The impact of inter-laboratory glucose bias on the diagnosis of gestational diabetes mellitus: Comparison of common automated central laboratory methods

Cathrine Munk Scheuer, Casper Duevang Tvarno, Charlotte Gils, Julie Dahl Ravn, H. David McIntyre, Dorte Møller Jensen, Peter Damm, Jeannet Kepp Bruun Lauenborg, Tine Dalsgaard Clausen, Martin Overgaard

Introduction

Gestational diabetes mellitus (GDM) is one of the most common pregnancy-related diseases and is characterized by hyperglycemia first noted during pregnancy [1]. Untreated, it can result in serious complications for both mother and offspring [2].

Abbreviations: CI, Confidence interval; EQA, External Quality Assurance; GDM, Gestational Diabetes Mellitus; ID-MS, Isotope Dilution Mass Spectrometry; L-ESB, Esbjerg Hospital Laboratory; L-HII, Hillerød Hospital Laboratory; L-NAS, Naestved Hospital Laboratory; L-OUH, Odense University Hospital Laboratory; L-RAN, Randers Hospital Laboratory; OGTT, Oral Glucose Tolerance Test; RFB, Reference Institute for Bioanalytics.

Material and methods: A prospective cohort study of women (n = 110) referred for second-trimester GDM diagnostics using a 75 g oral glucose tolerance test. Maternal glucose was assessed from venous plasma at fasting, 1 h and 2 h. Venous blood were collected in Fluoride Citrate tubes and frozen. Samples were analyzed at five central laboratories using four different automated glucose Hexokinase methods and GDM prevalence was evaluated according to WHO2013 diagnostic criteria.

Results: Maximum inter-laboratory bias was 0.19, 0.30 and 0.27 mmol/L in fasting, 1 h and 2 h samples, respectively. GDM prevalence ranged 30.0–41.1% across laboratories.

Conclusion: Inter-laboratory bias for mean venous glucose was low and within desirable limits. Nonetheless, the impact on GDM prevalence was considerable, which may inappropriately affect clinical practice.
The international WHO2013 guidelines for diagnosing GDM recommend a 2 h 75 g oral glucose tolerance test (OGTT) between 24 and 28 gestational weeks with the diagnostic thresholds of either a fasting plasma glucose ≥ 5.1, a 1 h plasma glucose ≥ 10.0 or a 2 h plasma glucose ≥ 8.5 mmol/L [3,4]. These strict diagnostic criteria have proven to identify pregnancies that may benefit from treatment [5]. However, the clinical value of implementation of the WHO2013 criteria has been questioned, as their use substantially increases the prevalence of GDM, which challenges healthcare systems [2,6]. Another issue using the relatively low WHO2013 criteria is that at lower glucose concentrations close to the population mean, even small variations in glucose measurements lead to large variations in the diagnostic rates of GDM [7]. Thus, a renewed focus on laboratory methods for glucose measurements is required due to their dominant role in GDM diagnostics [8-11].

Analytical procedures, e.g. analytical method, calibration and equipment, play an important role when assessing bias in glucose measurements. Previously, the Danish standard method for GDM diagnosis was to use capillary blood and point-of-care-testing equipment, which has proven less suitable in diabetes diagnostics [12-15]. Today, glucose measurements to diagnose GDM are predominantly performed with venous plasma using a hexokinase method performed on a fully automated analysis system.

The analytical imprecision for glucose measurements is generally low [12]. Despite the use of external quality assurance (EQA) systems, it is difficult for laboratories to accommodate the desired analytical performance specifications based on biological variation (i.e. imprecision < 2.5% and bias < 2.4%) [16]. In addition, the Danish hospital laboratories use fully automated chemistry analyzers from different manufacturers. Thus, differences in analytical procedures between laboratories may result in bias, which could affect the prevalence and uniformity of GDM diagnosis.

Previous studies have primarily assessed bias between sampling procedures (e.g. different types of blood collection tubes or blood specimens) and specific analytical methods, but no studies have yet investigated inter-laboratory differences in fully automated analytical methods used in GDM diagnostics. We aimed to determine inter-laboratory bias for venous plasma glucose measurements and evaluate the impact of analytical procedures across laboratories on the GDM prevalence defined by WHO2013 diagnostic criteria.

