Diagnostic accuracy of oral glucose tolerance tests, fasting plasma glucose and haemoglobin A1c for type 2 diabetes in women with polycystic ovary syndrome: A systematic review and meta-analysis

Yitayeh Belsti [a], Joanne Enticott [a], Rafiatu Azumah [b], Chau Thien Tay [a], Lisa Moran [a], Ronald C.W. Ma [c,d], Anju E. Joham [a,e], Joop Laven [f], Helena Teede [a,e,1], Aya Mousa [a,*,1]

[a] Monash Centre for Health Research and Implementation (MCHRI), Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia
[b] Robinson Research Institute, The University of Adelaide AHMS Building, North Terrace, Adelaide, South Australia, 5005, Australia
[c] Dept of Medicine and Therapeutics, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China
[d] Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, China
[e] Diabetes and Endocrine Units, Monash Health, Melbourne, Victoria, 3168, Australia
[f] Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynaecology, Erasmus University Medical Center, Rotterdam, Netherlands

ARTICLE INFO

ABSTRACT

Aims: To inform international guidelines, a systematic review and meta-analysis was conducted to assess the performance of diagnostic methods for type 2 diabetes in women with polycystic ovary syndrome (PCOS).
Methods: An updated systematic search was conducted on five databases from 2017 until October 2023 and combined with prior searches (from inception). Meta-analyses of diagnostic accuracy tests were conducted.
Results: Nine studies comprising 2628 women with PCOS were included. Against the oral glucose tolerance test, a haemoglobin A1C (HbA1c) $\geq$ 6.5% had a pooled sensitivity of 50.00% (95% confidence interval (CI): 35.53–64.47), specificity of 99.86% (95%CI: 99.49–99.98), and positive and negative predictive values of 92.59% (95%CI: 75.27–98.09) and 98.27% (95%CI: 97.73–98.68), respectively, with an accuracy of 98.17% (95%CI: 97.34–98.79). Fasting plasma glucose values $\geq$ 7.0 mmol/L had a pooled sensitivity of 58.14% (95%CI: 42.13–72.99), specificity of 92.59% (95%CI: 75.35–98.08), positive and negative predictive values of 92.59% (95%CI: 75.35–98.08) and 99.09% (95%CI: 98.71–99.36), respectively, and an accuracy of 99.00% (95%CI: 98.46–99.39) against the oral glucose tolerance test.
Conclusions: To our knowledge, this is the first systematic review assessing the performance of diagnostic methods for type 2 diabetes in women with PCOS. We demonstrate that using a cut-off for HbA1c of $\geq$6.5% in this population may result in misdiagnosis of half of the women with type 2 diabetes. Our results directly informed the recommendations of the 2023 International PCOS Guideline, suggesting that the oral glucose tolerance test is the optimal method for screening and diagnosing type 2 diabetes in women with PCOS and is superior to fasting plasma glucose and HbA1c.

1. Introduction

Polycystic ovary syndrome (PCOS) is a common endocrine disorder, with prevalence varying from 8 to 13% in different population groups [1–4]. Irregular menstrual cycles, hirsutism, polycystic ovaries, infertility, and psychological and metabolic features are common in PCOS, with varying genetic and environmental risk factors. PCOS is diagnosed using the 2018 International PCOS Guideline updated Rotterdam criteria, which require two of the following: oligo/anovulation, clinical/biochemical hyperandrogenism and/or polycystic ovaries by ultrasound [5].

Although there are variations by body mass index (BMI), approximately 75%–95% of women with PCOS have underlying insulin resistance, which is a key risk factor underpinning the development of impaired glucose tolerance and type 2 diabetes (T2D) [6,7]. The prevalence of T2D ranges from 1.5% to 12.4% among women with PCOS,

---

depending on age [8], BMI [9,10] and ethnic variation [11–13]. A recent ten-year retrospective study showed that the incidence of T2D was approximately 6.25 per 1000-person years in PCOS compared with 1.49 in non-PCOS populations [14]. Both PCOS itself and concomitant risk factors such as family history of T2D or gestational diabetes mellitus (GDM), age >40 years and/or obesity [15,16] contribute to T2D risk in PCOS. Timely screening and diagnosis among this population can therefore facilitate and maximise prevention and treatment efforts.

