Abstract

Objectives: Guidelines recommend the diagnosis of diabetes should be based on either plasma glucose or glycated hemoglobin (HbA1c) findings. However, lately studies have advocated glycated albumin (GA) as a useful alternative to HbA1c. We conducted a systematic review and meta-analysis to determine the overall diagnostic accuracy of GA for the diagnosis of diabetes.

Content: We searched for articles of GA diabetes diagnostic accuracy that were published up to August 2021. Studies were selected if reported an oral glucose tolerance test as a reference test, measured GA levels by enzymatic methods, and had data necessary for 2 × 2 contingency tables. A bivariate model was used to calculate the pooled estimates.

Summary: This meta-analysis included nine studies, totaling 10,007 individuals. Of those, 3,106 had diabetes. The studies showed substantial heterogeneity caused by a non-threshold effect and reported different GA optimal cut-offs for diagnosing diabetes. The pooled diagnostic odds ratio (DOR) was 15.93 and the area under the curve (AUC) was 0.844, indicating a good level of overall accuracy for the diagnosis of diabetes. The effect of the GA threshold on diagnostic accuracy was reported at 15.0% and 17.1%. The optimal cut-off for diagnosing diabetes with GA was estimated as 17.1% with a pooled sensitivity of 55.1% (95% CI 36.7%–72.2%) and specificity of 94.4% (95% CI 85.3%–97.9%).

Outlook: GA has good diabetes diagnostic accuracy. A GA threshold of 17.1% may be considered optimal for diagnosing diabetes in previously undiagnosed individuals.

Keywords: diabetes mellitus; diagnosis; diagnostic accuracy; glycated albumin; meta-analysis.

Introduction

Diabetes is a major health issue that has reached alarming levels: today, nearly half a billion people are living with diabetes worldwide [1]. The condition is chronic and requires continuous medical care with multifactorial risk-reduction strategies beyond glycemic control. The recommendations in the American Diabetes Association (ADA) Standards of Medical Care in Diabetes [2], include screening, diagnostic, and therapeutic actions that are known or believed to favorably affect the health outcomes of patients with diabetes. To date, there is no reference standard definition that captures the phenotypic complexity of diabetes and the risk of its complications. Currently, diabetes may be diagnosed based on either fasting plasma glucose (FPG), 2 h plasma glucose (2 h PG) after a 75 g oral glucose tolerance test (OGTT) or glycated hemoglobin (HbA1c). All tests are
equally appropriate and do not necessarily detect diabetes in the same individuals [3, 4].

OGTT is still a standard recommendation with great sensitivity for diabetes diagnosis. Its benefit is the 2 h PG cut point that diagnoses more people with diabetes compared with FPG and HbA1c cut points [3]. However, OGTT measurement lacks reproducibility, it is time-consuming, requires fasting and two blood samples [3, 4].

HbA1c, which is considered the reference for routine monitoring of patients with diabetes, is also a primary diagnostic tool for diabetes. HbA1c has several advantages compared with the FPG and OGTT, including greater convenience (fasting is not required), and greater preanalytical stability [3]. However, HbA1c is not suitable for conditions with altered erythrocyte turnover, such as hemoglobinopathies, chronic kidney disease and anemia [5]. Those conditions can interfere with the HbA1c measurement and adversely affect the interpretation of HbA1c results [5]. Furthermore, HbA1c ≥ 6.5% (48 mmol/mol) diagnoses only 30% of the diabetes cases identified collectively using HbA1c, FPG, and/or 2 h PG [6]. Therefore, it is important to consider alternative options in the diagnosis of diabetes.

Glycated albumin (GA), one of the validated tests as an alternative glycemic marker, is produced through the of glucose to albumin in a nonenzymatic reaction [7, 8]. Presently, GA can be measured by enzymatic assays in automated analyzers designed for high throughput. GA is hemoglobin/erythrocyte independent and reflects the average glucose concentration over the preceding 2–3 weeks, rather than 2–3 months observed for HbA1c [7, 8]. GA, with predictive values alike to HbA1c, it correlates with microvascular and macrovascular outcomes, and even death, especially in people with diabetes [8–12]. Additionally, studies have demonstrated the performance of GA in the diagnosis of diabetes when compared to the performance of HbA1c seems to be similar [13–22]. Therefore, in those studies GA has been proposed as a marker of glycemia that might complement or replace HbA1c under conditions wherein the latter does not reflect glycemic status accurately. However, regardless of the diabetes diagnostic reference standards or the GA thresholds, those studies have been published using varying levels of GA performance [13–22]. Consequently, the use of GA has not been completely endorsed in the diagnosis and screening of diabetes. Thus, to provide more precise summary estimates of clinical performance, we performed a systematic review and meta-analysis of studies that evaluated the performance of GA in the diagnosis of diabetes.

Materials and methods

The protocol of this systematic review was registered in the International Prospective Register of Systematic Reviews (PROSPERO) under the number CRD42021265628. In this systematic review and meta-analysis, we followed Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [23] and conducted the study according to the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement [24].

Search strategy and data sources

With assistance from our Institution’s library search specialist, we developed a searching strategy and searched the electronic databases PubMed (MEDLINE) without filter and complemented our search in EMBASE using database filters to remove MEDLINE results. Our search strategy looked for the combination of terms related to “glycated albumin” and “diabetes mellitus” in the title/abstract or across the record and in the medical subject heading (MeSH). This strategy was initially run against the databases in March 2020, and it was updated in August 2021. Details of all search terms are presented in Supplementary Material (Supplemental Table S1). Duplicate articles were removed from our initial search results, and the remaining articles were assessed for eligibility.

