Impact of enzymatic methods for glucose measurement on the performance of clinical laboratories in proficiency testing

Diogo José da Silva Jerônimo¹ ² and Elisabeth Costa Monteiro¹

¹Postgraduate Programme in Metrology, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, 22451-900, Brazil
²Control Lab - Controle de Qualidade para Laboratórios, Rio de Janeiro, Brazil

E-mail: beth@puc-rio.br

Abstract. The present study compares performance scores obtained by clinical laboratories participating in Proficiency Testing schemes for measuring three concentration items of glucose in lyophilized serum, according to the analytical method employed (Oxidase-Peroxidase or Hexokinase-Dehydrogenase). Two distinct scoring criteria guided the evaluation: one based on the consensus value from the participants’ results; and another one that uses parameters exclusively based on the assigned value defined by the reference material. The results indicate that the measuring method employed in laboratory practice significantly affects the uniformity of PT participants’ outcomes. The larger dispersion of data provided by participants using the Oxidase-Peroxidase enzymatic method caused a higher index of evaluation's inadequacies for this group when the laboratory performance score applies the consensus approach.

1. Introduction

Efforts toward reliable clinical laboratory results, aiming at guaranteeing the appropriate diagnoses and treatments, are continuously increasing [1-4]. An essential tool for quality assessment of measurement capability has been widely employed utilizing participation in proper proficiency testing (PT) schemes [5,6]. Medical laboratories apply a wide variety of techniques to quantify a large number of parameters, so the results may not be comparable across labs [7].

The same analyte can be measured by different analytical methods, according to the routine employed by each medical laboratory. The non-uniformity of the analytical procedure used by participants may affect their performance results from PT interlaboratory comparisons. If the performance of the different methods is not uniform, it would probably impact on the proficiency test evaluation, mainly when the consensus value derived from participants’ results is the basis for the assigned value, which is the most commonly employed procedure.

PT performance of clinical laboratories has been demonstrated to be affected by the scoring approach employed for their measurement capability evaluation [1]. Discrepancies in the PT participant score resulted from bias introduced by the use of statistics incorporating elements associated with the consensus value obtained from the outcome of participants when compared to strategies entirely based on the reference material value for determining assign value and acceptable ranges [1]. Characteristics inherent to the different analytical methods employed for the estimative of the measurand may play a role in the discrepancies produced by participants’ consensus in the target value and the standard deviation for the proficiency assessment.
Glucose measurement is among the most customary analyzes of the clinical laboratory’s routine. Measurement methods employed are based on enzymatic reactions, especially Hexokinase-Dehydrogenase (H-D) and Oxidase-Peroxidase (O-P). Both the approaches encompass spectrophotometric assays. In the H-D approach, the hexokinase facilitates the phosphorylation of glucose, being followed by an oxidation reaction by glucose-6-phosphate dehydrogenase, which produces reduced nicotinamide adenine dinucleotide (NADPH) proportionally to the amount of glucose in the sample. NADPH is, then, measured spectrophotometrically. Regarding the O-P method, glucose is oxidized to gluconic acid and hydrogen peroxide, followed by a reaction catalyzed by peroxidase that results in a product quantifiable by spectrophotometric detection [8].

The present study compares the proficiency testing performance of biomedical laboratories for detecting and quantifying glucose in human serum, which used the two different enzymatic methods: Oxidase-Peroxidase and Hexokinase-Dehydrogenase. The analysis compares all PT participants’ results according to the \( Z \) score, based on the consensus value obtained from the outcome of participants; and to an approach named \( Z'_{FL \, score\,(ref)} \) in [1], which comprises parameters exclusively based on the assigned value defined by the reference material.

2. Materials and Methods

The study evaluates the results reported by 1737 clinical laboratories taking part in a proficiency testing round for glucose determination in samples of lyophilized human serum with three different glucose concentrations (91.08 mg/dL, 301.96 mg/dL, 143.33 mg/dL) [1]. The data were provided by an accredited PT provider. All PT participants reported the employment, in their routine glucose measurements, combinations of analytical systems based on two possibilities of enzymatic methods: Oxidase, with Peroxidase reaction (O-P); or Hexokinase, with Dehydrogenase reaction (H-D).

