CLINICS ARTICLE TITLE PAGE TEMPLATE

UNDERSTANDING THE USE OF SIGMA METRICS IN HBA1C ANALYSIS

AUTHOR NAMES AND DEGREES

Erna Lenters-Westra[1,2], PhD, and Emma English, PhD[3]

AUTHOR AFFILIATIONS

[1]HbA1c researcher, Clinical Chemistry Department, Isala, Zwolle, The Netherlands

[2]European Reference Laboratory for Glycohemoglobin, location Isala, Zwolle, The Netherlands

[3]Lecturer in Health Science, University of East Anglia. Norwich, UK

AUTHOR CONTACT INFORMATION

[1]Dr. van Heesweg 2, 8024 AB, Zwolle, The Netherlands, w.b.lenters@isala.nl

[3]Faculty of Medicine and Health, School of Health Science, Norwich Research Park, emma.english@uea.ac.uk

CORRESPONDING AUTHOR

Erna Lenters-Westra

DISCLOSURE STATEMENT

The Authors have nothing to disclose

KEY WORDS

HbA1c, diabetes, sigma-metrics, analytical performance criteria

4-8 keywords to direct and optimize search results

KEY POINTS

1) Performance of HbA1c analyzers is, on the whole, very good with many attaining the international guidance target of sigma >2 at a TAE of 10%, in addition many analysers perform well in excess of these targets, as individual analyzers or in networks of analysers across a wide geographical area
2) In strict evaluation conditions, point of care test devices can perform as well as routine laboratory analysers and may in future be considered suitable for use in the diagnosis of diabetes
3) Using direct calibration to the primary reference measurement procedure we have demonstrated that there is the capacity to improve performance of analysers in routine clinical practice
4) Whilst fewer analysers currently meet tighter targets of a TAE of 10% and four sigma and TAE of 6% at two sigma, the outcomes for pass or fail are comparable for the two criteria.

5) Precision has a greater impact on the calculated sigma than bias does, in the current data set. There are many ways in which to calculate bias and imprecision and the method used to establish these values must be detailed in any evaluation or study.

SYNOPSIS

Provide a brief summary of your article (100 to 150 words; no references or figures/tables). The synopsis appears only in the table of contents and is often used by indexing services such as PubMed

This study utilizes three unique data sets to demonstrate the state of the art of HbA1c analyzers in a range of settings and compares their performance against the international guidance set by the IFCC task force for HbA1c standardization. The data is used to demonstrate the effect of tightening of those criteria and the study serves as a guide to the practical implementation of the sigma metrics approach in a range of clinical settings.

UNDERSTANDING THE USE OF SIGMA METRICS IN HBA1C ANALYSIS

Erna Lenters-Westra[1,2], PhD, and Emma English, PhD[3]

[1]HbA1c Researcher, Clinical Chemistry Department, Isala, Zwolle, The Netherlands

[2]European Reference Laboratory for Glycohemoglobin, location Isala, Zwolle, The Netherlands

[3]Lecturer in Health Science, Faculty of Medicine and Health, University of East Anglia, Norwich Research Park.

[1]Dr. van Heesweg 2, 8024 AB, Zwolle, The Netherlands, w.b.lenters@isala.nl

Corresponding author:

Erna Lenters-Westra

Disclose Statement:

The Authors have nothing to disclose

Key words:

HbA1c, diabetes, sigma-metrics, analytical performance criteria

Introduction

Global standardization of HbA1c methods, in particular through the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) primary reference method, has been the fore-runner to a marked improvement in the analytical performance of many HbA1c assays. The IFCC primary reference method has been proven to be stable in many inter-comparison studies with designated comparison methods (DCMs), (the NGSP and the Japanese Diabetes Society (JDS)/Japanese Society of Clinical Chemistry (JSCC)), method and is recognized in a 2010 consensus statement as the only valid anchor for the standardization of HbA1c (1, 2). The development of this reference method enabled the World Health Organization (WHO) and American Diabetes Association (ADA) to advocate the use of HbA1c for the diagnosis of type two diabetes (3). WHO guidance stipulates that "stringent quality assurance tests are in place and assays are standardized to criteria aligned to the international reference values". In addition to standardization, it is imperative that tests used for the diagnosis of diabetes meet strict analytical performance criteria as both bias and imprecision have the potential to affect reported values. This effect may be to the extent to cause mis-classification of a patient's result, and thus give rise to either an inappropriate positive diagnosis or a missed diagnosis.