Selection of studies

Studies were selected and included in our final analysis when they met the following criteria: (a) studies that assessed the performance of GA when solely OGTT (reference standard 1) or OGTT and/or HbA1c (reference standard 2) were diabetes diagnostic reference standards; and (b) studies with enrolled individuals older than 18 years. Studies were excluded when: (a) individuals with known diabetes diagnosis or who were receiving anti-diabetic medication were included; (b) GA was measured by a non-enzymatic method; (c) case-control studies; (d) review articles; (e) comments, letters and/or editorials; (f) language other than English, Spanish or Portuguese.

Two review authors (F.C.C. and P.A.C.F.) independently screened titles/abstracts of all reports identified by the literature search and using eligibility criteria coded them as either “potentially include” or “exclude”. Based on the screening results, “potentially include” articles had their full-text assessed for eligibility, using an eligibility assessment form. We reported all excluded studies, with reasons for exclusion, in the PRISMA flow diagram. If multiple publications on a same cohort were found, the latest and most complete publication was considered. Differences in opinion were resolved through discussion or, if required, arbitration by a third review author (J.L.C).

Data extraction and management

Two review authors (F.C.C. and P.A.C.F.) independently extracted data, using a data extraction form, similar to a form previously used by Renz, et al. 2019 [25]. Any disagreements were resolved through discussion, or by consulting a third review author (J.L.C or A.L.P). The following information was extracted from each report: (a) study details
(author, publication year, country of origin); (b) study design; (c) sample size; (d) diabetes incidence; (e) participant characteristics [age, gender (male/female), GA, OGTT and HbA1c results]; (f) test methods (details of methodology and equipment description for GA, OGTT and HbA1c); and (g) performance of different cut-offs of GA (sensitivity and specificity, if possible, TP – true-positive cases; FP – false-positive cases; TN – true-negative cases; and FN – false-negative cases). We also attempted to contact authors for further information when data to construct a 2 × 2 contingency table was unclear or additional data were required. When data were not available from the authors, the study was excluded.

Quality assessment in included studies

Two review authors (F.C.C. and P.A.C.F.) independently evaluated the risk of bias and applicability of primary studies, using the Quality Assessment of Diagnostic Accuracy Studies tool QUADAS-2. QUADAS-2 consists of four key domains [(i) patient selection; (ii) index test; (iii) reference standard; (iv) flow and timing]], where each is assessed in terms of risk of bias and the first three in terms of concerns regarding applicability. The risk of bias and concerns about applicability were rated as “low,” “high,” or “unclear” [26]. Disagreements were resolved by consensus or by involving a third reviewer (J.L.C. or A.L.P).

Statistical analysis and data synthesis

The standard methods recommended for diagnostic accuracy meta-analysis studies were followed [27]. For each study, 2 × 2 contingency tables were constructed with data extracted for TP, TN, FP, and FN rates. Summary estimates of sensitivity, specificity, positive and negative likelihood ratios (LR+ and LR−), and diagnostic odds ratio (DOR) with their 95% confidence intervals (CI) were assessed using the bivariate model with random effects approach [28]. Summary receiver operating characteristic (SROC) curves were derived to calculate the area under the curve (AUC) and the Q index. An AUC close to 1 indicates that the diagnostic tests have high discrimination and are meaningful. A high Q index indicates high accuracy of the diagnostic tests. Hierarchical summary ROC curves (HSROC) were used to summarize the GA performance for specific cut-offs if 4 or more studies were presented data for the same or rounded cut-off. Fagan’s nomogram was used to present the post-test probabilities for diabetes and pooled sensitivity and specificity were used to present the clinical applicability of the test [29, 30]. A global diabetes prevalence of 9.3% was used as a pre-test probability for diabetes [1]. The heterogeneity among studies was evaluated by visual inspection of forest plots and SROC, Spearman’s correlation coefficient of sensitivity and specificity (p<0.05 indicated significant threshold effect), Cochran’s Q, Chi-square (χ²) (p<0.10 indicated significant heterogeneity), and the inconsistency index test (I²). The I² was defined as: below 30% considered non-important heterogeneity; 30%–60%, moderate heterogeneity; 60%–90%, substantial heterogeneity; above 90%, considerable heterogeneity. A high I² (>50%) and a low p value (<0.05) suggested the presence of heterogeneity caused by the threshold effect. The potential publication bias was assessed using Deeks’ funnel plot, where p<0.1 indicated statistical significance. Data analysis was performed using Meta-Disc, version 1.4 (Universidad Complutense, Madrid, Spain) and Stata software, Version 12.1 (Stata, College Station, TX, USA) by METANDI command. The forest plots were constructed using Review Manager Version 5.3 (Cochrane Collaboration, Oxford, UK). All studies selected for this review were previously approved by an Ethical Review Board and consequently ethical approval was not required for the present study.