Materials and methods

1.1. Participants

Blood samples were collected at the department of Clinical Biochemistry at Odense University Hospital, Region of Southern Denmark, during August 2021. Pregnant women referred for a diagnostic OGTT were eligible. To be referred to an OGTT, women must present with at least one risk factor for GDM according to Danish guidelines: Glucosuria, pre-pregnancy BMI ≥ 27 kg/m². GDM in a previous pregnancy, family history of diabetes, previous birth of a child ≥ 4,500 g, polycystic ovarian syndrome or multiple pregnancy [17]. There were no exclusion criteria. All participants gave written consent. The study was approved by the National Committee on Health Research Ethics (ref. 78723) and conducted according to the Helsinki Declaration as revised in 2013.

1.1.1. Sampling and analytical procedures

Venous blood samples were drawn at fasting, 1 h and 2 h and collected in Fluoride-Citrate Mix tubes (Greiner BioOne). Samples were stored at room temperature until centrifugation within 1 h, whereafter plasma was frozen at −80 °C within 4 h. From each participant, samples were aliquoted to five tubes, one for each of five different laboratories. Samples were divided by date of inclusion into four batches (B1-B4) which were analyzed with one-month intervals (day 1, day 30, day 60 and day 90, respectively) (Fig. 1). All samples were analyzed using automated hexokinase methods performed on random-chemistry analyzer modules in each of the participating laboratories: Odense University Hospital (L-OUH), Roche Cobas 8000 (c701), Roche Diagnostics DenmarkRegionshospitalet Randers (L-RAN), Siemens Atellica (CH 930), Siemens Healthineers DenmarkSydvestjysk Sygehus Esbjerg (L-ESB), Abbott Alinity (Alinity c), Abbott Laboratories DenmarkKNaestved Sygehus (L-NAS), Siemens Dimension Vista, Siemens Healthineers DenmarkNordsjællands Hospital Hillersod (L-HIL), Siemens Dimension Vista, Siemens Healthineers Denmark.

At all laboratories, samples were handled according to a pre-planned pre-analytical protocol regarding freezing/thawing, centrifugation, storage, transport etc. All five laboratories are accredited according to DS/EN ISO 15189 and met the requirements for inter-laboratory glucose comparisons in External Quality Assurance (EQA) distributions during the study period. Inter-assay CV’s for automated glucose methods across all laboratories and glucose levels ranged from 1.3 to 3.4%.

1.1.2. Reference laboratory

The laboratory L-OUH was chosen as the reference laboratory. L-OUH runs the Roche Cobas system c701, which uses a hexokinase method with photometric detection.

The method is traceable to the National Institute of Standards and Technology standards and an isotope dilution mass spectrometry (ID-MS) reference method, as listed by the Joint Committee for Traceability in Laboratory Medicine [18]. The accuracy of the L-OUH method was monitored using the “KS survey program” by the Reference Institute for Bioanalytics (RFB) and the “HK19 program” by the Danish Institute for EQA for Laboratories in the Health Sector. During the study period, four RFB distributions to L-OUH (two samples per distribution) showed a mean bias of + 1.5% to the ID-MS reference method. The range of glucose concentrations was 3.9 – 19.9 mmol/L.

1.1.3. Statistical methods

L-OUH was chosen as reference in all inter-laboratory comparisons as data was collected at this site. Fasting, 1 h and 2 h mean glucose measurements across laboratories were evaluated using paired t-test or Wilcoxon test depending on normality of distribution assessed by Shapiro-Wilk test. Correlations between fasting, 1 h and 2 h glucose and the reference laboratory values were assessed by Pearson’s correlation. Inter-laboratory bias and 95% limits of agreement (from here, presented as 95% confidence intervals (CI)) between the five laboratories were evaluated by modified Bland-Altman analyses of fasting, 1 h and 2 h glucose. Grubbs z-score and visual evaluation of the Bland-Altman plots were used to detect potential outliers and identified a total of five samples as such (at L-RAN, L-NAS, L-NOH). However, the measurements were all within the range of the individual laboratory’s glucose values and were, therefore, not excluded from the analyses.

