Clin Chem Lab Med 2012;50(1):67–71 © 2011 by Walter de Gruyter • Berlin • Boston. DOI 10.1515/CCLM.2011.740

A risk-analysis approach to the evaluation of analytical quality

Ferruccio Ceriotti[1,*], Piero Cappelletti[2], Marco Caputo[3], Francesca Di Serio[4], Gianni Messeri[5], Cosimo Ottomano[6], Mario Plebani[7] and Giuliano Soffiati[8]

[1] Diagnostica e Ricerca San Raffaele, San Raffaele Scientific Institute, Milan, Italy
[2] Laboratorio di Patologia Clinica, Azienda Ospedaliera, Pordenone, Italy
[3] Department of Diagnostic Services, Orlandi General Hospital, Bussolengo, Italy
[4] Department of Clinical Pathology I, University Hospital of Bari, Italy
[5] Dipartimento di Diagnostica di Laboratorio, Laboratorio Generale, University Hospital Careggi, Florence, Italy
[6] Dipartimento di Medicina di Laboratorio, Ospedali Riuniti di Bergamo, Laboratorio Analisi, Bergamo, Italy
[7] Dipartimento Medicina di Laboratorio, University Hospital of Padua, Italy
[8] Laboratorio di Chimica Clinica ed Ematologia, Ospedale San Bortolo, Vicenza, Italy

Abstract

Background: Setting specifications for analytical quality is always difficult. The risk-management approach might be a way to do so. In this approach, the definition of the required analytical quality is based on the evaluation of patient risk. Risk derives from the probability of error and from the damage that such an error might cause.
Methods: Eight Italian laboratories took part in this experiment. Measurements of glucose and total calcium were taken as examples. Analytical quality was evaluated using a specific ring trial with a frozen serum pool and by means of internal quality-control data. The total allowable error was defined according to biological variation specifications. The probability of error was extracted from the imprecision and comparative bias data of each laboratory. The damage caused by a wrong result was evaluated using the absolute probability judgment approach.
Results: According to the iso-risk plots (standardized hyperboles on a graph where the x-axis represents damage and the y-axis represents probability) for glucose, all the laboratories were working with an analytical quality that guaranteed low risk for patients. On the contrary, for total calcium none of the laboratories exhibited sufficient quality to guarantee low risk for patients, the presence of bias being the most relevant problem.
Conclusions: The results seem to demonstrate the applicability of the risk approach to the analytical phase, indicating a new possible way to define analytical quality targets.

Keywords: analytical quality; quality specifications; risk analysis.

Introduction

Laboratory testing aims to provide clinicians with reliable results to correctly understand the patient's status. It can be used for diagnostic or monitoring purposes. If a clinical laboratory produces mistaken results patients might risk damage due to incorrect treatment. The reduced risk of adverse events due to laboratory results should be paramount to evaluate test analytical quality level. Recently, Krouwer and Cembrowski (1) underlined the usefulness of an approach based on risk analysis to define analytical quality. We present an approach (the "iso-risk plot"), which is less comprehensive and less complex than error grids, to verify the adequacy of analytical quality and indicate any necessary improvements.

Risk derives from two components: (a) probability of occurrence and (b) amount of damage related to the occurrence of a specific event. The concept of risk analysis, previously applied to the pre-analytical phase (2), is herein used to evaluate the adequacy of the analytical performance of eight Italian clinical laboratories. To assess this approach, we decided to focus on two of the most commonly requested tests, glucose and total calcium, as a paradigm for all other laboratory tests.

When dealing with the concept of "analytical error", we should first define what an analytical error is. We decided to consider as an "analytical error" any result obtained which presented a difference from the "true" value greater than the total allowable error (TE_a) "desirable" level (3), defined according to the biological variation approach. The probability of an analytical error was estimated using internal quality control and with an *ad hoc* proficiency testing survey. The damage caused by an erroneous result was evaluated by applying certain principles of absolute probability judgment (APJ) (4). In extreme synthesis, the method was based on the observations of a group of experts (the eight authors), who estimated the severity of damage resulting from the occurrence of an error. The experts were asked to estimate the consequences for a patient, using four levels of damage from none to severe, once an error had happened. A mathematical procedure was then used to obtain the probability of each level of damage and a

*Corresponding author: Ferruccio Ceriotti, Diagnostica e Ricerca San Raffaele, Via Olgettina 60, 20132 Milano, Italy
Phone: +39 02 26432282, Fax: +39 02 25432640,
E-mail: ceriotti.ferruccio@hsr.it
Received August 30, 2011; accepted September 17, 2011; previously published online October 6, 2011

weighted sum of the probabilities gave a value for the damage (D) caused by each error.