In 2015 the IFCC Task Force on Implementation of HbA1c Standardization published an investigation of two different models to set and evaluate quality targets for HbA1c (4). The biological variation model and sigma-metrics model were investigated. The IFCC Task force advocates the use of the sigma-metric model, as the model of choice, as the within biological variation of HbA1c is very small and analytical performance criteria derived from biological variation of HbA1c are too strict even for the best performing HbA1c methods currently on the market (5). The taskforce set out guidance for the application of the sigma metrics model for HbA1c in all countries with the purpose of engaging international stakeholders in the field of diabetes to further the development of analytical quality at a global level.

In the laboratory sigma-metrics is a quality management strategy that provides a universal benchmark for process performances. Sigma-metrics places analytical characteristics (bias and imprecision) in the form of Total Allowable Error (TAE) within a framework of clinical requirements. Sigma Metrics allows the user to define the TAE they wish to achieve and how often results can acceptably be outside of this target value. The higher the sigma unit, the fewer times a system is allowed to fall short of the required target. A sigma of two implies a 5% risk to fail the TAE, i.e. it is acceptable for 5% of results to not meet the TAE value that has been pre-set. TAE for HbA1c has been set by the IFCC Task Force on Implementation of HbA1c standardization as a default of 5 mmol/mol (0.46% DCCT) at an HbA1c level of 50 mmol/mol (6.7% DCCT) which corresponds with a relative TAE of 10**%** ((5/50)*100%) in SI units (6.9% DCCT units ((0.46/6.7)*100%)) (6).

The aim of this study was to explore the ways in which sigma metrics is applied in clinical laboratory testing using the criteria set by the IFCC Task Force. Three approaches were used as exemplars of range of different uses of the sigma metrics approach. These were:

1) data from an EQA program of 134 individual laboratories in the Netherlands using a variety HbA1c methods, demonstrating how sigma is used in 'real-life' routine settings (6)

2) data from a recent evaluation of the performance of 7 HbA1c POC instruments to calculate sigma, which demonstrates how sigma metrics can be used to evaluate new or existing instruments at a laboratory level (7).

3) data from the IFCC monitoring program from 6 certified secondary reference method procedures (SRMP) to the IFCC reference method procedure (RMP) to calculate sigma , demonstrating how well methods perform at higher order levels (8).

In addition, changing the TAE of 10% to 6% in SI units was investigated to indicate the potential impact of tightening the criteria in the future.

Methods

Part One) Data from the recent (2016.2) Stichting Kwaliteitsbewaking Medische Laboratoria (SKML) External Quality Assurance Services (EQAS) in the Netherlands were used to assess the individual laboratory performance of various HbA_{1c} methods using sigma-metrics (6). In the SKML EQA program four fresh whole blood EDTA samples are distributed to individual laboratories, six times per year. The samples should to be analyzed within forty-eight hours of receipt and the results are submitted to the website of SKML. The results of the latest survey were used to calculate sigma. The target values of the distributed samples were assigned with the following six IFCC certified secondary reference measurement procedures (SRMP) based at two European reference laboratories, with values determined on two individual days in duplicate (9):

European Reference Laboratory for Glycohemoglobin, location Isala, Zwolle, The Netherlands:

- Roche Tina-quant Gen.2 HbA1c on Integra 800, immunoassay, (Roche Diagnostics);
- Premier Hb9210, boronate affinity chromatography HPLC, (Trinity Biotech);
- Tosoh G8, cation-exchange HPLC, (Tosoh Bioscience).

European Reference Laboratory for Glycohemoglobin, location Queen Beatrix Hospital, Winterswijk, The Netherlands:

- Sebia Capillarys 2 Flex Piercing, capillary electrophoresis, (Sebia);
- Premier Hb9210, boronate affinity chromatography HPLC, (Trinity Biotech);
- Menarini HA8180, cation-exchange HPLC, (Menarini Diagnostics).

The six SRMP were all calibrated with IFCC secondary reference material and showed excellent performance in the IFCC monitoring program 2015.

