Peer Review File

Article information: https://dx.doi.org/10.21037/jlpm-23-21

Reviewer A

This is a small study investigating the analytical performance of QuikRead go HbA1c analyser. The paper is short and well written but there are some critical information lacking. We know from external quality assessment (EQA) data that Cobas Tina-quant shows slightly higher HbA1c results than the IFCC reference measurement procedure (LC-MS) on average. Therefore, it is critical for this study that the authors give information about the bias and imprecision of the comparison method used. If the laboratory (Cobas) has a positive bias compared with the IFCC reference, could this in fact explain the negative bias of QuikRead? And further, we can see from Figure 1 that the variability is greater in high level compared with the low level, could this be explained by poorer repeatability of the two methods in high level? The author should reflect on these issues and discuss the results accordingly. When a method-comparison study is performed it is critical to have a much information as possible about the comparison method so that the evaluation of the POC instrument in question is done fair and correctly.

The performance of comparison method (Roche cobas c513 and Tina-quant haemoglobin A1cDx Gen 3) was verified following procedures meeting the needs of ISO 15189. Comparative instrument in this verification was Roche cobas c501 (Gen 3 reagent). Verification was done with a large sample set covering the measuring range 25-110 mmol/mol.
Current comparison method is routinely assessed and compared against target values assigned by reference laboratory (EQA provided by Labquality). Performance during 2022-2023 has been acceptable bias varying between -5,5% and 2,6%.
Changes in the text: The verification of the comparison method is briefly described on line 115 forward. The performance of the comparison method is described on line 222 forward. We modified conclusions according to the questions raised here.

Figure 1 shows Cobas minus QuikRead on the y-axis. This means that a positive deviation in the Figure actually means a negative deviation for QuikRead. I suggest turning this around (showing QuikRead minus Cobas on the y-axis) to avoid confusions. As an example, the authors stat that the method deviation was high above 48 mmol/mol with a deviation of -11 to +20 mmol/mol. In fact, the correct should be +11 to -20 mmol/mol. This mistake should be corrected in abstract and results.
Reply: We redraw Figure 1 with the suggested change to present QuikRead go minus Cobas on the y-axis. We also corrected the
Changes in the text: Figure 1 and lines 98-99 and the abstract.

Figure 1. Why is the 95 confidence interval (CI) of the regression line included in this figure? I do not see the relevance. Instead, the legend should include the mean bias and the limit of agreement.
Reply: We removed the CI from the regression line, but added 95 % CI on the mean difference together with the line of agreement, which was not printed properly on the original figure.
Changes in the text: Figure 1 redrawn.

I miss a Table which summarize the data. I miss information about the mean bias in level below and above 48 mmol/mol (with 95 CI, or p-values) and standard deviations. It is not possible to see from the figure whether the bias above 48 is different from the bias below 48.
Reply: We added more specific data of the results in table 1.
Changes in the text: Added table 1.

The authors state that POC devices are generally not recommended in diabetes diagnostics (line 136), but is this true? They should give a reference for this statement.

Reply: Despite of development and standardisation of POC HbA1c devises, the American Diabetes Association recently (2022) updated recommendation for classification and diagnosis of diabetes does not support their use in DM diagnosis.
Changes in the text: We added the reference (line 167, ref. 27).

I suggest deleting the sentence on lines 46-48 (Quantitative measurement of C-reactive protein and detection of Streptococcus A bacteria with QuikRead go correlates well or adequately to laboratory reference method). This is not relevant for the study.
Reply: We wish to keep the sentence, as we'd like to express that QuikRead go performance on certain analytes is rather well established, but there are no publications on its performance with HbA1c. We tried to rephrase the sentence a bit.
Changes in the text: With quantitative measurement of C-reactive protein and detection of Streptococcus A bacteria the device correlates well or adequately to laboratory reference methods (19–22). (Previously: Quantitative measurement of C-reactive protein and detection of Streptococcus A bacteria with QuikRead go correlates well or adequately to laboratory reference method (19–22).).