The effect of sample freezing on glucose stability was evaluated by comparing frozen and unfrozen samples from 53 participants at all OGTT time-points at L-OUH by paired t-test, and glucose stability during the study period was estimated by one-way ANOVA to assess intra-laboratory variation over time.

GDM prevalence by WHO2013 diagnostic criteria (fasting glucose ≥ 5.1 mmol/L and/or 1 h glucose ≥ 10.0 mmol/L, and/or 2 h glucose ≥ 8.5 mmol/L) was compared using McNemar’s test for correlated proportions.

Complete-case analyses (complete results for glucose analyses across all laboratories and all OGTT time-points) were performed. Two-sided P values < 0.05 were considered statistically significant. All statistical analyses were performed using IBM SPSS Statistics, Version 25.

Results

Of 110 women included, a total of 90 women had complete glucose analyses results across laboratories and were included in the
comparisons. Participation rate and study design are illustrated in Fig. 1.

1.2. Inter-laboratory comparison

We evaluated venous plasma glucose during diagnostic OGTTs in pregnancy (fasting, 1 h and 2 h). Frozen plasma aliquots were distributed for glucose measurements across participating laboratories with one-month intervals. Mean glucose bias with L-OUH as reference was low across laboratories regardless of glucose level (OGTT time-point) ranging from to \(0.03 \text{ mmol/L} (95\% \text{ CI } 0.22 - 0.16 \text{ mmol/L})\) to \(0.24 \text{ (95\% CI } 0.63 - 0.14 \text{ mmol/L})\) (Fig. 2 and Table 1). L-NAS demonstrated the largest bias at all glucose levels and most pronounced
Glucose data are given as mean (±SD) and bias as mean (95% CI). Inter-laboratory comparisons of mean glucose and mean bias across central laboratories with L-OUH as reference.

For the 1 h glucose (Table 1). However, from the modified Bland-Altman plots, glucose levels ranging 8 – 12 mmol/L exhibited a concentration-dependent trend of increasing negative bias in laboratories using Siemens methods (L-RAN, L-NAS and L-HIL) compared to the Roche Cobas method (L-OUH) (Fig. 2).

The level of agreement evaluated by linear regression analysis was high across automated glucose methods ranging from \( r = 0.997 \) to \( r = 0.997 \), with less than ten values outside the 95% CI for all analyses and only a few outliers (Fig. 2 and Supplementary Table S1). All methods showed the highest agreement for 1 h and 2 h glucose (r ranging 0.989 – 0.997).

Two laboratories (L-NAS and L-HIL) used the same analytical equipment, but the difference in mean glucose was nonetheless significantly different at all time-points (Table 1). Therefore, to explore potential batch-dependent differences, we performed an explorative post-hoc sensitivity analysis stratified by batch (number one to four), comparing the mean glucose values from L-NAS and L-HIL with the reference, respectively, which overall showed no batch-dependent differences (Supplementary Table S4).

In the assessment of the sample freezing effect, mean glucose concentrations did not differ significantly (ranging 0.01 – 0.07 mmol/L) (Supplementary Table S2). Similarly, the glucose stability, where selected samples (n = 13) were re-analyzed four times with one-month intervals, showed a non-significant variation over time (i.e. measured glucose concentration remained stable over time) (\( p = 0.995 \)) (Supplementary Table S3).

1.2.1. Variations in GDM prevalence

The GDM prevalence across laboratories ranged from 30.0% (95% CI 20.5% – 39.5%) to 41.1% (95% CI 30.9% – 51.3%) corresponding to a difference of 11% points (Table 2). Only L-NAS was significantly different from the reference L-OUH driven by differences in fasting values, which also accounted for a mean of 82% (27/33) of GDM cases across laboratories (Table 2).

Discussion

In the present study, we demonstrated low bias in glucose measurements across laboratories, methods and glucose levels. However, the resulting GDM prevalence ranged considerably from 30.0% to 41.1% across laboratories.