The mean of the four samples minus the mean of the target value set by the six SRMPs was used to calculate bias. The within-laboratory SD was calculated from the residual SD regression line through the laboratory results versus the target values and was used to calculate the coefficient of variation

(CV) (10). Sigma was calculated with the formula: σ=(TAE – B)/CV where TAE is total Allowable Error, B is bias compared to a reference method and CV is imprecision of the method. The TAE used was 5 mmol/mol (0.46% DCCT) at an HbA1c level of 50 mmol/mol (6.7% DCCT) which corresponds with a relative TAE of 10**%** ((5/50)*100%) in SI units (6.9% DCCT units ((0.46/6.7)*100%)) (4).

Part two) Data from an evaluation study of seven HbA1c point-of-care instruments done in our lab in 2014 (7) and the latest results of an evaluation of the Quo-Test after the manufacturer claimed to have adjusted the calibration are included. In these studies the CLSI EP-5 and EP-9 protocols were used to establish precision and accuracy (bias) for the POCT instruments. The CLSI EP-9 protocol was performed twice, with two different reagent lot numbers, and the bias was determined between the POC instruments and the mean of the first three SRMPs detailed above. The sigma value was calculated three different ways using the CV, of a sample with an HbA1c value of approximately 48 mmol/mol (6.5% DCCT), established in the EP-5 protocol and the CVs calculated from the duplicates in the EP-9 protocol using two different lot numbers. The bias was calculated at 48 mmol/mol using, Deming regression analysis compared to the mean of the first three SRMPs.

Part three) The design of the IFCC monitoring program is based on twenty four interconnected EDTA whole blood samples. The samples are sent annually to IFCC certified reference labs and subsequently stored at or below -70°C. One sample is then analyzed every fortnight, and the results are to be submitted to the website of HbA1c/IFCC (8). The twenty four samples are in fact twelve blinded samples in duplicate. From these duplicates the CV was calculated. Values of samples were assigned by the whole IFCC network (eighteen approved IFCC primary reference laboratories) each running the IFCC primary reference method (9). Data from the 2015 monitoring program were used to calculate the performance of six IFCC SRMPs mentioned above in part one.

In addition to the above, data was reanalyzed to investigate the impact of changing targets from a TAE of 10% to a TAE 6% to elucidate the likely impact of ‘tightening’ the performance criteria. The

pass and fail rate calculated with a TAE of 10% and four sigma was compared with the pass and fail rate with a TAE of 6% and a two sigma.

Statistics

Calculations were performed using Microsoft® Excel 2010 (Microsoft Corporation). Statistical analyses were performed using Analyse-It® (Analyse-It Software) and EP Evaluator Release 9 (Data Innovations LLC) (11).

For the duplicates in the EP-9 protocol and in the IFCC monitoring program, CV was calculated with the following formula:

$$CV_a = \frac{\sqrt{\frac{\sum(\Delta)^2}{n}}}{\bar{x}\sqrt{2}} \times 100\%$$

where CV_a is the analytical CV, Δ is the difference between duplicates, *n* is the number of duplicates, and $\bar{x}$ is the mean of the duplicates.

Results

Part One) Of the 134 laboratories who participated in the SKML-EQAS in the Netherlands, 90.3%, of the methods used in these laboratories, met the criteria of having σ >2, and 74.6% actually met standards for σ >4 with a TAE of 10%. With a TAE of 6%, 70.1% met the criteria for σ >2 and 41.0 % for σ >4 (Table 1 and Figure 1A and 1B). As an example, the Menarini HA8180 in one laboratory, had a calculated sigma of 32.7 due to a CV of 0.30% and a bias of 0.2%. Using the formula: σ=(TAE –

B)/CV resulted in a calculated sigma value of 32.7 ((10-0.2)/0.30=32.7). A very high sigma is generated with a low CV, generally below 1%. The opposite effect is also possible. A bias greater than the TAE of 10% will result in a negative sigma value. In practice a sigma of 32.7 is essentially equal to a sigma of 6 because world class performance is world class performance no matter how high the calculated sigma. A negative sigma indicates that the method fails to meet the set criteria more often than it achieves these set criteria. Table 1 details the range of sigma values that were calculated for each method, representing the distribution of the performance of that particular method between laboratories. However, this is better shown in Figure 1A. Most of the methods cluster around a certain sigma, except for the Roche method group, which shows the widest range of sigma values amongst users. Figure 1C compares the distribution of performance when using the alternative criteria of either a TAE of 10% and 4 sigma or a TAE of 6% and 2 sigma as criteria. Fewer individual laboratories would meet the criteria if the TAE was reduced than if the sigma target was increased.