Results

Selection of the studies

The initial search identified a total of 1,382 records (1,022 from PubMed and 360 from Embase). Of these, 1,358 records were excluded after screening the title/abstract and we fully assessed the remaining 24 records for the eligibility criteria. After full-text assess, 15 articles were excluded (three for different language, 1 used non-enzymatic method for measuring GA, 1 duplicate study population, 9 did not meet the research question or based on eligibility criteria, 1 had insufficient data for 2 × 2 contingency table) (Supplemental Table S2). The remaining 9 articles were eligible for data extraction, of which one article [18] reported two different diagnostic reference standards (OGTT solely and OGTT and/or HbA1c), which was included in the meta-analysis accordingly. Flow diagram is presented in Figure 1.

Characterization of the studies

The characteristics of each study included in the meta-analysis are shown in Table 1. The included studies were published between 2010 and 2021 and were predominantly performed in Asian countries (four from China; one, Japan; one, Korea; one, Taiwan; one, Brazil; and one, South Africa). Eight studies had cross-sectional design [13, 15–21] and one study was community-based cohort study [15]. The number of participants from included studies was 10,007, of those 3,106 (31.0%) were diagnosed with diabetes by the reference method of the individual studies. Six out of 9 included studies assessed the performance of GA in the diagnosis of diabetes by OGTT as the reference test, and had 5,933 participants. Of those, 1,422 (23.9%) were diagnosed with diabetes [13–18]. Four studies assessed GA performance using OGTT and/or HbA1c as reference standard and the number of enrolled individuals was 4,316. Of those, 1,770 (41.0%) were diagnosed with diabetes [18–21]. The included studies evaluated cut-offs of GA ranging from 13.0% to 17.5%. The Lucica GA-L assay (Asahi Kasei Pharma, Tokyo, Japan) was the most frequently used GA assay (n=7, 63.6%).
Quality assessment

The summary of our assessment of the quality of the studies included is reported in Table 2. Five studies had an overall low risk of bias and applicability concerns in all domains of the QUADAS-2 instrument.

One study [20] scored “high” risk of bias in the patient selection, because the flow of participants through the study excluded all individuals with first FPG <7.0 mmol/L and only those who presented a first FPG ≥7.0 mmol/L underwent OGTT and HbA1c with GA. Due to these inclusion criteria, this study scored “high” applicability concerns in the patient selection domain. It also scored “high” risk of bias in the index test domain, because this study used a predefined threshold for optimal cut-off value for GA, obtained by different diagnostic reference standards.

Another study [13] scored “high” risk of bias and applicability concerns in the patient selection because from 908 eligible individuals, 676 were excluded before performing OGTT and GA (633, as normoglycemic with FPG≤5.5 mmol/L, and 43, as newly diagnosed diabetes with FPG≥7.0 mmol/L and/or HbA1c≥6.5%). It also scored “high” risk of bias in the flow and timing domain, because in this study 29 participants who firstly were eliminated with FPG≥7.0 mmol/L criteria were added for receiver operating characteristic (ROC) analyses.

Finally, two studies [16, 21] were not clear in relation to which criterion was used in the patient selection.
### Table 1: Characteristics of selected studies.