The comparative analysis of performance evaluation was based on two statistical designs. One of them, the \( Z \) score described in [9], is frequently employed by proficiency testing procedures (equation 1).

\[
Z\,\text{score} = \frac{x - X_{\text{cons}}}{\sigma_{\text{cons}}}
\]  

(1)

where \( x \) is the result reported by a participant; \( X_{\text{cons}} \), the assigned value obtained from the interlaboratory consensus amongst participants; and \( \sigma_{\text{cons}} \), the standard deviation for proficiency assessment derived from the results reported by the participants in the same round.

The second approach for evaluating the clinical laboratories performance, proposed in [1], is based on parameters entirely independent of the participants’ measurement capability (equation 2).

\[
Z'_{FL\,\text{score\,(ref)}} = \frac{x - X_{\text{ref}}}{\left(\sigma^2_{\text{ref,FL}} + u^2_{X_{\text{ref}}}\right)^{1/2}}
\]  

(2)

where \( X_{\text{ref}} \) is the assigned value derived from the reference value of the provided PT item; \( \sigma_{\text{ref,FL}} \), the standard deviation for proficiency, determined using a fixed limit (FL) of 13 % of the assigned value defined by the reference value of the PT item (\( \sigma_{\text{ref,FL}} = X_{\text{ref}} \, \times (FL/2) \)) [10]; and \( u_{X_{\text{ref}}} \) is the standard uncertainty for reference values (\( X_{\text{ref}} \)) associated with each round item [9, 11], determined by:

\[
u_{X_{\text{ref}}} = \sqrt{\frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left( \frac{x_i - n_h \bar{X}_{\text{ref}}}{n_h} \right)^2}
\]  

(3)

where \( x_i \) is the \( i_{th} \) value obtained from the sample being analysed for the demonstration of sufficient homogeneity; and \( n_h \) is the sample size employed for the homogeneity assessment.
The standard deviation for proficiency assessment using the $Z_{FL \text{ score}(\text{Ref})}$ approach was established on a fit for purpose value, which is recommended by the International Union of Pure and Applied Chemistry (IUPAC) and described in the Harmonized Protocol for the PT in Analytical Chemistry [1, 9-11].

Outliers may be responsible for differences between means and variances of results provided by the two groups of participants [8], subdivided according to the analytical methods they used for measuring glucose in human serum (O-P or H-D). The Grubbs’ test evaluated the statistical outlier of data reported for each item of the PT round by the two groups of participating laboratories. After eliminating the outliers, the Kruskal-Wallis’ test (K-W) analysis investigated the difference between means of the two groups of laboratories. Under the same conditions, using the Levene’s test, the variances of results for each group of analytical method were checked for equality. Differences between assigned values designated by consensus or by the reference, associated with each PT item, were investigated by applying the one-sample $t$-test.

### 3. Results and Discussion

The study evaluates PT performance of clinical laboratories for quantifying glucose in lyophilized serum using two frequently employed enzymatic methods, Oxidase-Peroxidase (O-P) and Hexokinase-Dehydrogenase (H-D). The criteria for PT assessment derived from the participants’ consensus ($Z_{\text{score}}$), or were entirely independent ($Z_{FL \text{ score}(\text{Ref})}$) and based on the assigned value defined by the reference material. The reference values ($X_{\text{ref}}$) for each round item, and their standard uncertainties ($u(X_{\text{ref}})$), were obtained from the data provided by the assessment of homogeneity of the reference material, corresponding to items of round PT.

The $t$-test indicated differences between $X_{\text{const}}$ and $X_{\text{ref}}$ for PT items CHM001 and CHM003, at a significance level of 5%. Table 1 presents the results of tests to investigate the differences between outcomes from the two distinct analytical methods (O-P and H-D) used by participants for quantification of glucose.

| Item    | test  | Statistics | $p$-value |
|---------|-------|------------|-----------|
| CHM001  | K-W   | 19.07900   | 0.00001   |
|         | Levene| 116.86000  | 0.00000   |
| CHM002  | K-W   | 48.76600   | 0.00000   |
|         | Levene| 101.89000  | 0.00000   |
| CHM003  | K-W   | 0.10081    | 0.75090   |
|         | Levene| 97.08300   | 0.00000   |

The Levene’s test was significant at the 5% level for all three items of the round (table 1), indicating that there was a difference between the variances of the two sets of analytical methods employed by the participants. These differences, observed for the three PT items, impact on the standard deviation for proficiency when it derives from the participants’ consensus.