Part two) Table 2 shows the results of the sigma calculations using data from a previous POCT evaluation study. All POCT methods, except the first evaluations of the Quo-Lab and the Quo-Test, had a sigma > 2 (TAE 10%) and this was independent of which CV was used (EP-5 or CV from duplicates in EP-9 lot number A or B). Only the B-analyst had a sigma >4 with TAE of 10% and sigma > two with a TAE of 6%. Table 2 and Figure 2A show that a method can pass the criteria of having a sigma >2 with TAE of 10% even if the bias is 5.7% and 6% (InnovaStar) due to a very low CV (1.9, 1.4 and 0.9%). Taking a TAE of 6% and two sigma as the criteria, this method would have failed (Figure 2B), as would the DCA Vantage and, potentially, the Afinion analysers. Figure 2C shows that the pass and fail rate for the POCT methods when using criteria of TAE of 10% and four sigma was almost equal to a TAE of 6% and two sigma.

Part three) Table 3 shows the current criteria of the IFCC monitoring program. The six SRMP evaluated showed excellent performance in 2015 concerning deviation from IFCC target, reproducibility and linearity. High sigma values were calculated at HbA1c values of 30, 60 and 90 mmol/mol, which are the levels at which deviation from IFCC target values are detailed on the annual monitoring certificate (Table 4). Again it can be seen from Table 4 that a CV <1.0% leads to sigma values > 6 (Tosoh G8 and Menarini HA8180). Figure 3A shows the results of the six SRMP using the bias at 60 mmol/mol and a TAE of 10% and Figure 3B shows the same results but with a TAE of 6%. All SRMP had a sigma > four with a TAE of 10% and a sigma > two with a TAE of 6% (Figure 3C).

Discussion

The global context This study used data from three different clinical settings in order to highlight the current state of the art of HbA1c analyzers that are in routine use around the world. The TAE of 10% (5 mmol/mol at an HbA1c value of 50 mmol/mol) was set as a default by the IFCC Task force and these goals should serve as a starting point for discussion with international stakeholders in the field of diabetes. This cut-point was chosen as it is based on the difference in HbA1c results, in two consecutive HbA1c tests that clinicians use as a guide to change therapy and is therefore a clinical decision limit (4). Previous guidance focused on the imprecision of methods with instruments expected to perform at a within laboratory CV of <3% for SI units and <2% for % (NGSP) units. The sigma metrics approach allows for both bias and imprecision to be taken into account when assessing analytical performance, generating a more comprehensive view of performance.

Performance within and between countries Results from part one of this study clearly show that the vast majority of individual laboratories (90%) are meeting these targets of a sigma value of >2 at a TAE <10%. Two method groups (Roche and Bio-Rad) performed less well with a wide range of sigma values calculated for the within method group. A reason for this large distribution might be that in this EQA program the Roche method was taken as one group (whole blood application and hemolysate

application, Gen. 2 and Gen. 3 reagent, using different instruments (Cobas Integra 800, Cobas C8000, Hitachi), giving considerable heterogeneity to the group. In addition, as with all methods factors such as technical/analytical skills required to perform the analysis and maintenance of instrument etc can have an impact on performance and thus sigma values. In general one can say that the smaller the distribution (range) of sigma values the more robust and reproducible that particular method is likely to be at that sigma level.

It is difficult to draw conclusions for some method groups due to small user numbers, however it can be seen that increasing the sigma cut point to >4 sigma leads to a reduction in the number of laboratories meeting that target (80%). This falls further to 75% with a TAE of 6% and sigma of >2.

Individual laboratories or networks of laboratories in any country can use the results of their EQA data to calculate their bias and imprecision and use this model to assess the performance of their own analyzer(s). The minimum expected standard is a sigma value of 2 or greater at a TAE of 10%.

How do POC devices compare? Part two data shows that, under strict evaluation conditions, most POCT devices perform to the same limits that are expected of laboratory analyzers. Only the Quo-Lab and Quo-Test did not reach these targets, however the Quo-Test did meet the targets after a manufacturer re-calibration. This is a significant finding as the question as to whether or not POC devices can be used for the diagnosis of diabetes is regularly raised. These findings would indicate that the majority of methods could be used in this setting however, it should be cautioned that the performance demonstrated here may not reflect the performance of these devices in the clinical setting, with multiple users and multiple lot numbers and further evidence is required before advocating the use of POCT devices for the diagnosis of diabetes. Indeed when the criteria are tightened to 4 sigma at a TAE of 10%, only one analyzer demonstrated a consistent ability to meet that target and with a reduced TAE of 6% and 2 sigma, only two instruments met the targets, indicating that at TAE of 10% at 2 sigma is the current state of the art of these analyzers.

