Clinica Chimica Acta 433 (2014) 259–263

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

Comparing analytic performance criteria: Evaluation of HbA_{1c} certification criteria as an example

Curt L. Rohlfing [a,*,1], Curtis A. Parvin [c,1], David B. Sacks [d], Randie R. Little [a,b] for the NGSP Steering Committee

[a] Department of Pathology & Anatomical Sciences, University of Missouri, Columbia School of Medicine, 1 Hospital Drive, Columbia, MO 65212, United States
[b] Department of Child Health, University of Missouri, Columbia School of Medicine, 1 Hospital Drive, Columbia, MO 65212, United States
[c] Quality Systems Division, Bio-Rad Laboratories, Plano, TX 75074, United States
[d] Department of Laboratory Medicine, National Institutes of Health, Bethesda, MD 20892, United States

ARTICLE INFO

Article history:
Received 31 January 2014
Received in revised form 26 March 2014
Accepted 27 March 2014
Available online 8 April 2014

Keywords:
Analytical bias
Analytical imprecision
Hemoglobin A_{1c}
HbA_{1c}

ABSTRACT

Background: Direct comparison of analytical performance criteria that utilize different statistical approaches can be problematic. We describe a mathematical approach to compare performance criteria for hemoglobin A_{1c} (HbA_{1c}) analysis used by the NGSP standardization program and the College of American Pathologists (CAP) to enhance consistency between the schemes.
Methods: The imprecision (CV) and bias combinations required to pass each criterion at probabilities of 0.95, 0.99 and 0.999 were calculated and used to construct contour plots to compare them. The CV/bias requirements were calculated mathematically for the 2011–2012 CAP (3/3 results within ±7% of the target) and different proposed NGSP (33/40 to 40/40 results within ±7% of the target) criteria, and using computer simulations for the existing NGSP criterion (95% confidence interval of the differences between the method and NGSP within ±0.75% HbA_{1c}).
Results: Requiring 37 of 40 results to be within ±7% of the NGSP target best matched the CAP criterion at zero bias (95% chance of passing).
Conclusions: The NGSP Steering Committee recommended a certification criterion of 37 of 40 results within ±7% of the NGSP (reduced to ±6% in 2014). The described evaluation approach may be useful in other situations where comparison of different performance criteria is desired.

1. Introduction

For critical analytes, defined analytical performance goals are vital to facilitate the optimal use of the test in patient care. In the process of developing these goals direct comparisons between performance criteria that utilize different schemes and/or statistical methodologies may be useful; however, such comparisons may be difficult to perform mathematically.

Hemoglobin A_{1c} (HbA_{1c}) is a well-established and important indicator of outcome risks in patients with diabetes. The American Diabetes Association (ADA) and other clinical organizations, including the World Health Organization, the European Association for the Study of Diabetes, and the International Diabetes Federation, now recommend HbA_{1c} for diagnosing diabetes. Based on the results of large-scale clinical trials that established the relationships between HbA_{1c} and clinical outcomes [1,2], the ADA began recommending specific HbA_{1c} levels for use in the treatment of patients in 1994 [3]. However, the College of American Pathologists (CAP) proficiency survey data clearly showed large discrepancies in HbA_{1c} results among assay methods and laboratories [4], making it extremely difficult to incorporate the treatment guidelines in clinical practice. Thus the National Glycohemoglobin Standardization Program (NGSP) was implemented in 1996 with the goal of harmonizing HbA_{1c} results to those of the clinical trials that established the relationships between HbA_{1c} and outcome risks [5], specifically the Diabetes Control and Complications Trial [1] and United Kingdom Prospective Diabetes Study [2]. The NGSP assists manufacturers with calibration, certifies them as traceable to the DCCT via sample comparisons with NGSP network laboratories, and monitors the effectiveness of the program via the CAP GH-2 whole blood proficiency survey for HbA_{1c}. The program has been highly successful in its efforts to improve the quality of HbA_{1c} testing [5].