Materials and methods

The experiment involved a group of eight Italian clinical laboratories in Bari, Bergamo, Bussolengo, Florence, Milan, Padua, Pordenone and Vicenza.

Process analysis

Extensive information on analytical instruments, reagents, calibration and quality control procedures was collected from each laboratory (Supplementary data that accompanies the online version of this article at http://dx.doi.org/1015/cclm.2011.740).

Analytical quality data

A pool of sera with a glucose and calcium concentration close to the selected values was prepared at one center. It was aliquoted and frozen at –80°C and then sent to the other seven laboratories in dry ice. This pool was analyzed in triplicate on three different days. The imprecision and the bias from the overall mean concentrations were calculated for each laboratory.

Evaluation of possible damage for patients caused by analytical error

The laboratory directors (the authors) were asked to estimate the consequences for a patient, using a four-level damage scale, once an error had happened. We tried to select concentrations of the two analytes that were close to decisional limits. In this way, an analytical error could move the patient from one group to another, inducing (or not inducing) various therapeutic interventions. A glucose concentration of 6.38 mmol/L, within the impaired fasting glucose (IFG) gray zone, and a total calcium concentration of 2.00 mmol/L (close to the lower reference limit) were chosen. Two levels of error were set: bias >TE_a, but <2-times the TE_a and bias >2-times TE_a where TE_a was set at ±6.9% for glucose and ±2.4% for calcium (5). The four levels of damage were defined as follows: 0=no damage, 1=minimal damage (e.g., repetition of the test), 2=medium damage (delay in diagnosis and/or treatment), and 3=relevant damage (wrong diagnosis or wrong treatment).

To this end, the participants filled in a matrix that numerically estimated the probability of certain damage for both levels of error and both directions of bias (over-estimation or under-estimation). The opinions were combined and a geometrical mean was calculated to obtain an estimate of the level of damage in the four possible conditions of error (over-estimation or under-estimation >TE_a or twice the TE_a).

Probability of analytical errors

From bias and imprecision data, a sigma value was calculated for each laboratory. Based on this sigma value, the probability of obtaining a glucose or total calcium result outside TE_a and twice the TE_a was calculated. Moreover, each laboratory obtained from its statistics the percentage of glucose results that fell in the interval 6.38 mmol/L±6.9% (5.94–6.83 mmol/L) and the percentage of total calcium data within the interval 2.00 mmol/L±2.4% (1.95–2.05 mmol/L). Data were relatively similar for all eight laboratories, ranging between 11.4% and 18.0% for glucose and between 7.1% and 16.0% for total calcium. The probability of error was estimated using these data.

Risk of patient damage

This was obtained by multiplying the estimate of damage by the probability of occurrence. The probability of occurrence and damage were plotted on "iso-risk" plots, which are standardized hyperboles on a graph where the x-axis represents damage and the y-axis represents probability.

Results and discussion

From the process analysis, almost all possible common sources of error were excluded: (a) reagents were ready for use and stored properly (the correctness of temperature was controlled); (b) internal quality control programs were active, although with different approaches; and (c) errors due to inadequate samples were minimized (clot detectors in all laboratories but one, serum indices) (see supplementary data).

Analytical performance of the eight laboratories is shown in Table 1.

Evaluation of damage produced by analytical error

Given a true glucose value of 6.38 mmol/L, introducing an error >–6.9% translates into a result <5.94 mmol/L (error mode 1) while an error >–13.8% would produce a result <5.50 mmol/L (error mode 2). On the contrary, errors of +6.9% and +13.8% produce results >6.83 mmol/L and >7.28 mmol/L (error modes 3 and 4, respectively). None of the participants evaluated any of these error modes as being able to significantly damage the patient, mainly due to the consideration that international guidelines on diabetes management (6) require a second measurement to confirm diagnosis. Error mode 2 was considered the most dangerous, moving the patient from the IFG area into the non-diabetic range (not requiring confirmation); error modes 1 and 3 were considered of modest relevance, because they do not cause misclassification, while error mode 4 classifies the patient as diabetic (Figure 1A).

Given a true total calcium concentration of 2.00 mmol/L, introducing an error >–2.4% would produce a result <1.95 mmol/L (error mode 1) while an error >–4.8% would produce a result <1.91 mmol/L (error mode 2). On the contrary, errors of +2.4% and +4.8% would produce results >2.05 mmol/L and >2.10 mmol/L (error modes 3 and 4, respectively).