Establishing the limit of performance of routing analyzers Part three data demonstrates how well routine laboratory analyzers can perform when directly calibrated to the IFCC RMP. Each of the six SRMPs performed exceptionally well meeting both the sigma >2 and sigma >4 targets at a TAE of 10%, in addition all also met the sigma >2 at a TAE of 6% target. This shows that routine analyzers have the capacity to meet these stringent performance targets and the aim should be to increase performance to these levels, in the clinical setting.

Future aims Overall, the performance of the majority of laboratory and POC analyzers for HbA1c is very good. Whilst not all analyzers are currently meeting the proposed target of >2 sigma at a TAE of 10%, some analyzers are already performing far in excess of this target, clearly indicating the capacity for excellent performance in routine clinical settings. Manufacturers and users of all instruments should strive to further improve the quality and consistency of the results they produce, to ensure that 100% of methods meet the basic quality targets. Once this has been achieved then the criteria can be further tightened, this will allow HbA1c analysis to become the benchmark for performance of all laboratory testing. It should be noted that the IFCC task force also advocated the use of analyzers that perform to the level of four sigma with a TAE of 10% when used for clinical trials, this is achievable with a range of current analyzers, but not all and care should be taken when selecting and instrument to be used for collecting trial data.

Should the performance guidelines but changed? This study also evaluated the impact of tightening the performance criteria to >4 sigma at 10% TAE or >2 sigma at 6% TAE. Both targets produced similar findings with slightly fewer methods attaining the latter target. Whilst increasing the sigma level will still allow some methods to pass with a considerable bias it could be used to give the

clinician the confidence to know that the result they have is true, within a specific range, and it is very unlikely that the value is ever likely to be outside of that range. However decreasing the TAE means that the range of values around the true value will be much smaller, with the caveat that up to 5% of results will not fall in that range. For the clinician this does leave some uncertainty as they would not know which results were outside of the expected range, however for a situation where minimum deviation from the true value is required then this would be the preferred choice.

The primary focus for the immediate future should be to ensure that all analyzers perform to the minimum standards set by the IFCC task force and this performance should be considered when deciding on which analyzer to choose, once this goal is achieved further tightening of the criteria can be discussed.

Acknowledgments

This manuscript could not have been written without having the results of the Dutch SKML. We Thank the SKML for sharing these results with us.

References

1. Weykamp C, John WG, Mosca A, Hoshino T, Little R, Jeppsson JO, et al. The IFCC Reference Measurement System for HbA1c: a 6-year progress report. Clin Chem. 2008;54(2):240-8.

2. Hoelzel W, Weykamp C, Jeppsson JO, Miedema K, Barr JR, Goodall I, et al. IFCC reference system for measurement of hemoglobin A1c in human blood and the national standardization schemes in the United States, Japan, and Sweden: a method-comparison study. Clin Chem. 2004;50(1):166-74.

3. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. Diabetes Care. 2009;32(7):1327-34.

4. Weykamp C, John G, Gillery P, English E, Ji L, Lenters-Westra E, et al. Investigation of 2 models to set and evaluate quality targets for hb a1c: biological variation and sigma-metrics. Clin Chem. 2015;61(5):752-9.

5. Lenters-Westra E, Roraas T, Schindhelm RK, Slingerland RJ, Sandberg S. Biological variation of hemoglobin A1c: consequences for diagnosing diabetes mellitus. Clin Chem. 2014;60(12):1570-2.

6. SKML. Available from: http://www.skml.nl/en/home

7. Lenters-Westra E, Slingerland RJ. Three of 7 hemoglobin A1c point-of-care instruments do not meet generally accepted analytical performance criteria. Clin Chem. 2014;60(8):1062-72.

8. Medicine IFoCCaL. Available from: http://www.ifcchba1c.net/node/1.

9. Medicine IFoCCaL. IFCC HbA1c Network. Available from: http://www.ifcchba1c.net/network/approved?page=1.

10. SKML. Scoring and Reporting System. Available from: http://www.skml.nl/en/home/schemes/reportings/inputs/great-britain-frgi.

11. Innovations D. EP Evaluator Available from: https://www.datainnovations.com/products/ep-evaluator.