstance, making a diagnosis is more or less like any other human activity, thereby strongly relying on innate skill, expertise and, occasionally, on luck. None of us will ever be able to write a tragedy like William Shakespeare, a symphony like Ludwig van Beethoven, or even driving a car like Ayrton Senna. All these men are well-recognized "outliers" in their disciplines and their incomparable skill is virtually unreachable. Nevertheless, training is often effective to attenuate the difference between a modest and an excellent performer, throughout all human activities. Medicine makes no exception to this rule.

It is clear and unquestionable that perception plays an essential role in physician's ability to make early diagnosis, even without support from diagnostic testing (9). There will always be physicians able to make an accurate diagnosis much earlier than their colleagues. Diagnostic algorithms, recommendations and even guidelines are valuable contributions, although they only provide a general guidance, since different patients with the same pathology may have rather different signs and symptoms (acute myocardial infarction is a paradigmatic example) (10). Therefore, experience remains central for developing and improving clinical competency. As brilliantly underscored by Adam S. Cifu (8), amplifying clinical experience is

more or less like calibrating physicians' brain to correct interpretation of what they can see, hear, touch and even smell. Unlike instrument calibration, however, the possible approaches for improving "physicians' calibration" are more challenging, since humans are not machines, and the learning process is more difficult than using neural networks or expert systems, which easily and rapidly learn by themselves. Moreover, there are no universal "standard materials" that can be used in clinics since, as previously discussed, there is no patient completely similar to another due to the influence of age, sex, comorbidities, pain sensitivity, etc. (11). Physicians' calibration can hence be defined as a set of activities aimed to define, under specified clinical conditions, the correlation between specific signals (signs and symptoms) present in a given patient and the likelihood of a certain pathology. The results of this unusual calibration process should hence allow aligning the clinical decision making with an accurate diagnosis. Quite understandably, this is not an easy enterprise, begging the obvious search for reliable approaches aimed to improving the individual diagnostic ability.

Some computer-assisted diagnostic systems have been proposed for reducing misdiagnosis, but they have not been validated against patient outcomes, and none of them has achieved widespread popularity in clinics (12). Even in a high technologic and informatics era, the famous words of William Osler remain pervasive: "*Medicine is a science of uncertainty and an art of probability*" (13). In the medical diagnostic process, clinical estimate of probability—even when Gestalt-guided or somewhat instinctive—strongly affects physicians' belief as to whether a patient has a certain disease, thus triggering further clinical actions such as rule out, treat, or ordering additional tests. Moreover, the impact of probability assessment on medical decision-making largely depends on the possible consequences of missing the disease. For example, missing a finger fracture or a myocardial infarction is not the same in physician's perspective. Obviously and understandably, probability evaluation cannot be the same. Throughout the clinical decision making process, probability must be continuously re-evaluated whenever new information becomes available. This represents a sort of mix of ongoing Bayes and Gestalt processes, whose degree of complexity is hardly appraisable.

Physicians are human exactly as their patients, and are hence vulnerable to multiple confounding factors such as unique genetic profiles and specific personal history, which would make them variably tolerable to risks and uncertainty. The tiredness of a physician is another source

of vulnerability, as recently demonstrated by a study showing that physicians, at the end of their shifts, are more prone to prescribe potentially inappropriate medicines (14).

According to the Institute of Medicine (IOM), diagnosis should be seen as a collaborative effort. The stereotype of a single physician contemplating a patient case and discerning a diagnosis is not always true, since the diagnostic process often involves intra- and inter-professional teamwork. This implies that diagnostic errors often occur as a combination of multiple errors throughout the healthcare industry, as admirably depicted by James Reason in his distinguished "Swiss cheese model" (15).

The IOM and the national Academies of Sciences have recently defined the diagnostic error as "*failure to (i) establish an accurate and timely explanation of patient's health problem(s) or (ii) communicate that explanation to the patient*" (16). Even communicating with patients has thus been included in the definition of diagnostic error, so reinforcing the efforts to improve communication, reducing legal litigation and, consequently, lowering the risk of practicing defensive medicine (17,18).

Conclusions

The achievement of substantial improvements in diagnostic accuracy needs a multifaceted approach, entailing recovered enthusiasm on traditional clinical skills teaching, systematic exploration of new educational approaches to diagnostic reasoning, a facilitating process for more effective teamwork among health care professionals, patients and their families and, finally, a substantial focus on investments in basic science of clinical diagnosis. Whenever possible, easily available evidence-based knowledge storage should be developed for helping clinicians in diagnostic decision making. Moreover, systematic enhancement of feedback is necessary for improving physician's calibration, thus encouraging the discussion of diagnostic errors, autopsies and "morbidity and mortality" conferences (19). All these changes will need a paradigm shift, organizational rebuilding and political actions.

One additional, albeit essential, improvement should also involve the definition of acceptable error rates in healthcare system, with focus at least on some major and/or most common pathologies.

Acknowledgments

Footnote

References

1. Lippi G, Plebani M, Graber ML. Building a bridge to safe diagnosis in health care. The role of the clinical laboratory. Clin Chem Lab Med 2016;54:1-3.
2. Plebani M, Lippi G. Closing the brain-to-brain loop in laboratory testing. Clin Chem Lab Med 2011;49:1131-3.
3. Plebani M, Lippi G. Improving diagnosis and reducing diagnostic errors: the next frontier of laboratory medicine. Clin Chem Lab Med 2016;54:1117-8.
4. National Institute of Standards and Technology (NIST). The Impact of Calibration Error in Medical Decision Making. Planning Report 04-1. NIST, Gaithersburg, MD, USA; 2004.
5. Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. Clin Chem 2007;53:1338-42.
6. Rolfe A, Burton C. Reassurance after diagnostic testing with a low pretest probability of serious disease: systematic review and meta-analysis. JAMA Intern Med 2013;173:407-16.
7. Plebani M, Lippi G. Improving the post-analytical phase. Clin Chem Lab Med 2010;48:435-6.
8. Cifu AS. Diagnostic Errors and Diagnostic Calibration. JAMA 2017. [Epub ahead of print].

9. Cervellin G, Borghi L, Lippi G. Do clinicians decide relying primarily on Bayesians principles or on Gestalt perception? Some pearls and pitfalls of Gestalt perception in medicine. Intern Emerg Med 2014;9:513-9.
10. Lippi G, Sanchis-Gomar F, Cervellin G. Chest pain, dyspnea and other symptoms in patients with type 1 and 2 myocardial infarction. A literature review. Int J Cardiol 2016;215:20-2.
11. Lippi G, Bassi A, Bovo C. The future of laboratory medicine in the era of precision medicine. J Lab Prec Med 2016;1:7.
12. Garg AX, Adhikari NK, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005;293:1223-38.
13. Silverman M, Murrary TJ, Bryan CS. The Quotable Osler. Philadelphia, PA: American College of Physicians, 2008.
14. Linder JA, Doctor JN, FriedbergMW, et al. Time of day and the decision to prescribe antibiotics. JAMA Intern Med 2014;174:2029-31.
15. Reason J. Human error. New York: Cambridge University Press, 1990.
16. National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health Care. Washington, DC: The National Academies Press, 2015. Available online: http://nap.edu/21794
17. Montagnana M, Lippi G. The risks of defensive (emergency) medicine. The laboratory perspective. Emerg Care J 2016;12:5581.
18. Cervellin G, Cavazza M. Defensive medicine in the emergency department. The clinicians perspective. Emerg Care J 2016;12:5615.
19. Shiff GD. Minimizing Diagnostic Error: The Importance of Follow-up and Feedback. Amer J Med 2008;121:S38-42.