In turn, the K-W test indicated that there was a significant difference between the means of the analytical methods for items CHM001 and CHM002 (table 1). For these two items, one expects that the differences between the means and variances of the analytical systems O-P and H-D will result in inconsistencies in the performance evaluation of the laboratories when using consensus-based approaches. For item CHM003 (table 1), discrepancies are expected to arise only from the differences between the variances of criteria.

Figure 1 presents the laboratory outcomes for CHM001 item, according to the analytical method used by participants of the proficiency test round of glucose dosing. From the total number of 1737 results reported, after applying the Grubbs’ test for identifying and removing outliers, this number was
reduced to 1392 results (19.86% sample reduction). The great majority of data originated (1038 laboratories) employing Oxidase-Peroxidase (O-P), which consists of the most commonly used method. Participants employing Hexokinase-Dehydrogenase (H-D) method provided 354 outcomes. In figure 1, the dashed blue line and the gray region represent the acceptable ranges for outcomes of glucose concentration according to Z score and Z'LF-score(Ref), respectively. The solid red and blue lines indicate, respectively, the X_ref and X_cons assigned values (figure 1).

The Oxidase-Peroxidase method shows a larger dispersion of the participants’ results when compared to the outcomes reported by laboratories using Hexokinase-Dehydrogenase method. As a result, the great majority of participants using H-D enzymatic method presented a satisfactory classification when laboratory performance considered the independent criterion Z'LF-score(Ref), proposed in [1]. In turn, a significant number of results derived from O-P measurements were outside this acceptable range, which is exclusively based on the reference value (X_ref), limited by the gray region. Even though there are differences between the O-P and H-D measurements' means (table 1 and figure 1), both shift away from the X_ref. This difference is responsible for the discrepancy between X_cons and X_ref assigned values. Despite the higher adequacy of the results' set obtained using H-D, with reduced dispersion of outcomes, their average value is even more distant from X_ref than the bias associated with the O-P results (figure 1).

The discrepancy between the acceptable range for the Z score (dashed blue lines of figure 1), and for the Z'LF-score(Ref) reference approach (gray region) indicates a potential risk of an incorrect decision of clinical laboratories’ classification when based on Z score. For the proficiency testing with the CHM001 item, 6% of participants received improper classification by Z score, being 5.7% from the laboratories employing the O-P method, and 0.3% from the participants using the H-D approach. These results demonstrate the significant impact of the wider dispersion of O-P results, risking their users of an inappropriate classification when laboratory performance evaluation employs scoring approaches strongly dependent on the participants’ consensus.

Figure 1. Distribution of reported results for glucose concentration of item CHM001, according to the analytical method employed by laboratories (O-P and H-D). The region in gray shows the acceptable range for Z'LF-score(Ref), and dashed blue lines indicate the acceptance limits for Z score.
The $t$-test of both the PT items CHM001 (table 1 and figure 1) and CHM003 indicated differences between $X_{\text{ons}}$ and $X_{\text{ref}}$, at a significance level of 5%. However, the impact on the metrological compatibility of PT results was more evident for the item with the lowest concentration (CHM001), for which both K-W and Levene tests showed significant differences between analytical systems. In turn, item CHM003, whose K-W indicates a good correlation between the enzymatic methods, the PT compatibility remained in the edge of an acceptable limit.

For the item with the highest glucose concentration (CHM002), although the $t$-test indicated an agreement between $X_{\text{ons}}$ and $X_{\text{ref}}$, both the K-W and Levene tests pointed toward significant differences between the means and variances of the analytical systems O-P and H-D. Considering the fewer performance discrepancies and the satisfactory PT compatibility observed for the CHM002 item, differences between methods appear to be less relevant than agreement between consensus and reference assigned values.

According to literature, O-P and H-D approaches present good correlation [12], as also observed in this study for the CHM003 item measurements, but differences between their consensus and the reference values generate the need of applying independent criteria for performance evaluation.

The significant difference observed between the results obtained from these two distinct analytical methods, and their impact on the PT performance evaluation, particularly for the CHM001 item (table 1 and figure 1), corroborates with [7], which shows that laboratory results are strongly dependent on the selected method to perform measurements, with risks of making their results incomparable.