Also in this case, the degree of damage caused by the errors was judged to be modest, because although outside the reference intervals (error modes 1 and 2) the results were far from the critical value for hypocalcaemia (1.75 mmol/L) (7) (Figure 1B).

Probability of analytical error

Glucose Table 1 shows sigma values above 2 for all the laboratories for each error mode but one, thus indicating low probability (<5%) of a result exceeding the allowable

Table 1 Analytical performance of the eight laboratories taking part in the experiment.

	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8
Glucose								
Mean, mmol/L	5.90	6.02	6.09	6.05	6.05	5.81	6.16	6.04
Overall CV, %	1.62	1.03	2.08	1.62	1.55	1.03	1.79	1.13
Bias % from overall mean: 6.01 mmol/L	–1.9	0.1	1.2	0.6	0.6	–3.4	2.3	0.4
Median of monthly CV, %	1.8	1.4	1.1	1.6	1.8	2.0	1.1	2.2
SIGMA values[a]								
Error mode 1 TE_a –6.9%	2.8	5.1	7.5	4.7	4.3	1.7	4.6	3.3
Error mode 2 TE_a –13.8%	6.6	10.1	13.9	8.9	8.2	5.2	8.1	6.5
Error mode 3 TE_a +6.9%	4.9	5.0	5.3	3.9	3.6	5.1	2.3	3.0
Error mode 4 TE_a +13.8%	8.7	10.0	11.7	8.2	7.5	8.6	5.8	6.1
Calcium								
Mean, mmol/L	2.12	2.18	2.23	2.06	2.10	2.21	2.27	2.12
Overall CV, %	0.92	0.76	2.37	1.92	1.65	1.53	1.74	0.86
Bias % from overall mean: 2.16 mmol/L	–1.7	1.0	**3.3**	**–4.6**	**–3.0**	2.3	**4.9**	–2.1
Median of monthly CV, %	1.3	1.5	1.9	1.6	2.0	1.2	2.5	2.3
SIGMA values[a]								
Error mode 1 TE_a –2.4%	0.5	2.3	3.1	N.C.	N.C.	3.9	2.9	0.1
Error mode 2 TE_a –4.8%	2.4	3.9	4.4	0.1	0.9	5.8	3.9	1.2
Error mode 3 TE_a +2.4%	3.2	0.9	N.C.	2.7	2.7	0.1	N.C.	2.0
Error mode 4 TE_a +4.8%	5.0	2.6	0.8	3.6	3.9	2.1	N.C	3.0

The mean and CV% of each laboratory were obtained by measuring frozen aliquots of a pool of sera in triplicate on three different days. As a comparison median values of the monthly CV of the internal quality control program are shown. [a]Calculated taking into consideration the sign of the bias (error modes 1 and 2 under-estimation, error modes 3 and 4 over-estimation). N.C., calculation impossible, bias (in bold) greater than the TE_a.

error. Moreover, if this amount of analytical error occurs in patients with glucose values higher or lower than that selected (6.38 mmol/L), damage for the patient would probably be negligible, as clinical judgment would remain unchanged (except in very uncommon situations such as insulinomas). We therefore decided to correct this probability by multiplying it by the fraction of patients with glucose values in the range 5.94–6.83 mmol/L that would be more affected by these analytical errors (Table 2). For each error mode, we plotted the risk (R) for each laboratory on iso-risk graphs (2) (Figure 2A).

As shown in Figure 2A, all the laboratories appear to operate within the "control" area. The only situations closer to the planning area are the two laboratories with a more relevant analytical bias (labs 6 and 7).

Total calcium On the contrary, when using ±2.4% as TE_a, none of the laboratories showed acceptable performance and four out of eight had a bias from the overall mean that exceeded the TE_a. If considering only the monthly CV, no laboratories had a sigma value >2. This fact clearly explains