Reviewer B

Point of care (POC) instruments are increasingly being used in all parts of the health care system – both in hospitals (on the wards, in the ER, etc.) and in primary care settings. POC can be extremely useful in acute settings, or in decentralized settings where getting a result from a central laboratory can delay diagnosis and treatment, or when long transportation time of samples may introduce preanalytical issues. However, it is vital that we know the POC instruments work as intended so that we can trust the results, and manufacturer independent evaluations are therefore essential. I therefore warmly welcome this study. However, there are some shortcomings, and I believe some improvements could be made.
Major comments:
1. In the Introduction, POC is described as "one thing". However, there are many systems, many manufacturers, and analytical quality varies a lot between the systems. This should not be ignored.
Reply: We rephrased the section describing HbA1c POC in the Introduction. We wish this improved the text. We hope the analyser dependent nature of the data is expressed better now.
Changes in the text: However, method evaluation and comparison studies show variable, even unacceptable, performance for various HbA1c POC devices (12,13). Some studies claim certain POC devices to perform well compared to reference laboratory method (14), while others show bias, most often negative (15–17). The data depends on the analyser and the study setting, and should not be generalised. (line 41 forward)
Previously: However, method comparison studies show variable, even unacceptable, performance for HbA1c POC devices (12,13). Some studies claim POC devices to perform well compared to reference laboratory method (14), while others show bias, most often negative (15–17).

2. It is perhaps not explicitly stated, though it is implicit, that the comparison method is the "gold standard". How did you verify the trueness of the comparison method? How did this method perform on EQA during the evaluation period? How does the general level of HbA1c on this measurement system compare to a reference measurement system? If you don't know the trueness of this method, it is not possible to judge the performance of the POC method either. A perceived "negative bias" on the POC method may actually be due to a "positive bias" in the comparison method.
Reply: The performance of comparison method (Roche cobas c513 and Tina-quant haemoglobin A1cDx Gen 3) was verified following procedures meeting the needs of ISO 15189. Comparative instrument in this verification was Roche cobas c501 (Gen 3 reagent). Verification was done with a large sample set covering the measuring range 25-110 mmol/mol.
Current comparison method is routinely assessed and compared against target values assigned by reference laboratory (EQA provided by Labquality). Performance during 2022-2023 has been acceptable bias varying between -5,5% and 2,6%.

Changes in the text: The verification of the comparison method is briefly described on line 74 forward. The performance of the comparison method is described on line 132 forward.

3. Why did you not perform duplicate measurements of all the samples? This is a weakness that needs to be addressed. I do not think the conclusion in line 103 is warranted.
Reply: Due to lack of time and supplies, we were not able to measure all samples in duplicates with QuikRead go. It is true that duplicate measurements would have strengthened the data.
Changes in the text: Due to limited resources, we did not systematically evaluate method repeatability with blood samples, but the samples (n=6) with deviating results compared to cobas c513 gave 1-5 % (1-3 mmol/mol) difference in reanalysis. Lack of duplicate QuikRead go measurements is a weakness of this study, and comprehensive repeatability for the device cannot be calculated from these data. (line 111 forward).
Previously: We did not systematically evaluate method repeatability with blood samples, but the samples (n=6) with deviating results compared to cobas c513 gave 1-5 % (1-3 mmol/mol) difference in reanalysis. The repeatability may be considered adequate based on these limited data.

4. It would have been preferable to define quality goals apriori. You simply state that the performance is not good enough – but what would have been good enough? Could you use accuracy as a quality goal, since you only have singlicates? What would be good enough analytical quality for monitoring diabetes? What would be good enough to diagnose diabetes? You need to say something about this in order to support your conclusion.
Reply: It is true that specific quality goals were not set. However, we now related the data and findings to eg.other publications.
Changes in the text: Rows 123-137 added.

5. Limits of allowable percentage deviation is sometimes defined in such studies, at least above a certain threshold, as it is hard, and not necessarily useful, to define absolute (not relative) limits in the higher ranges. It could be useful to report relative deviation in addition to absolute deviation.
Reply: We presented the data also with a percentage difference (Bland Altman) in Figure 2. In figure 1 and 2 the 95 % CI of the mean difference is shown together with the line of agreement. This emphasizes the significance of the bias (95 % CI not intercepting the line of agreement).
Changes in the text: Figure 2 and 95 % CI in the mean difference in Figure 1 and 2. Added the lines of agreement in Figure 1 and 2.

6. You describe that performing the study under optimal conditions is a strength because it minimizes variation. This is true - but it can also be considered a weakness. What if the system is somehow difficult to use and prone to mistakes, and intended end users make more mistakes than professionals? Evaluating user friendliness is an important aspect of evaluating POC systems, and evaluation among end users is therefore important. This should be discussed in the conclusion.
Reply: We were not able to evaluate user friendliness in this study. This, together with possible errors in sample handling or measurement should be done in another study. We discussed this as below.
Changes in the text: Added text: A sample collector was used with each sample according to manufacturer's written instructions with no excess blood outside the collector. The analysis is validated on only 1 ul of blood in the collector and any excess blood can cause result deviation. This was acknowledged in this study. User friendliness was not widely evaluated, as a single person performed all the analysis. This should be tested in a separate study with an emphasis on the possible effect of misuse of sample collector and imprecise amount of blood collected. (line 156 forward).