4. Conclusion

The present study characterizes the differences in performance scores obtained by clinical laboratories participating in Proficiency Testing schemes for measuring three concentration items of glucose in lyophilized serum, according to the analytical method employed (Oxidase-Peroxidase or Hexokinase-Dehydrogenase). The evaluation was guided by two distinct scoring criteria, the $Z$ score, based on the consensus value from the participants’ results; and the $Z’_{FL}$ score(ref) [1], which considers parameters exclusively based on the assigned value defined by the reference material.

The results indicate that the measuring method employed in the laboratory practice significantly affects the agreement between outcomes provided by PT participants. The laboratories using the Oxidase-Peroxidase enzymatic method presented a much higher dispersion of reported data than Hexokinase-Dehydrogenase outcomes. This scattering of provided O-P results induced their high index of incorrect classification of performance (as satisfactory or unsatisfactory) when using the consensus approach for evaluation. On the other hand, due to the lower dispersion of the results provided by laboratories using the H-D method, the opposite was observed for this group of participants.

The widespread use of analytical methods with considerable variability of their results, coupled with a significant number of measurement systems with effectively unsatisfactory performance, generates an expanded distribution of the results reported in a proficiency test. This wide dispersion of results induces non-homogeneities of the performance of PT participants using these technologies, which increases their risk of unacceptable performance classification. Besides, the larger scattering of results directly affects the values of the parameters associated with the $Z$ score evaluation. As a result, a bias in the acceptance range can occur when a proficiency test analysis employs parameters based on the consensus of the participants ($Z$ score), generating a higher risk of incorrect decision regarding the performance of the laboratory.

This non-uniformity of the quality of laboratory results according to measurement methods employed emphasizes the relevance of using reliable targets to set the PT assigned value; ideally, traceable to SI [1, 3, 4, 13]. The use of proficiency testing criteria thoroughly independent of the results reported by the participants contributes to the appropriate decision of clinical laboratories’ performance [1, 13].

Given the different combinations of instrumentation and reagents, which the two enzymatic approaches may employ, future work consists of extending the present investigation to evaluate the impact of the nuances of the complete set of analytical system's elements in the PT performance evaluation.
Considering the possibility of a biased consensus due to the use of faulty methodology by participants, the score inadequacy can be prevented by applying reliable references to set the assigned value, ideally traceable to SI, utterly independent of the results reported by the PT participating laboratories.

Acknowledgment
The authors acknowledge the Control Lab management for providing the data analyzed in this study.

References
[1] Jerônimo D J S and Costa Monteiro E 2018 J. Phys.: Conf. Ser. 1065 132003.
[2] Plebani M, Sciacovelli L and Aita A 2017 Clin. Lab. Med. 37 187–205
[3] Costa Monteiro E 2007 Metrologia e Instrumentação 6 6-12
[4] Costa Monteiro E and Leon L F 2015 J. Phys.: Conf. Ser. 588 012032
[5] ISO/IEC 2010 ISO/IEC 17043:2010 Conformity Assessment — General Requirements for Proficiency Testing (International Organisation for Standardization)
[6] ISO 2012 ISO 15189:2012 Medical Laboratories – Particular Requirements for Quality and Competence (International Organisation for Standardization)
[7] Delatour V, Lalere B, Saint-Albin K, Peignaux M, Hattchouel J M, Dumont G, De Graeve J, Vaslin-Reimann S and Gillery P 2012 Clin. Chim. Acta 414 182–184
[8] Moodley N, Ngxamngxa U, Turzyniecka M J and Pillay T S 2015 J. Clin. Pathol. 0 1–7
[9] ISO 2015 ISO 13528:2015 Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparisons (International Organisation for Standardization)
[10] Ehrmeyer S S and Laessig R H 1987 Clin. Chem. 33(10) 1901–1902
[11] Thompson M, Ellison S L R and Wood R 2006 Pure Appl. Chem. 78(1) 145–196
[12] Ayyanar K, Pichandi S, Janakiraman P 2018 Int. J. Biotechnology and Biochemistry 14 51–58
[13] De Bièvre P 2016 Accredit. Qual. Assur. 21 167–169