1.3. Inter-laboratory bias

Inter-laboratory bias evaluated in the present study has not been investigated before in a prospective cohort setting nor in relation to GDM. We used L-OUH as the reference method for each sample and determined a low inter-laboratory bias ranging from –0.03 to –0.24 mmol/L (–0.6% – 3.1%) for all time-points of the OGTT. During the study period of 4 months, L-OUH displayed a mean bias of + 1.5% as compared to an ID-MS reference method. Thus, the observed inter-laboratory bias was overall lower than the desired analytical performance (bias < ±2.4%) [16].

A number of factors may have contributed positively to these findings. Firstly, we examined analytical bias robustly over a four-month

---

**Table 1**

Inter-laboratory comparisons of mean glucose and mean bias across central laboratories with L-OUH as reference.

<table>
<thead>
<tr>
<th>Equipment/system</th>
<th>L-OUH</th>
<th>L-RAN</th>
<th>L-ESB</th>
<th>L-NAS</th>
<th>L-HIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting glucose (mmol/L)</td>
<td>Roche Cobas</td>
<td>Siemens Atellica</td>
<td>Abbott Alinity</td>
<td>Siemens Dimension Vista</td>
<td>Siemens Dimension Vista</td>
</tr>
<tr>
<td>Bias fasting glucose (mmol/L)</td>
<td>Reference</td>
<td>–0.03 (±0.27 - 0.21) *</td>
<td>–0.03 (±0.22 - 0.16) *</td>
<td>–0.16 (±0.36 - 0.04) *</td>
<td>–0.07 (±0.31 - 0.17) *</td>
</tr>
<tr>
<td>1 h glucose (mmol/L)</td>
<td>9.30 (±1.78)</td>
<td>7.72 (±1.73) *</td>
<td>7.74 (±1.78)</td>
<td>7.56 (±1.72) *</td>
<td>7.65 (±1.71) *</td>
</tr>
<tr>
<td>Bias 1 h glucose (mmol/L)</td>
<td>Reference</td>
<td>–0.08 (±0.35 - 0.38) *</td>
<td>–0.06 (±0.35 - 0.23) *</td>
<td>–0.24 (±0.63 - 0.14) *</td>
<td>–0.16 (±0.52 - 0.21) *</td>
</tr>
<tr>
<td>2 h glucose (mmol/L)</td>
<td>9.01 (±1.39)</td>
<td>6.52 (±1.36) *</td>
<td>6.55 (±1.40) *</td>
<td>6.39 (±1.35) *</td>
<td>6.49 (±1.40) *</td>
</tr>
<tr>
<td>Bias 2 h glucose (mmol/L)</td>
<td>Reference</td>
<td>–0.09 (±0.45 - 0.21) *</td>
<td>–0.05 (±0.33 - 0.22) *</td>
<td>–0.22 (±0.50 - 0.07) *</td>
<td>–0.12 (±0.52 - 0.28) *</td>
</tr>
</tbody>
</table>

---

**Table 2**

Prevalence of gestational diabetes across central laboratories with L-OUH as reference.