7. Line 148-150. There are already studies out there, for instance SKUP reports (www.skup.org), which the authors do not seem aware of.
Reply: The authors are aware of the thorough work of SKUP. With the final sentence we emphasized testing QuikRead Go HbA1c in a bigger scale, which to our knowledge has not been performed.
Changes in the text: We rephrased the last sentence to a less ambiguous form.

Minor comments:

1. Line 110: biological variation.
Reply: Within-subject corrected to biological variation
Changes in the text: The term biological variation is used (line 121).
2. Line 112: I don't understand this sentence, rephrase?
Reply: We rephrased the sentence: Thereafter, the bias observed can be considered clinically relevant. Previously published data has shown similar negative bias for other HbA1c POC devices (24).
Changes in the text: Rephrased sentences (above) on line 123 forward.

Reviewer C

A very robust study showing the comparison of a point of care device to laboratory device.

Please include more information in the figure caption to explain the results independently including how the regression was performed.
Reply: We improved the figure caption and added more specific information. Figure 1 was also modified.

Changes in the text: Figure 1.

Reviewer D

This study is a method comparison of HbA1c for QuikRead vs Roche cobas TinaQuant. It is important for POC HbA1c to be analytically sound, especially as HbA1c is a standardised measurand with guideline based decision limits, and so correct management can be achieved.

The relatively large number of samples in the method comparison is a strength of this article.

Suggestions to incorporate:

- It is unclear if the "control repeatability" is within day or between day. Please specify.
Reply: Control repeatability was between day control repeatability. This is now specified in the text.
Changes in the text: "Between day" (…control repeatability…) added (line 86).

- Try to conform to single type of spelling system - eg "standardized" (with z) used on page 4 but "anonymised" (with s) used in the next paragraph.
Reply: We eliminated US spelling and used UK spelling.
Changes in the text: UK spelling used throughout the document.

- though Passing-Bablok regression was mentioned in the methods, this was not presented in the results or figures. Include the equation in the results and a scatter plot of this in the figures.
Reply: This was an unfortunate typo. We did not include Passing-Bablok regression model in the data presented here. Instead, we used Bland Altman analysis.
Changes in the text: Correction on line 79.

- It is worth noting that though results were compared against Roche TinaQuant, which is standardised, this is not a "true" gold standard. In particular, see the paper "EurA1c: The European HbA1c Trial to Investigate the Performance of HbA1c Assays in 2166 Laboratories across 17 Countries and 24 Manufacturers by Use of the IFCC Model for Quality Targets" where the between lab CV for Roche was 4.4% (in IFCC units), which was more than many other methods. It would be useful to quote the imprecision seen from the Roche assay locally. It appears that most of the difference between assays is either from imprecision or "matrix effects" rather than bias per se, as overall bias appears minimal.
Reply: We took this comment and paper into consideration and added data and references into the manuscript.

Changes in the text: Line 119-131.

- It could be interesting to see a version of Bland Altman plot also in percentage differences on the y-axis given the point that the bias was unacceptable at higher HbA1c; given that there appears to be more scatter at higher HbA1c in a concentration dependent manner. My own assessment of the graph is there doesn't appear to be a clinically significant bias, especially so relative to the scatter itself.
Reply: We presented the data also with a percentage difference (Bland Altman) in Figure 2. In figure 1 and 2 the 95 % CI of the mean difference is shown together with the line of agreement. This emphasizes the significance of the bias (95 % CI not intercepting the line of agreement).
Changes in the text: Figure 2 and 95 % CI in the mean difference in Figure 1 and 2. Added the lines of agreement in Figure 1 and 2. Added discussion on lines 116-117.

- When comparing acceptable deviation between methods, it would be useful to use total allowable error from an EQA or proficiency testing guideline as a benchmark to see how many samples were outside of this difference
Reply: From 151 samples analyzed in this study the bias between POC and comparison method was more than +/- 5 mmol/mol with 31 samples. However, with only four samples the bias observed with QuikRead Go was positive. The samples showing negative bias on QuikRead Go varied both below and over the limit of 48 mmol/mol. As focusing on the clinical outcome (DM diagnosis), the bias values were not discussed. However, it is noteworthy that there is a variation in results below the cut-off value 48 mmol/mol, although not as prominent as with high level samples. The individual deviating values can be seen in the figures.
Changes in the text: Added lines 74-77. Added information on comparison method EQA (lines 132-135).