However, the optimal tool/method for screening and diagnosing T2D in PCOS remains controversial, and the reported performance of existing tests has not previously been synthesised in PCOS. The 2018 International Evidence-based PCOS Guideline [5] recommended assessing glycaemic status at the time of PCOS diagnosis and every 1–3 years thereafter, with the frequency informed by the presence of other diabetes risk factors. Here, the use of either oral glucose tolerance tests (OGTT) or levels of fasting plasma glucose (FPG) or glycated haemoglobin A1C (HbA1c) was recommended to assess glycaemic status [5]. As no eligible studies were identified in the 2018 guideline literature search, these recommendations were based on expert consensus, highlighting the need for updated evidence-based recommendations derived from newly published evidence. Further, an OGTT was recommended at pre-conception in women with PCOS seeking fertility treatment or planning pregnancy and in high-risk women with PCOS. High-risk groups include those with a BMI >25 kg/m^2 (or >23 kg/m^2 for Asian ethnic groups), hypertension, high-risk ethnicity, history of impaired fasting glucose, impaired glucose tolerance or GDM or a family history of T2D [5].

Whilst the OGTT is arguably the most accurate method for detecting T2D, it is also an inconvenient, time-consuming and expensive method, with variability in collection processes and diagnostic cut-offs [17–19]. The American Diabetes Association (ADA) suggests that either FPG or HbA1c could be used instead of the OGTT to screen for T2D [20]. However, FPG has been classified as an insufficient screening tool in women with PCOS [21]. Similarly, HbA1c has been recommended as a diagnostic marker for T2D due to its many advantages. These include its ability to indicate average blood glucose levels over the preceding two to three months, its relative simplicity requiring only a single blood sample, and its stability and resistance to alterations caused by fasting or postprandial states or biological variability. However, its measurement accuracy is affected by genetic variants, some diseases including anaemia, as well as recent blood transfusions or the use of some medications [22–25]. Agreement among experts on whether FPG, OGTT or HbA1c is the best method for diagnosing T2D among women with PCOS is yet to be reached [26,27]. To date, there have been no published systematic reviews assessing or comparing the performance of these diagnostic tests for T2D in women with PCOS.

Given the current uncertainties around the most optimal tool for T2D diagnosis in PCOS and the absence of prior evidence synthesis addressing this question, we aimed to conduct a systematic review and meta-analysis evaluating and comparing the diagnostic accuracy metrics of different tests for detecting T2D in individuals with PCOS. We hypothesised that the diagnostic accuracy of FPG and HbA1c will be comparable to the OGTT for detecting T2D in PCOS. This study was performed in conjunction with the 2023 International Evidence-based Guideline for the Assessment and Management of PCOS [28] and directly informed the evidence-based recommendations therein, guiding clinicians, researchers, and policy-makers in this field.

2. Methods

We conducted a systematic review and meta-analysis of all available diagnostic studies, comparing the diagnostic accuracy of HbA1c and FPG against the standard OGTT for T2D diagnosis in PCOS. The review is reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses of Diagnostic Test Accuracy studies (PRISMA-DTA) guidelines [29]. The protocol was developed by the expert evidence team in consultation with the guideline development group, using gold-standard methodology endorsed by the Australian National Health and Medical Research Council [30]. This followed the same methodological protocol as the 2018 guideline [31], which is publicly available in the online technical report (https://www.monash. edu/__data/assets/pdf_file/0020/1412282/PCOS-Guideline_Technical-report.pdf). Additional detailed methodology for this specific review, including *a priori* eligibility criteria, is publicly available in the 2023 guideline technical report published online (https://doi.org/10.26180/23625288.v1).

2.1. Search strategy

This is an update of a systematic search conducted in 2017 (and previously in 2010 as part of the PCOS Guideline literature reviews), which found no suitable articles based on prior guideline eligibility criteria.

To identify relevant studies for this review using updated eligibility criteria (described below), we followed a two-step process. First, search results from the prior searches covering literature from inception to 2017 were re-screened against current criteria. Second, we conducted a new updated search from 2017 until October 3, 2023 using five electronic databases: Medical Literature Analysis and Retrieval System Online (MEDLINE) (Ovid), Psychological Information (PsycINFO) (Ovid), Excerpta Medica Database (EMBASE) (Ovid), All Evidence-Based Medicine (EBM) (Ovid), and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). Using both the prior and current searches, we captured all literature from inception to October 2023, and these studies were assessed for eligibility against the criteria for this review.

Manual searching was also conducted, using reference lists of relevant articles and reviews to identify additional eligible studies. The full details of the search strategy are included in the Supplementary Material.