Since its inception in 1996, the NGSP has tightened several times the criteria for certification of manufacturers' methods with the goal of improving the quality of HbA_{1c} testing. Similarly, CAP replaced peer-group grading of HbA_{1c} for the GH-2 HbA_{1c} survey with accuracy-based grading in 2007, and has since tightened the acceptable performance limits from the initial ±15% to ±7% in 2011–2012 [5] and ±6% in 2013. At the time CAP adopted ±7% as the acceptability criterion, the NGSP criterion was based on Bland/Altman statistical methods [6]. The NGSP

Abbreviations: ADA, American Diabetes Association; SRL, Secondary Reference Laboratory.

* Corresponding author at: Diabetes Diagnostic Laboratory, M767A Department of Pathology & Anatomical Sciences, University of Missouri School of Medicine, 1 Hospital Dr., Columbia, MO 65212, United States. Tel.: +1 573 884 2385; fax: +1 573 884 4748.
E-mail address: RohlfingC@missouri.edu (C.L. Rohlfing).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.cca.2014.03.034
0009-8981/

required that the 95% confidence interval of the differences between a manufacturer's method (analyzed in duplicate) and the NGSP Secondary Reference Laboratory (SRL, mean of duplicates) in a 40 sample comparison must fall within ±0.75% HbA_{1c} [5]. In deciding whether to tighten the NGSP criterion, the NGSP Steering Committee felt it important to compare the existing and any proposed new criteria to the CAP criterion to ensure that the two are comparable.

Unfortunately, the 2 sets of criteria could not be compared directly. This is due to NGSP certification requiring that 40 samples be analyzed in duplicate, while for CAP surveys three samples are each measured only once. Moreover, the NGSP criterion was based on the mean and standard deviation (SD) of the differences, while the CAP criterion is based on the percentage deviation from the target value. Therefore, we developed a new statistical strategy to compare the two different schemes. Here, we describe the statistical method used to compare the CAP and NGSP criteria. These analyses enabled the development of an NGSP certification criterion that was comparable to the CAP criterion. The statistical approach presented here can be applied to other situations where comparison of different approaches to analytic acceptance criteria is required.

2. Methods

2.1. NGSP and CAP criteria

Both the existing (2010–2012) NGSP criterion (95% confidence interval of the differences between laboratory method and SRL within ±0.75% HbA_{1c}) and proposed NGSP criteria (33 to 40 out of 40 single results within ±7% of the means of duplicate SRL results) were compared to the 2011–2012 CAP limits of ±7% of the target value.

2.2. Calculations

Both analytical bias and imprecision (CV) influence the accuracy of individual HbA_{1c} measurements; thus, the probability of a laboratory passing a given criterion depends upon the laboratory's bias and CV. For the CAP and proposed NGSP criteria, the probability of passing as a function of the laboratory's bias and CV was mathematically derived and computed over a range of bias (−7% to +7% in 0.2% increments) and CV (0% to 5% in 0.1% increments) combinations. Computer simulations were employed to compute the probabilities of passing the existing NGSP criterion. Forty HbA_{1c} levels were randomly generated to match the distribution of HbA_{1c} results used for NGSP certification (8 samples between 4% and 5.5%, 12 samples between 5.5% and 7%, 12 samples between 7% and 8.5%, 8 samples between 8.5% and 10% HbA_{1c}). Duplicate measurements were randomly generated for each level for the NGSP SRL. Singleton measurements were generated for a hypothetical laboratory that reflected a specified bias and CV combination. The simulated data were evaluated by the existing NGSP criterion to determine pass or fail. This process was repeated one million times and the fraction of the million simulations that passed the criterion was used to estimate the probability of a laboratory passing given the specified bias and CV combination.