<table>
<thead>
<tr>
<th>Equipment</th>
<th>L-OUH</th>
<th>L-RAN</th>
<th>L-ESB</th>
<th>L-NAS</th>
<th>L-HIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDM diagnosis b</td>
<td>Roche Cobas</td>
<td>Siemens Atellica</td>
<td>Abbott Alinity</td>
<td>Siemens Dimension Vista</td>
<td>Siemens Dimension Vista</td>
</tr>
<tr>
<td>95% CI</td>
<td>41.1 % (37)</td>
<td>41.1 % (37)</td>
<td>36.7% (33)</td>
<td>30.0% (27)</td>
<td>37.8 % (34)</td>
</tr>
<tr>
<td>95% CI (≥5.1 mmol/L)</td>
<td>30.9 % (51.3)</td>
<td>30.9 % (51.3)</td>
<td>26.7 % (46.7)</td>
<td>20.5 % (39.5) *</td>
<td>27.7 % (47.9)</td>
</tr>
<tr>
<td>1 h glucose threshold (≥5.1 mmol/L)</td>
<td>34.4 % (31)</td>
<td>34.4 % (31)</td>
<td>30.0 % (27)</td>
<td>24.4 % (22) *</td>
<td>28.9 % (26)</td>
</tr>
<tr>
<td>95% CI</td>
<td>24.6 – 44.3</td>
<td>24.6 – 44.3</td>
<td>20.5 – 39.5</td>
<td>15.5 – 33.4 *</td>
<td>19.5 – 38.3</td>
</tr>
<tr>
<td>1 h glucose threshold (≥10.0 mmol/L)</td>
<td>41.4 % (13)</td>
<td>13.3 % (12)</td>
<td>14.4 % (13)</td>
<td>10.0 % (9)</td>
<td>12.2 % (11)</td>
</tr>
<tr>
<td>95% CI</td>
<td>7.1 – 21.7</td>
<td>6.3 – 20.4</td>
<td>7.1 – 21.7</td>
<td>3.8 – 16.2</td>
<td>5.4 – 19.0</td>
</tr>
<tr>
<td>2 h glucose threshold (≥8.5 mmol/L)</td>
<td>12.2 % (11)</td>
<td>7.8 % (7)</td>
<td>10.0 % (9)</td>
<td>7.8 % (7)</td>
<td>6.7 % (6)</td>
</tr>
<tr>
<td>95% CI</td>
<td>5.4 – 19.0</td>
<td>2.2 – 13.3</td>
<td>2.8 – 16.2</td>
<td>2.2 – 13.3</td>
<td>(1.5 – 11.9)</td>
</tr>
</tbody>
</table>

**Prevalence of gestational diabetes based on glucose measurements from venous whole blood collected in FC-Mix tubes, frozen and thawed prior to analysis.**

Laboratories: L-OUH = Odense, L-RAN = Randers, L-ESB = Esbjerg, L-NAS = Naestved, L-HIL = Hillerød. GDM: Gestational diabetes mellitus. CI: Confidence Interval. Data given as percentages (N) and (95% CI).

* Significantly different from L-OUH.

b GDM diagnosis by WHO2013 criteria.
period with multiple samples. By contrast, EQA schemes only distribute a few samples, which the participating ISO15189 accredited laboratories routinely use to access accuracy. Secondly, the comparisons in the present study were made using track-automated chemistry analyzer modules with random access and the same enzymatic principle for glucose analysis (i.e. hexokinase/glucose-6-phosphate dehydrogenase). Further, all participating laboratories used the same sample handling protocol and random chemistry analyzer modules within each laboratory. Thus, our study design has likely minimized systematic errors related to instrument module bias.

Calibration issues caused by calibrator/reagent lot-to-lot variation may have affected our results [19]. Among the five participating laboratories, both L-HIL and L-NAS used the Siemens Dimension Vista platform for automated glucose analyses, but L-NAS measured the lowest glucose level among all participating laboratories. Further, the Bland-Altman plots indicated a method-dependent bias increasing with glucose concentration. Nevertheless, the difference between L-HIL and L-NAS was < 0.1 mmol/L across OGTT time-points and thus within the range generally accepted as inter-laboratory instrument module variation.

Two confounding factors, sample freezing and storage over time, could potentially diminish or increase possible time-dependent glucose variations during our study. However, neither the comparison of frozen versus unfrozen samples nor the glucose stability test over time with replicate-sample analyses showed any significant influence of these factors in our study. These results were in line with a previous study reporting sustained stability of plasma glucose despite long-term storage [20].

In conclusion, we conducted a robust prospective study to evaluate the inter-laboratory glucose bias between central laboratories. The results from the present study highlight that standardization of glucose measurements in central laboratories – with regard to blood sampling procedures (i.e. venous plasma in Fluoride-Citrate Mix tubes) and automated analyzer systems using identical assay principles – constitute an important prerequisite for high diagnostic accuracy in GDM.