2.2. Eligibility criteria

The inclusion and exclusion criteria were based on the PICOS (population, intervention, comparison, outcome, and study type) framework, developed by international content experts within the guideline development group. Studies reporting any measurement relating to the diagnostic accuracy of tests for T2D, including FPG, OGTT (2-h glucose) and HbA1c among women with PCOS of any age, weight and ethnicity were included. All available retrospective, cross-sectional and cohort studies, as well as clinical trials, evidence-based guidelines, systematic reviews and health technology assessments were included, provided they reported relevant data for extraction and evaluation of the diagnostic performance of FPG and/or HbA1c against the OGTT. The reference test for all analyses was the OGTT, and the index test was either FPG or HbA1c.

Exclusion criteria were as follows: studies conducted in participants who did not have diagnosed PCOS or who had pre-existing T2D, as well as studies written in languages other than English. Additionally, narrative reviews, non-evidence-based guidelines, case series, editorials, letters and commentaries were excluded.

2.3. Screening

Studies were imported into EndNote 20 for deduplication. After deduplication, the remaining articles were imported into Covidence web-based software for screening. Two independent reviewers (YB and RA) performed screening by title and abstract, followed by full-text screening based on the above eligibility criteria determined *a priori*. Disagreements were resolved by discussion, with arbitration by the guideline evidence team (AM and CTT) where required.

2.4. Assessment of methodological quality

Methodological quality at the study level was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) appraisal tool [32]. Using QUADAS-2 [32], risk of bias was assessed across four categories with a total of 10 questions (patient selection [3 questions], index test [2 questions], reference standard [2 questions], flow and timing signalling questions [3 questions]); while applicability assessment was undertaken using three categories and three questions (whether the patients, index tests and condition [reference standard] matched the review question). The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach was employed in evidence interpretation, with consideration of factors such as implementability, acceptability, and cost [33].

2.5. Data extraction and management

Study details and diagnostic accuracy data were extracted using a structured template, which includes the following details: first author, year, country, population/setting, study design, a summary of findings, PCOS diagnosis method and criteria, T2D diagnosis method and criteria, sample size, reference standard used for OGTT, HbA1c and FPG, and true positive (TP), false positive (FP), true negative (TN), false negative (FN), sensitivity and specificity values. In the case of missing data, the

MedCalc diagnostic test evaluation calculator and Review Manager V.5.3 software were used to calculate diagnostic accuracy measurements where possible using the available data.