<table>
<thead>
<tr>
<th>Author (reference)</th>
<th>Study location</th>
<th>Study design</th>
<th>Sample size, n</th>
<th>Age, years</th>
<th>GA, %</th>
<th>Reference standard</th>
<th>Incidence of diabetes, %</th>
<th>GA cut-off (S &amp; E)a</th>
<th>GA method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ma, et al. 2010</td>
<td>China</td>
<td>Cross-sectional</td>
<td>1971</td>
<td>53.1 ± 14.6</td>
<td>17.86 ± 4.5</td>
<td>OGTT</td>
<td>38.30</td>
<td>17.1% (76.82% &amp; 76.89%)</td>
<td>Enzymatic method (Lucica GA-L, Asahi Kasei Pharma, Tokyo, Japan)</td>
</tr>
<tr>
<td>Hwang, et al. 2014</td>
<td>Korea</td>
<td>Cross-sectional</td>
<td>852</td>
<td>52.5 ± 10.3</td>
<td>14.2 ± 5.6</td>
<td>OGTT + HbA1c</td>
<td>37.08</td>
<td>14.3% (66.4% &amp; 88.3%)</td>
<td>Enzymatic method (Lucica GA-L, Asahi Kasei Pharma, Tokyo, Japan)</td>
</tr>
<tr>
<td>Ikezaki, et al. 2015</td>
<td>Japan</td>
<td>Cross-sectional</td>
<td>176</td>
<td>Men 60 (53, 63)</td>
<td>Men 13.8</td>
<td>OGTT</td>
<td>16.5</td>
<td>15.2% (62.1% &amp; 61.9%); 16.5% (34.5% &amp; 87.1%)</td>
<td>Enzymatic method (Lucica GA-L, Asahi Kasei Pharma, Tokyo, Japan)</td>
</tr>
<tr>
<td>Wu, et al. 2016</td>
<td>Taiwan</td>
<td>Community-based cohort</td>
<td>1,559</td>
<td>50.4 ± 12.6</td>
<td>14.0 ± 2.6</td>
<td>OGTT</td>
<td>8.5</td>
<td>14.0% (83.33% &amp; 63.28%); 14.5% (78.03% &amp; 76.94%); 15.0% (74.0% &amp; 85.0%); 15.5% (68.94% &amp; 90.96%); 16.0% (62.12% &amp; 94.81%); 16.3% (56.06% &amp; 96.71%); 16.5% (55.3% &amp; 96.92%); 17.0% (67.73% &amp; 98.18%); 17.1% (66.21% &amp; 98.32%); 17.5% (41.67% &amp; 98.95%)</td>
<td>Enzymatic method (Lucica GA-L, Asahi Kasei Pharma, Tokyo, Japan) on Beckman Coulter AU2700 Chemistry analyzer (Beckman Coulter systems Co., Nyon, Switzerland)</td>
</tr>
<tr>
<td>He, et al. 2017</td>
<td>China</td>
<td>Cross-sectional</td>
<td>1,287</td>
<td>55 (47–62)</td>
<td></td>
<td>OGTT + HbA1c</td>
<td>77.08</td>
<td>17.1% (63.41% &amp; 95.93%)</td>
<td>Enzymatic method (Lucica GA-L, Asahi Kasei Pharma) on 7600 chemistry analyzer (Hitachi)</td>
</tr>
<tr>
<td>Su, et al. 2018</td>
<td>China</td>
<td>Cross-sectional</td>
<td>691</td>
<td>50.5 ± 13.3</td>
<td>16.2 ± 3.1</td>
<td>OGTT</td>
<td>48.5</td>
<td>16.3% (67.5% &amp; 83.4%)</td>
<td>Enzymatic method (Lucica GA-L, Asahi Kasei Pharma, Tokyo, Japan) on Hitachi 7600–120 (Hitachi, Tokyo, Japan)</td>
</tr>
<tr>
<td>Chume, et al. 2019</td>
<td>Brazil</td>
<td>Cross-sectional</td>
<td>242</td>
<td>53.4 ± 13.4</td>
<td>14.9 ± 2.2</td>
<td>OGTT</td>
<td>By OGTT 31.8</td>
<td>13.0% (93.5% &amp; 15.2%); 14.0% (84.4% &amp; 44.2%); 14.5% (70.1% &amp; 57.6%); 14.8% (64.9% &amp; 65.5%); 15.0% (62.3% &amp; 69.7%); 15.5% (48.1% &amp; 77.6%); 16.0% (42.9% &amp; 84.8%); 16.3% (42.9% &amp; 87.9%); 16.6% (36.4% &amp; 90.3%); 16.8% (31.2% &amp; 93.3%)</td>
<td>Enzymatic method (GlycoGap, Diazyme Laboratories, Poway, CA) in Cobas c702 (Roche diagnostics, Germany)</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Author (reference)</th>
<th>Study location</th>
<th>Study design</th>
<th>Sample size, n</th>
<th>Age, years</th>
<th>GA, %</th>
<th>Reference standard</th>
<th>Incidence of diabetes, %</th>
<th>GA cut-off (S &amp; E)</th>
<th>GA method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zemlin, et al. 2019</td>
<td>South Africa</td>
<td>Cross-sectional</td>
<td>1,294</td>
<td>47.8 ± 15.5</td>
<td>13.3 ± 2.7</td>
<td>OGTT</td>
<td>7.3%</td>
<td>17.0% (29.9% &amp; 93.9%); 17.1% (28.6% &amp; 93.9%); 17.5% (20.8% &amp; 96.4%); 14.7% (64.0% &amp; 64.1%); 16.6% (33.7% &amp; 90.4%)</td>
<td>Enzymatic method (quantILab Glycated albumin assay, Werfen™, Italy) in Roche cobas 6,000 analyzer (Roche diagnostics, Germany)</td>
</tr>
<tr>
<td>Li, et al. 2021</td>
<td>China</td>
<td>Cross-sectional</td>
<td>1935</td>
<td>NGT 28.11 ± 5.44 Pre-DM 37.15 ± 12.81 DM 47.63 ± 13.44</td>
<td>NGT 12.36 ± 0.81 Pre-DM 13.69 ± 1.45 DM 18.35 ± 5.00</td>
<td>OGTT + HbA1c</td>
<td>19.431%</td>
<td>15.15% (90.7% &amp; 78.9%)</td>
<td>Enzymatic method (Lucica GA-L, Asahi Kasei Pharma, Tokyo, Japan) on Cs400 B (Dirui Industrial Co., Ltd., Changchun, China)</td>
</tr>
</tbody>
</table>

Data are expressed as mean ± SD or median (interquartile range); GA designated optimal threshold in diagnosis of diabetes in bold; Data supplied by the author after contact; OGTT and/or HbA1c are reference; GA, glycated albumin; OGTT, oral glucose tolerance test; HbA1c, glycated hemoglobin; S & E, sensitivity and specificity; DM, diabetes mellitus; NGT, normal glucose tolerance.
Table 2: Quality assessment using QUADAS-2 criteria.

<table>
<thead>
<tr>
<th>Study</th>
<th>Risk of bias</th>
<th>Applicability concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient selection</td>
<td>Index text</td>
</tr>
<tr>
<td>1 Ma, et al. 2010</td>
<td>😊😊😊😊😊</td>
<td>😊😊😊😊😊</td>
</tr>
<tr>
<td>2 Hwang, et al. 2014</td>
<td>😊😊😊😊😊</td>
<td>😊😊😊😊😊</td>
</tr>
<tr>
<td>3 Ikezaki, et al. 2015</td>
<td>😊😊😊😊😊</td>
<td>😊😊😊😊😊</td>
</tr>
<tr>
<td>5 He, et al. 2017</td>
<td>😊😊😊😊😊</td>
<td>😊😊😊😊😊</td>
</tr>
<tr>
<td>6 Su, et al. 2018</td>
<td>😊😊😊😊😊</td>
<td>😊😊😊😊😊</td>
</tr>
<tr>
<td>7 Chume, et al. 2019</td>
<td>😊😊😊😊😊</td>
<td>😊😊😊😊😊</td>
</tr>
<tr>
<td>8 Zemlin et al. 2019</td>
<td>😊😊😊😊😊</td>
<td>😊😊😊😊😊</td>
</tr>
<tr>
<td>9 Li, et al. 2021</td>
<td>😊😊😊😊😊</td>
<td>😊😊😊😊😊</td>
</tr>
</tbody>
</table>

*: low; **: high; ?: unclear.