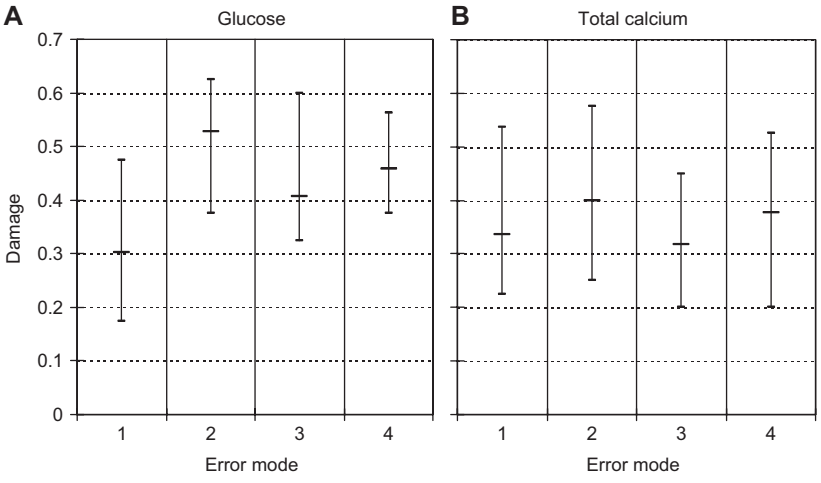

Figure 1 Evaluation of patient damage related to the various error modes.
Minimum, maximum and geometric means are indicated. For explanation of the error modes see text.

Table 2 Probability of analytical errors for the different error modes.

	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8
Glucose								
Percentage of results in the range 5.94–6.83 mmol/L	15.3	11.4	15.8	14.5	13.7	14.0	16.1	18.0
Probability of error (ppm)								
Error mode 1	837	<1	<1	<1	2	11,428	1	163
Error mode 2	<1	<1	<1	<1	<1	<1	<1	<1
Error mode 3	<1	<1	<1	13	44	<1	3,349	564
Error mode 4	<1	<1	<1	<1	<1	<1	<1	<1
Calcium								
Percentage of results in the range 1.95–2.05 mmol/L	16.0	7.1	11.7	11.3	11.6	10.0	8.2	8.1
Probability of error (ppm)								
Error mode 1	94,441	1,602	255	113,300	116,000	12	287	72,594
Error mode 2	2,736	7	2	106,353	42,702	<1	9	19,475
Error mode 3	258	24,737	117,000	804	804	93,467	82,000	4,083
Error mode 4	<1	766	49,138	34	11	4,044	82,000	219

why an analytical error (as defined in this paper) has a very high probability of occurrence (Table 2) and all patients with results within the focused range (1.95–2.05 mmol/L) are actually in error.

When drawing the iso-risk plot, most of the laboratories appear to be in the "urgency" zone, where a corrective intervention is needed.

Conclusions

The obtained results (Figure 2A) indicate that the measurement of glucose is "under control", at least in the group of laboratories involved in this study. The real problem is the possible presence of a significant bias, and how this bias might be identified and eventually corrected.

The case of total calcium was completely different (Figure 2B), as all the laboratories appear to be in the "urgency" zone. Furthermore, if we consider only the larger TE_a (±4.8%), four laboratories still remain in this zone.

It is clear that the analytical performances for this analyte are not in line with the quality specifications based on biological variation; this is not surprising, and is in accordance with the results recently presented by Carobene et al. (8). The question is whether the quality specifications are wrong, or if the analytical quality is insufficient and, in turn, if this could harm patients. Even if we consider the lowest estimation of damage (as shown in Figure 1B), the laboratories with the larger negative bias still remain in the urgency zone for the second error mode (under-estimation above 4.8%), clearly indicating the need for improved precision and a significant reduction of the analytical bias.

The proposed approach is a new way of dealing with analytical error and setting analytical quality specifications starting from a patient-centered scenario and from the risk of harming the patient, as suggested by Plebani and Lippi (9). It aims to combine the two components of the second level of the Stockholm hierarchy (biological variation and clinicians' opinions) (10), but asking the clinicians to consider the level of damage possibly caused by a certain error, rather than what amount of error will probably cause damage.

Some weaknesses in the study design should be clearly underlined, the first being identification of true error rates.