2.6. Data synthesis and statistical analysis

A two by two table was constructed for TP, FP, TN and FN for all studies where the sensitivity and specificity estimates were reported. These data were then entered into Review Manager V.5.3 software, and forest plots and summary receiver operating characteristic curves (sROC) were created by assuming the reference test was 100% specific and sensitive. Studies that utilised the diagnostic criteria for T2D as outlined by the World Health Organization (WHO), the ADA, or a combination of both were included in the analysis. The diagnostic accuracy of FPG with a cut-off point of ≥ 7.0 mmol/l (≥ 126 mg/dl) and HbA1c with a cut-off point of $\geq 6.5\%$ (≥ 48 mmol/mol) for diagnosis of T2D were compared against the current gold-standard OGTT cut-off point of ≥ 11.1 mmol/l (≥ 200 mg/dl) [34,35]. For all included studies and corresponding diagnostic tests cut-off points, we presented individual TP, FP, TN, FN, sensitivity, specificity, positive and negative predictive values, accuracy, positive likelihood ratio, negative likelihood ratio, and kappa (k). For all studies, paired forest plots were created to present study-level measures of sensitivity and specificity for HbA1c $\geq 6.5\%$ and FPG ≥ 7.0 mmol/l. We also presented the sROC for
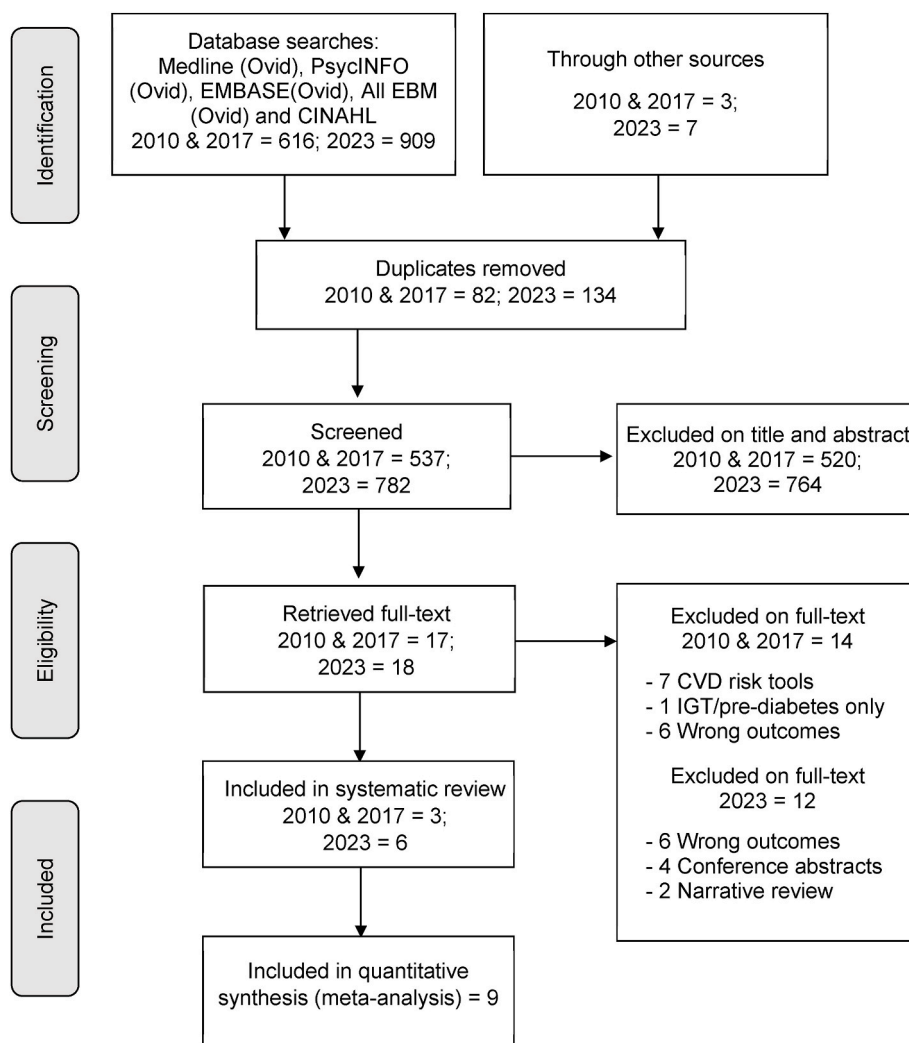

Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart describing the review process (updated database search was conducted on October 3, 2023). Abbreviations: CVD, cardiovascular disease; IGT, impaired glucose tolerance.

each analysis against the OGTT. Data synthesis was performed narratively to summarise studies or outcomes that could not be pooled in the meta-analysis.

3. Results

3.1. Characteristics of included studies

The systematic review process is presented in Fig. 1. Electronic and manual searches yielded 1319 unique articles, of which 1284 articles were excluded based on title and abstract screening, leaving 35 studies

to be assessed by full-text. Following full-text review, a further 26 articles were excluded (Fig. 1). Finally, nine studies reporting diagnostic accuracy tests of T2D in women with PCOS were eligible for inclusion in this review.

The included studies were conducted between 2011 and 2021 in eight countries; one each from Iraq [36], Austria [37], Denmark [27], Spain [38], USA [39], Turkey [26], and the Netherlands [40], and two from China [41,42] (Table 1). Three studies assessed the diagnostic accuracy of FPG alone against the OGTT [38,40,41], while three assessed HbA1c alone against the OGTT [26,27,42], and the remaining three studies assessed both HbA1c and FPG against the OGTT [36,37,

Table 1

Characteristics of studies included in a systematic review of glycaemic tests for type 2 diabetes in women with polycystic ovary syndrome.