Meta-analysis

Overall diagnostic accuracy

For this analysis, we considered GA cut-offs designated as optimal for diagnosing diabetes by the authors of each article [13–21]. GA optimal threshold for the diagnosis of diabetes ranged from 14.3% to 17.1%. A total of 10,007 individuals were included in this analysis. Of those, 3,106 were diagnosed with diabetes by the reference method of the individual studies. The pooled DOR was 15.93 (Supplemental Table S3) and the AUC was 0.844 (Supplemental Table S3). The summary estimate of sensitivity was 0.69 (95% CI: 0.68–0.71) and specificity was 0.87 (95% CI: 0.86–0.87). There was considerable heterogeneity between studies in terms of sensitivity (Chi-square: 201.77; p<0.0001) and specificity (Chi-square: 552.14; p<0.0001) (Supplemental Table S3). We assumed there was no threshold effect among included studies, since SROC was not shoulder-shaped (Supplemental Figure S1) and Spearman correlation coefficient of sensitivity and specificity was ~0.5 (p=0.2). Besides, very low p value and a very high I2 of summary estimates indicated the heterogeneity due to non-threshold effect. The number of studies available has inadequate power to detect the impact of individual quality items as potential sources of heterogeneity. Therefore, we were unable to refine our investigation using meta-regression analyses. However, we performed subgroup analysis, according to the reference test.

Subgroup pooled diagnostic accuracy

As the diagnostic reference standard differed among the studies, we performed subgroup analysis, according to the reference test. The subgroup with OGTT and/or HbA1c as reference standard [18–21] had a considerably higher pooled DOR (18.51 vs. 11.91) and diagnostic accuracy (AUC 0.908 vs. 0.772) when compared to the subgroup with OGTT solely [13–18]. Pooled estimates of sensitivity, specificity, LR+ and LR− were similar. After re-running the meta-analysis by removing one study at a time, no article explained the persisting high heterogeneity for all summary estimates, regardless of reference standard, and we were unable to elucidate the reasons for this. Detailed accuracy estimates, SROC curves, and heterogeneity test results, according to the subgroup, are provided in the Supplemental Tables S4 and S5, and Supplemental Figures S2 and S3.

Effect of the GA threshold on diagnostic accuracy

The metandi command in Stata software requires a minimum of four studies to compute data [31]. For this reason, to perform this analysis, we used rounded GA cut-offs regardless of the reference test, as a result, only the cut-offs of 15.0% and of 17.1% each gathered at least 4 studies.
GA≥15.0% for the diagnosis of diabetes

Two studies assessed the performance of GA ≥15.0% to diagnose diabetes by OGTT [15, 18]. The study by Ikezaki, et al. evaluated GA ≥15.2% and Zemlin et al. GA ≥14.9% to diagnose diabetes by OGTT [13, 17]. All cut-offs were rounded to 15.0%, totaling 3,271 individuals. The HSROC curve is presented in Figure 2A. The AUC was 0.72 (Q*=0.659) (Supplemental Table S6). Forest plots of sensitivity and specificity of the four studies are shown in Figure 3A and the summary of diagnostic accuracy of GA ≥15.0% is presented in Supplemental Table S6. Sensitivity ranged from 62% to 74% and specificity from 62% to 94% (Figure 3A). The pooled sensitivity for these studies was 67.1% (95% CI 60.5%–73.0%, I²=32.1%) and the pooled specificity was 80.9% (95% CI 68.4%–90.6%, I²=98%) (Supplemental Table S6). The pooled LR+ was 3.51 (95% CI 1.74–7.05; I²=97.0%), LR− was 0.4 (95% CI 0.3–0.54; I²=72.7%) and DOR was 8.61 (95% CI 3.36–22.07; I²=92.69%). Due to the limited number of pooled studies to this meta-analysis, we were unable to perform sensitive analysis to explore the reasons for the considerable heterogeneity among the studies, despite the low p value and high I2 of specificity and DOR indicating heterogeneity due to non-threshold effect. However, the Deeks’ funnel plot revealed that there was no significant publication bias (p=0.19), Supplemental Figure S4A. Considering GA≥15.0% (with present pooled LR+ and LR−) as diabetes diagnostic criterion and inferring in global population with pre-test probability of 9.3% for diabetes [1], after a positive test (GA≥15.0%) the post-test probability for diabetes would increase to 26%, while a negative test (GA<15.0%) would decrease the post-test probability for diabetes to 4% (Figure 4A).