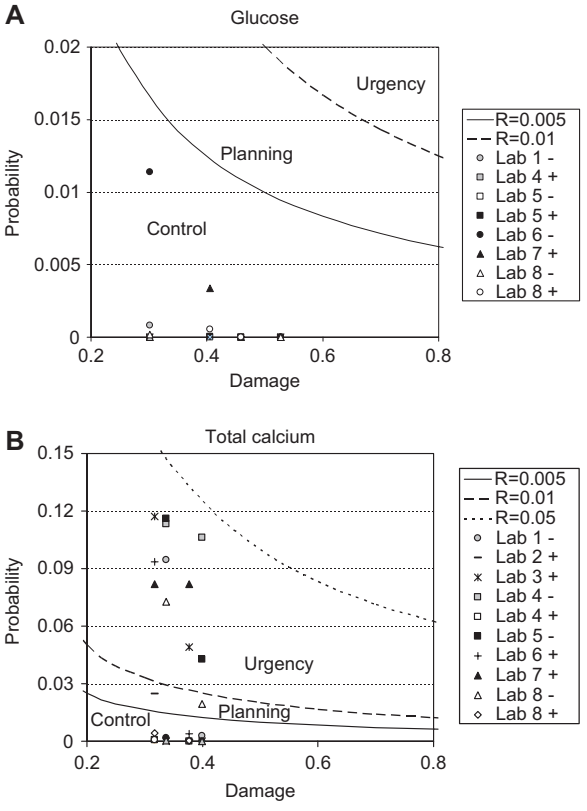

Figure 2 Iso-risk curve. Only the laboratories with a probability of error occurrence >0.0001% (A) or 0.001% (B) are indicated in the legend.
For the level of damage the average estimation shown in Figure 1 was used. The symbol (+) or (–) after the laboratory number indicates error mode over-estimation (+) or under-estimation (–).

Plebani and Carraro (11, 12), following clinical feedback from the wards, re-analyzed suspect cases. However, that approach cannot be applied systematically. Examining the analytical process of each laboratory, it is clear that the majority of causes of possible mistakes have been eliminated through complete automation, including centrifugation, ready-for-use bar coded reagents, clot detection, serum indices, and a careful internal quality control program that can detect errors or drifts in the calibration process. However, an accident may still occur, or interferences from drugs may affect the reliability of results, yet with a very low frequency that we were unable to estimate. For this reason, it was decided to extract the probability of an erroneous result from the quality control data. The second limitation is in relating the laboratory results to patient outcomes. There are several reasons for this: the main ones are the setting in which the result is used (diagnosis or monitoring) can influence the effect of an error; the type of patient (outpatient or patient in a critical care unit); the clinician's behavior (will he/she follow the guidelines?). In fact, the evaluations of the eight laboratory directors involved in this study were significantly divergent (Figures 1A and B).

In spite of these weaknesses, which could be overcome with a consensual approach to estimate patient damage, we believe this to be a good way to set targets for analytical quality.

Acknowledgments

This work was possible thanks to the valuable support provided by Siemens Healthcare Diagnostics, Italy. The authors wish to thank Michael John of the Vita-Salute San Raffaele University for the English language editing of the manuscript.

Conflict of interest statement

Authors' conflict of interest disclosure: The authors stated that there are no conflicts of interest regarding the publication of this article. Research support 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.

Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.

References

1. Krouwer JS, Cembrowski GS. Towards more complete specifications for acceptable analytical performance – a plea for error grid analysis. Clin Chem Lab Med 2011;49:1127–30.
2. Signori C, Ceriotti F, Sanna A, Plebani M, Messeri G, Ottomano C, et al. Process and risk analysis to reduce errors in clinical laboratories. Clin Chem Lab Med 2007;45:742–8.
3. Fraser CG, Hyltoft Petersen P, Libeer JC, Ricós C. Proposal for setting generally applicable quality goals solely based on biology. Ann Clin Biochem 1997;34:8–12.
4. Kirwan B. A guide to practical human reliability assessment. London: Taylor & Francis, 1994.
5. Ricós C, Alvarez V, Cava F, García-Lario JV, Hernández A, Jiménez CV, et al. Current databases on biological variation: pros, cons and progress. Scand J Clin Lab Invest 1999;59:491–500.
6. Sacks DB, Bruns DE, Goldstein DE, Maclaren NK, McDonald JM, Parrott M. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Clin Chem 2002;48:436–72.
7. 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.
8. Carobene A, Franzini C, Ceriotti F. Comparison of the results from two different external quality assessment schemes supports the utility of robust quality specifications. Clin Chem Lab Med 2011;49:1143–9.
9. Plebani M, Lippi G. Closing the brain-to-brain loop in laboratory testing. Clin Chem Lab Med 2011;49:1131–3.
10. Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Consensus agreement conference on strategies to set global quality specifications in laboratory medicine. Stockholm. April 24–26, 1999. Scand J Clin Lab Invest 1999;59:585.
11. Plebani M, Carraro P. Mistakes in a stat laboratory: types and frequency. Clin Chem 1997;43:1348–51.
12. Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. Clin Chem 2007;53:1338–42.