Authors	Population/Setting	Study design	n	Age (years) mean ± SD or median (IQR)	BMI (kg/m^2) mean ± SD or median (IQR)	Methods/tools used	Outcomes	Summary of findings
Altemimi et al. [36]	Premenopausal women with PCOS in Faiha Specialized Diabetes, Endocrine, and Metabolism Center, University of Basrah, Iraq	Cross-sectional	129	26.30 ± 6.85	31.37 ± 7.69	2-h OGTT, HbA1c, FPG	Glycaemic disorders (IGT, prediabetes, T2D, FPG)	Screening of glycaemic disorders using 2-h OGTT is crucial for PCOS regardless of risk factors, and HbA1c seems to be an unsatisfactory screening tool to predict glycaemic disorders in women with PCOS
Lerchbaum et al. [37]	Women with PCOS in the Medical University of Graz, Austria	Cross-sectional	671	27 (23, 31)	24.2 (21.30, 30.10)	2-h OGTT, HbA1c, FPG	Glucose metabolism (prediabetes, T2D)	Findings do not support the recommendation that FPG or HbA1c can be used to screen prediabetes in women with PCOS. Instead, OGTT should be performed for screening of prediabetes
Li, 2015 et al. [41]	Women with PCOS, at the Family Planning Association of Hong Kong and the Department of Obstetrics and Gynaecology, Queen Mary Hospital, Hong Kong	Cross-sectional	467	30 (27, 33)	22.1 (19.9, 25.6)	OGTT, FPG	Dysglycaemia	A full OGTT should be recommended as the screening method for dysglycaemia in women with PCOS, regardless of BMI or family history of T2D
Magnussen et al. [27]	Premenopausal women with PCOS, at Odense University Hospital, Odense, Denmark	Retrospective	208	NR	NR	2-h OGTT, HbA1c	IGT, T2D	HbA1c is a relatively poor diagnostic marker in PCOS
Ortiz-Flores et al. [38]	Women with PCOS in Hospital Universitario Ramón y Cajal, Spain	Retrospective	400	26 (20, 30)	28.6 (22.90, 34.20)	OGTT, FPG	Dysglycaemia (IFG), T2D	An OGTT is the most accurate method for the diagnosis of disorders of glucose tolerance in women with PCOS in the clinical setting. FPG, on the contrary, is less accurate in predicting IGT and T2D in these women
Zhen et al. [42]	Women with PCOS in the Fifth Affiliated Hospital of Zhengzhou, China	Cross-sectional	161	23.68 ± 4.23	27.40 ± 2.20	OGTT, HbA1c	Prediabetes (IGT), T2D	FPG or HbA1c are not optimal indicators for screening abnormal glucose metabolism. However, their combination may reduce the misdiagnosis rate of glucose metabolic disorders to some extent. High-risk groups may still need to undertake OGTT to confirm diagnosis
Hurd et al. [39]	Women with PCOS in University Hospitals Case Medical Center reproductive endocrinology outpatient unit at Case Western Reserve University, USA	Prospective	111	28.00 ± 7.00	35.00 ± 8.00	2-h OGTT, HbA1c, FPG	Prediabetes, T2D	Women with PCOS should be screened for Prediabetes and T2D using OGTT or HbA1c
Celik et al. [26]	Women with PCOS at Gynaecological outpatient department of Namik Kemal University Hospital, Turkey	Case Control	252	24.80 ± 5.50	26.10 ± 5.70	2-h OGTT, HbA1c	Prediabetes, T2D	The only way to reliably detect abnormal glucose metabolism in Turkish women with PCOS appears to be using the OGTT
Veltman-Verhulst et al. [40]	Women with PCOS at a tertiary outpatient clinic for reproductive medicine, University Medical Centre Utrecht, the Netherlands	Cross-sectional	226	29.60 ± 4.30	27.30 ± 6.70	2-h OGTT, FPG	IFG, T2D	Compared with the OGTT, HbA1c has some potential weaknesses for IGT and T2D screening in women with PCOS

Abbreviations: BMI, body mass index; FPG, fasting plasma glucose; HbA1c, haemoglobin A1c; IFG, impaired fasting glycaemia; IGT, impaired glucose tolerance; NR, not reported; OGTT, oral glucose tolerance test; PCOS, polycystic ovary syndrome; SD, standard deviation; T2D, type 2 diabetes.

39]. Diagnostic test calculators were used to retrieve relevant unreported diagnostic accuracy metrics from the available information. The number of participants with PCOS ranged from 111 [39] to 671 [37]. All studies were conducted in University hospitals, where hospital services were combined with education and medical research (Table 1).

3.2. Methodological quality and publication bias

A summary of risk of bias and applicability concerns is presented in Fig. 2. In the risk of bias assessment, the patient selection category was deemed low risk of bias for only two of the nine studies, while the rest had high risk of bias [37,40]. The main issue underlying the high risk of bias pertained to the question of whether a consecutive or random sample of patients was enrolled, which was only satisfied by two studies [37,40]. The index test category was scored as high risk of bias for all studies. This is because most studies did not indicate whether index test results were interpreted without knowledge of the results of the reference standard (i.e. there was no explicit mention of blinding), despite all studies using prespecified standard cut-offs. For the reference standard category, the standard test used was the OGTT for all studies. Thus, for the question of whether the reference standard was likely to correctly classify the target condition, all studies satisfied this criterion. However, lack of blinding was again a factor influencing this category; hence, it was ultimately scored as unclear risk of bias across the included studies.