Apart from the above cited studies [13, 15, 17, 18], we also performed a meta-analysis to assess diabetes diagnostic accuracy of GA ≥15.0%, including one study that evaluated GA ≥15.15% using OGTT and/or HbA1c as reference standard [21]. Pooling these studies together, the total of individuals was 5,206 (Figure 2B), the combined sensitivity was 74% (95% CI 60%–84%; I²=93.5%) and combined specificity was 81% (95% CI 68%–89%; I²=98.1%). Those results are similar to the one found without the study by Li, et al. [21], but worsen the heterogeneity between studies. Furthermore, the Deeks’ funnel plot revealed that there was significant potential publication bias (p=0.04), Supplemental Figure S4B. Therefore, the results from the primary meta-analysis for this cut-off (GA ≥15.0%) were considered.

GA≥17.1% for the diagnosis of diabetes

Three studies reported the performance of GA ≥17.1% for diagnose diabetes by OGTT [14, 15, 18]. One study evaluated the threshold of GA ≥17.1% to diagnose diabetes using
Figure 3: Forest plots of estimates of sensitivity and specificity in each study.
(A) GA ≥ 15.0% to diagnose diabetes by OGTT; (B) GA ≥ 15.0% to diagnose diabetes regardless of the reference standard: OGTT solely or OGTT and/or HbA1c; (C) GA ≥ 17.1% to diagnose diabetes regardless of the reference standard: OGTT solely or OGTT and/or HbA1c. TP, true positive; FP, false positive; FN, false negative; TN, true negative. GA, glycated albumin; OGTT, oral glucose tolerance test; HbA1c, glycated hemoglobin.

Figure 4: Fagan’s nomogram for GA, showing post-test probabilities for diabetes.
(A) GA ≥ 15.0% and (B) GA ≥ 17.1%. GA, glycated albumin.
OGTT and/or HbA1c as reference standard [17]. All four studies totaled 5,059 individuals. The HSROC curve is shown in Figure 2C. The AUC was 0.85 (95% CI 0.82–0.88; \( Q^* = 0.7775 \)) (Supplemental Table S6). Forest plots of sensitivity and specificity of the four studies are shown in Figure 3C and the summary of diagnostic accuracy of GA \( \geq 17.1% \) is presented in Supplemental Table S6. Sensitivity ranged from 29% to 77% and specificity from 77% to 98% (Figure 3C). The pooled sensitivity was 55.1% (95% CI 36.7%–72.2%), \( I^2 = 97.3% \) and specificity was 94.4% (95% CI 85.3%–97.9%, \( I^2 = 99.2% \)) (Supplemental Table S6). The pooled LR+ was 9.78 (95% CI 4.29–22.34; \( I^2 = 97.5% \)), LR− was 0.47 (95% CI 0.33–0.69; \( I^2 = 97.3% \)) and DOR was 20.56 (95% CI 9.01–46.94; \( I^2 = 93.1% \)). Again, we were unable to perform sensitive analysis to explore the reasons for the considerable heterogeneity between studies in pooled indexes. The Deeks’ funnel plot showed no significant publication bias (\( p = 0.76 \)), Supplemental Figure S4C. Applying the Fagan’s nomogram with pre-test probability of 9.3% for diabetes [1], the post-test probability for diabetes would increase to 50% after a positive test (GA \( \geq 17.1% \)), while a negative test (GA \( < 17.1% \)) would decrease the post-test probability for diabetes to 5% (Figure 4B).

**Discussion**

**Summary of main results**

Our results showed that when examining GA at designated as optimal cut-offs (by the authors of each primary study) for the diagnosis of diabetes, pooled sensitivity and specificity were 0.69 and 0.87, respectively. The diagnostic test exhibited high discrimination (AUC=0.8442 with a \( Q^* \) value of 0.7757) and good determination effect (DOR=15.93). When we splitted the studies into subgroups according to reference standard, the subgroup with OGTT and/or HbA1c had a considerably higher pooled DOR and AUC than the subgroup with OGTT solely. Pooled sensitivity, specificity, LR+ and LR− of the two subgroups were similar to each other and were almost equal to the overall pooled estimates of primary analysis.

We presented the effect of the GA threshold at rounded values of 15.0% and 17.1%. For a rounded cut-off of 15.0%, the pooled sensitivity and specificity was 0.671 and 0.809, respectively. This accuracy implies 0.329 of false-negative and 0.191 of false-positive. The AUC and DOR suggested good determination effect and acceptable diagnostic accuracy. The pooled LR+ and LR− indicated that the pre-test to post-test probabilities would generate a minimal change, though significant. In comparison to the cut-off of 15.0%, the threshold of 17.1% showed lower pooled sensitivity (0.551) and false-positive (0.056), but greater pooled specificity (0.944) and false-negative (0.449). The AUC was 0.85 and DOR was 20.56, suggesting good determination effect and great diagnostic accuracy. The pooled LR+ indicated that the post-test probability for diabetes would moderate increase after a positive test, while the pooled LR− indicated that the pre-test to post-test probabilities, though significant, would generate a minimal change after a negative test.

**Our results compared with other reports**

As far as we know, this is the first systematic review with meta-analyses to evaluate the accuracy of the GA at the cut-offs of 15.0% and of 17.1% in the diagnosis of diabetes. In a recent systematic review and meta-analyses [32] that aimed to summarize the available data on GA measurements for the diagnosis of diabetes authors reported the accuracy of the GA at the cut-offs of 14.0%. The summary estimate of sensitivity was 0.766, specificity was 0.687, an AUC of 0.80 and DOR of 7.176. However, meta-analyzed data included sample from select populations, such as kidney transplant recipients [33] and youths 10 to 18 years-old with cystic fibrosis [34].