2.3. Contour plots

Contours of constant probability (0.95, 0.99 and 0.999) were derived from the computed probabilities of passing a given criterion over the grid of relative bias and CV combinations evaluated. Contour plots were then constructed to directly compare the different criteria. Details of the mathematical derivations and simulations are given in Appendix A. All analyses were performed using Matlab software (MathWorks).

2.4. Target value assignments and uncertainties

CAP target values are assigned by all of the NGSP SRLs (n = 7) where each SRL analyzes each CAP sample three times on each of two separate days. NGSP certification target values are based on the means of duplicate SRL results. Estimates of the uncertainties for the CAP value assignments and the NGSP certification target values (CVs of 0.5% and 1.5%, respectively) were obtained based on value assignments from four previous CAP surveys performed in 2009 and 2010 (2 surveys per year, each with 3 HbA_{1c} concentrations) and incorporated into the models. Although CAP requires two of three survey samples to be within ±7% in order for a laboratory to pass, our calculations were based on all three samples falling within these limits.

3. Results

3.1. Existing (2010–2012) NGSP vs. CAP 2011–2012 criteria

Fig. 1 plots solid contour lines for the existing NGSP acceptance criterion and dashed contour lines for the CAP ±7% acceptance criterion. The NGSP criterion was less stringent than CAP at all probability levels as demonstrated by each solid contour curve dominating the dashed contour curve of the same shade. This implies that there is a wider range of possible laboratory bias and CV combinations that would pass the existing NGSP criterion at a given probability level compared to the CAP criterion.

3.2. Proposed NGSP (2013) vs. CAP 2011–2012 criteria

In order to more directly align the proposed NGSP criterion with the CAP requirements, the new criterion should be based on the number of results (out of 40) that are required to fall within ±7% of the SRL results (mean of duplicates) rather than Bland/Altman. Fig. 2 plots contours for the CAP criterion (dashed lines) and the proposed NGSP criteria (solid lines) requiring 36 out of 40 (Fig. 2A), 37 out of 40 (Fig. 2B) and 38 out of 40 (Fig. 2C) results to be within ±7% of the SRL.

Fig. 2A shows results similar to those in Fig. 1. An NGSP criterion that would require 36 out of 40 results to fall within ±7% of the SRL is less stringent than the CAP criterion at all probability levels for any laboratory with a CV > ~1.5%. Fig. 2B shows that for laboratory bias < ~±3% the 37/40 criterion is comparable to the CAP criterion at a 0.95 probability of passing (black curves). At higher probabilities of passing (0.99 and

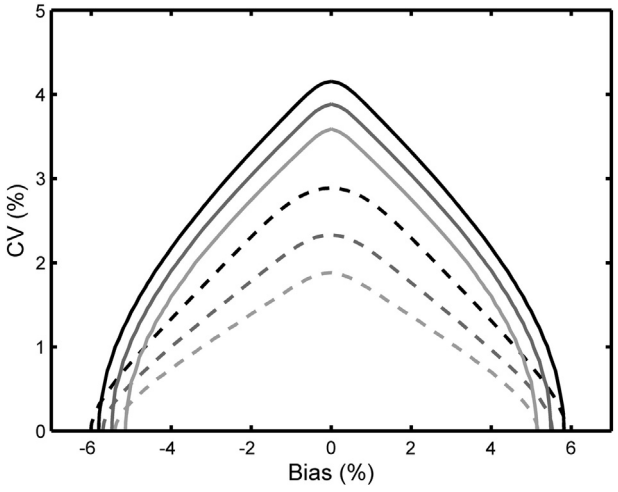

Fig. 1. The existing (2010–2012) NGSP certification criterion (95% confidence interval of the differences between laboratory method and SRL within ±0.75% HbA_{1c}) compared to the 2011–2012 CAP criterion of ±7%. The lines represent the bias and CV combinations required to pass the NGSP (—) and CAP (– –) criteria with 0.95 (top, black), 0.99 (middle, dark gray) and 0.999 (bottom, light gray) probabilities.