1.3.1. Consequences of bias on GDM prevalence

Although the observed inter-laboratory bias was low, glucose values showed some variation across platforms, which resulted in significant differences in WHO2013 GDM prevalence in the present study. The largest inter-laboratory variation in GDM diagnosis arose from the fasting glucose values, which also accounted for 82% of GDM cases according to the WHO2013 criteria. Thus, for countries/sub-regions implementing the WHO2013 criteria, the choice of glucose measurement regime may significantly influence GDM diagnosis rates. Previous studies have reported that GDM prevalence may differ considerably as the result of uncorrected pre-analytical variations, in particular, the efficacy of glycolysis inhibition and differences in the analytical methods [7,9,14,21–23]. However, such confounding factors should be negligible in the present inter-laboratory comparison. Nonetheless, the inescapable bias in glucose measurements stresses the dilemma of GDM diagnostics: With diagnostic thresholds at high glucose concentrations, a large analytical bias will not greatly affect GDM prevalence. However, when applying strict diagnostic criteria (e.g. WHO2013) using low thresholds closer to the population mean and several glucose time-points for diagnosis, large variations in prevalence will occur even with low analytical bias. Taken together, the high diagnostic accuracy of GDM hinges on standardization of glucose measurements in central laboratories, both with respect to blood sampling procedures and automated analyzer systems. Such factors should be prioritized whenever possible while achieving a reasonable balance between high diagnostic sensitivity and clinical implications [24]. Nevertheless, with current methods, we inevitably face the challenge of achieving a “true” precise GDM diagnosis [9].

1.3.2. Clinical impact of findings

The low inter-laboratory bias in glucose measurements fulfilled the high standards of current Danish recommendations. Despite low bias, the clinical impact on GDM prevalence was considerable for one laboratory, which diagnosed 11 percent points fewer women as having GDM than the reference laboratory. Such differences may cause inequality in the local burden on healthcare providers. These findings stress the importance of continuous EQA and advocate using only the venous plasma automated central laboratory methods when diagnosing GDM to reduce inter-laboratory bias.

1.3.3. Strengths and limitations

The strength of the present study pertains to the batch-wise glucose comparisons over time, producing a more robust method comparison analysis compared to a cross-sectional analysis. A limitation of the bias assessment was the comparison of each laboratory glucose result to L-OUH as the reference laboratory. Although we assessed the bias of L-OUH glucose to the EQA material, which is traceable to a certified reference method, we cannot directly determine the absolute bias of the common automated clinical chemistry platforms used in this study. However, during the study period, all laboratories adhered well to reference-method-determined glucose values in EQA performance reports (data not shown), and all methods employed calibration standards traceable to the certified EQA reference material.

Finally, our results on inter-laboratory bias in glucose measurement are applicable to the general population of pregnant women, however, the same does not apply to our result on the diagnostic rates of GDM since these numbers are based on samples from women with risk factors.

Conclusion

Inter-laboratory bias of venous glucose measurements was in general low and within desirable limits. However, despite low degree of inter-laboratory variation, the clinical impact on GDM diagnosis was considerable. Such differences may inappropriately affect clinical practice in the current setting, where the diagnosis of GDM relies solely on the yet imperfect measurement of glucose.

CRediT authorship contribution statement

Cathrine Munk Scheuer: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. Casper Duevang Tvarns: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Writing – review & editing. Charlotte Gils: Methodology, Resources, Writing – review & editing. Julie Dahl Ravn: Investigation, Writing – review & editing. H. David McIntyre: Methodology, Writing – review & editing. Dorte Møller Jensen: Methodology, Writing – review & editing. Peter Damm: Methodology, Writing – review & editing. Jeannet Kepp Bruun Lauenborg: Funding acquisition, Writing – review & editing. Tine Dalsgaard Clausen: Conceptualization, Methodology, Writing – review & editing. Martin Overgaard: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.
Acknowledgements

The authors would like to thank Niels Strømsvig Larsen and Ida Hvam for excellent technical assistance.

Funding

The study was funded by the Joint Regional Research Pool between the Zealand Region and the Southern Region of Denmark.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.cca.2023.117414.

References