The results of GA diagnostic accuracy in our meta-analysis showed similar results to another meta-analysis conducted to evaluate the accuracy of the HbA1c in the diagnosis of diabetes, where both GA and HbA1c at optimal thresholds presented higher values of pooled specificity than sensitivity [35–37]. Summary estimates of GA\( \geq 17.1% \) compared with other reports summary estimates of HbA1c\( \geq 6.5% \) for the diagnosis of diabetes are presented in Supplemental Table S7. Our findings in terms of pooled sensitivity (0.551) for the GA\( \geq 17.1% \) are slightly higher than those reported elsewhere in meta-analysis that assessed the diagnostic value of HbA1c\( \geq 6.5% \) for diabetes by Xu et al. (0.518) [35] and Kaur et al. (0.50) [36], it was even higher when compared with pooled sensitivity reported by NCD-RisC group (0.305) [37]. So, GA\( \geq 17.1% \) had lesser false-negative cases compared to HbA1c\( \geq 6.5% \). However, our pooled sensitivity is lower than the one reported by Hoyer et al. (0.551 vs. 0.684) [38]. On the contrary, our finding of pooled specificity for GA \( \geq 17.1% \) (0.944) is lower than that reported in HbA1c\( > 6.5% \) by Xu et al. (0.956) [35], Kaur et al., (0.973) [36], NCD-RisC group (0.997) [37] and by Hoyer et al. (0.959) [38]. Thus, GA\( \geq 17.1\% \) had higher false-positive cases than HbA1c\( \geq 6.5\% \). Further, GA 17.1% presented lower diagnostic accuracy (AUC=0.85 vs. 0.93), and determination effect (DOR=20.7 vs. 40.6) than those for HbA1c 6.5%
reported by Xu et al. [37]. The pooled LR+ was also lower (9.78) than that reported by Xu et al. (19.0) [35] and by Kaur et al. (18.32) [36], which indicates that HbA1c, of 6.5% presents greater post-test probability for diabetes than GA of 17.1% after a positive test result. The GA 17.1% had similar pooled LR– (0.47 vs. 0.48) as for HbA1c 6.5% estimated by Xu et al. [35], and slightly lower than those estimated by Kaur et al. (0.51) [36]. The pooled LR– indicates both GA and HbA1c would generate a minimal change of pre-test to post-test probabilities after a negative test result.

**Applicability of findings to the review question**

To make sense of the results of the meta-analysis and its applicability in clinical practice, we explored pooled sensitivity and specificity, and the post-test probabilities for diabetes applying the Fagan’s nomogram. A global diabetes prevalence of 9.3% was used as pre-test probability for diabetes [1] with pooled LR+ and LR– for GA cut-offs 15.0% and 17.1%. After a test, the post-test probability for diabetes would increase to 26% for GA≥15.0% and 50% for GA≥17.1%. The post-test probability would decrease to 4% for GA <15.0% and 5% for GA <17.1%. Using GA ≥15.0% to diagnose diabetes with the pooled sensitivity of 0.671 and specificity of 0.809, for every 1,000 individuals tested, 62 cases of diabetes would be detected, 31 cases would be missed, and there would be 173 false diabetes diagnoses. For GA≥17.1% as diabetes diagnostic criterion with the pooled sensitivity of 0.551 and specificity of 0.944, we estimate for every 1,000 individuals tested 51 cases of diabetes would be detected, 42 cases would be missed, and there would be 51 false diabetes diagnoses. Even though GA≥17.1% presents lower diagnostic accuracy with higher false-positive results than HbA1c≥6.5%, its higher sensitivity than HbA1c≥6.5% [35–37] may have important implications from both clinical and healthcare policy perspectives. The alarming increase in the prevalence of diabetes worldwide warrants tests with greater sensitivity without meaningful loss of specificity for the early identification of the disease [1]. Thus, based on our findings the GA thresholds of 17.1% for screening purposes may be considered, once an early preventive intervention for people at high risk and treatment for newly diagnosed can help in reducing the incidence of diabetes complications, including cardiovascular morbidity and mortality [39].

Fang et al. analyzed the data from a multiethnic community-based cohort (n=4,785), and suggested GA had excellent diagnostic accuracy, with the AUC ranging from 0.824 to 0.951. GA cut-offs of 16.5% and 17.8% were, respectively, equivalent to an FPG of 126 mg/dL (97th percentile) and HbA1c of 6.5% (98th percentile) and had low to moderate sensitivity (0.273–0.707) but high specificity (0.980–0.992) for detecting undiagnosed diabetes. However, the reference definitions adopted in this study were without OGTT [FPG (≥126 mg/dL), HbA1c (≥6.5%), either FPG or HbA1c increased, or both FPG and HbA1c] [40]. Another study by Araki et al. in Japanese people reported a very efficient strategy to improve the metabolic control status of a general population using GA measurement as a screening tool for diabetes [41]. In the study, traditional glycemic tests were dispensed, and GA values were used to define the glycemic status and clinical practice in approximately 3 million people [41]. Based on Araki et al. definition, our finding of optimal threshold of GA as 17.1% for the screening for diabetes in previously undiagnosed population lies within the range of prediabetes (16.5–18.3%) [41]; and that is close to the “optimal” cut-offs estimated by Fang et al. and by several included studies [14, 15, 18, 20, 40]. It is noteworthy to mention that the risk of all-cause and cardiovascular mortality starts in the prediabetes stage even before clinical diabetes sets in and may also lead to significant morbidities as well [12, 42].

This behavior is essentially explained by the fact that there is no reference standard definition that captures the phenotypic complexity of diabetes and the risk of its microvascular and macrovascular complications. Consequently, all tests are equally appropriate to diagnose diabetes, although OGTT normally ranks high with great sensitivity for diabetes diagnosis [3, 4, 43, 44].

Although the GA is also relatively easy to use (fasting not required and measurement stability) and presents higher sensitivity for diabetes than the HbA1c, when GA is used, traditional glycemic tests should ideally also be measured. Because the number of false-negative for GA persisted considerably high, therefore, using GA alone in health surveys might miss some previously undiagnosed people who would be considered as having diabetes using a glucose-based test and/or HbA1c, and under these circumstances, could benefit from lifestyle and treatment interventions. This does not diminish the importance of GA in the diagnosis of diabetes, because adding GA to traditional glycemic test instruments could improve the clinical pathway of individuals with diabetes and healthcare systems.

**Strengths and weaknesses of the review**

A major strength of this review is that we conducted an extensive and systematic literature search without filter,
which ensured we included all studies that met the inclusion criteria, and, in the case of missing data, we attempted to contact the authors to improve the data extraction. Three diagnostic test accuracy studies [40, 45–47] that assessed the performance of GA without OGTT in reference standard did not meet eligibility criteria but after re-running the meta-analysis including those studies, the summary estimates were not significantly different from the primary meta-analysis (results not shown).

This study presents certain limitations. Although the findings were generally similar in the studies included, the meta-analysis revealed that there was considerable heterogeneity among them. Even after excluding two studies [13, 20] that we judged to be at high risk of bias and applicability concerns, and omitting the other studies one at a time, the analysis persisted very similar to our initial results. The attempt of performing a subgroup analysis according to the diagnostic reference standard or using rounded or same diagnostic cut-off values of GA was also not able to decrease heterogeneity. Our analysis suggested that the presence of heterogeneity was caused by a non-threshold effect. The small number of studies available hampered other types of subgroup analyses and a full explanation for the significant amount of heterogeneity found among studies. The minimum number of studies required for regression analysis is ten, otherwise we would have inadequate power to detect the potential sources of heterogeneity [48–50].

It should also be noted that our meta-analysis results are based on test accuracy data reported by primary studies conducted in settings with a disease prevalence exceeding that in most national/local prevalence [1]. Another limitation found is that most studies included in the present meta-analysis were undertaken in Asian countries, most notably in China. This may limit the generalization of our findings and indicates a need for further evaluations of test performance in different ethnicities.

We could not assess all objectives planned for this review due to limitations in data availability, highlighting an information gap. Studies were not consistent in using the same thresholds for GA in the diagnosis of diabetes. As a result, we were unable to fully assess the effect of different GA thresholds on diagnostic accuracy. We created subgroups with a rounded cut-off value and/or neglected the diagnostic reference standard, which enabled us to evaluate the effect at two GA thresholds. Therefore, our findings should be interpreted with caution. And, not to perpetuate missing data, it is extremely important that future studies are designed and reported according to the Standard for Reporting Diagnostic Accuracy (STARD) statement [51]. We also suggest reporting data of sensitivity and specificity of multiple GA cut-off points (e.g., 14.0%; 14.5%; 15.0%; 15.5%; 16.0%; 16.5%; 16.6%; 16.8%; 17.0%; 17.1%; 17.2%; ...).

Conclusions

GA performance in the diagnosis of diabetes is similar to HbA1c. Both GA and HbA1c result in few false-positive diabetes cases, but high number of false-negatives diabetes cases. The GA threshold of 17.1% may be considered optimal for diagnosing diabetes in previously undiagnosed individuals and would be more sensitive than HbA1c≥6.5%, with no meaningful loss of specificity. Since the number of false-negatives for GA 17.1% persisted considerably high, a negative result should ideally go for further investigation through a different test for diagnosis confirmation. Thus, GA may be used more of an additional test than an alternative to traditional glycemic tests, including HbA1c. The use of GA in surveillance requires further consideration of how it predicts and helps prevent diabetes complications and it is beyond the scope of this review. Furthermore, careful consideration about standardization of GA assays would be necessary, as has been done for HbA1c, to yield highly consistent GA results and increase precision.

Acknowledgments: We are grateful to Ms. Viviane Carrion Castanho and Ms. Raquel da Rocha Schimitt Domingos (librarians) at the Biblioteca FAMED/HCPA for their assistance for searches. We thank all the authors of the primary studies who contributed to this work.

Research funding: This study was supported by the Research Incentive Fund (FIPE) of the Hospital de Clínicas de Porto Alegre (HCPA) (FIPE/ HCPA, GPPG 190320), and was also financed in part by the Programa de Excelência Académica da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES-PROEX) and by the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS). F.C.C. received scholarship from CAPES-PROEX. L.G.S. received an undergraduate scholarship from FAPERGS. The funding organizations played no role in study design, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Author contribution: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Informed consent: Not applicable. However, all included studies declared that they have obtained informed consent from all individuals included in the study.
Ethical approval: The local Institutional Review Board deemed the study exempt from review.

References


Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2022-0105).