Finally, for the flow and timing category, the interval between index test(s) and the reference standard was appropriate in five cross-sectional studies, but not in the remaining four studies. Within all studies, all patients received the same reference standard. Similarly, all patients were included in the analysis within all studies. Overall, studies that satisfied all three criteria were deemed low risk of bias, while those not satisfying one of the criteria were high risk of bias, or unclear if the necessary information was not provided.

In the applicability assessment, there were no concerns regarding the included patients not matching the review question in all studies, except for the study by Celik et al. [26]. Across all studies, there were no concerns that the index test, its application or interpretation deviated from the review question or that the target condition as defined by the reference standard did not match the review question.

3.3. Quantitative synthesis and meta-analysis

Using OGTT performance as the reference standard, the forest plots in Fig. 3 show the sensitivity, and specificity of included studies in comparing OGTT vs. HbA1c and OGTT vs. FPG for diagnosing T2D. Six studies were included for each comparison. Although three studies included both HbA1c and FPG, combined analysis of HbA1c plus FPG against the OGTT was precluded due to the absence of information regarding whether testing was sequential or parallel, the order of testing, the number and definition of positive cases (requiring one or both tests to be positive) and the degree of overlap between tests.
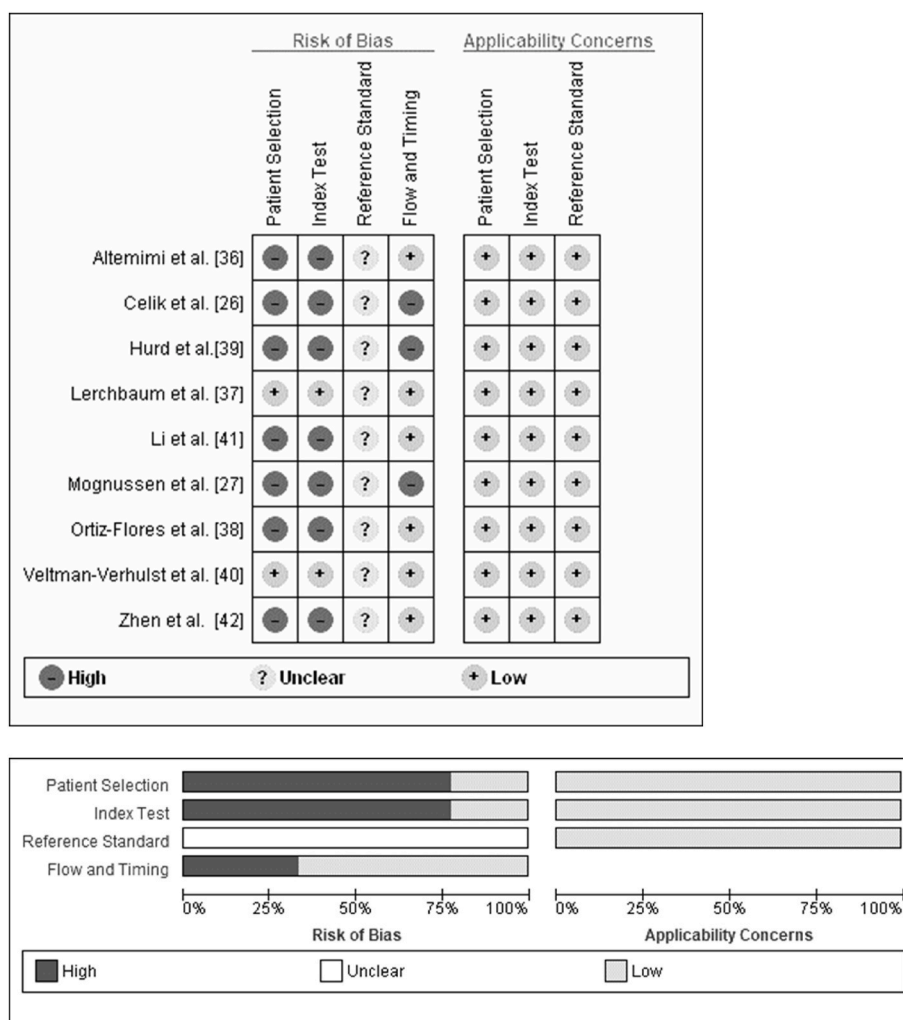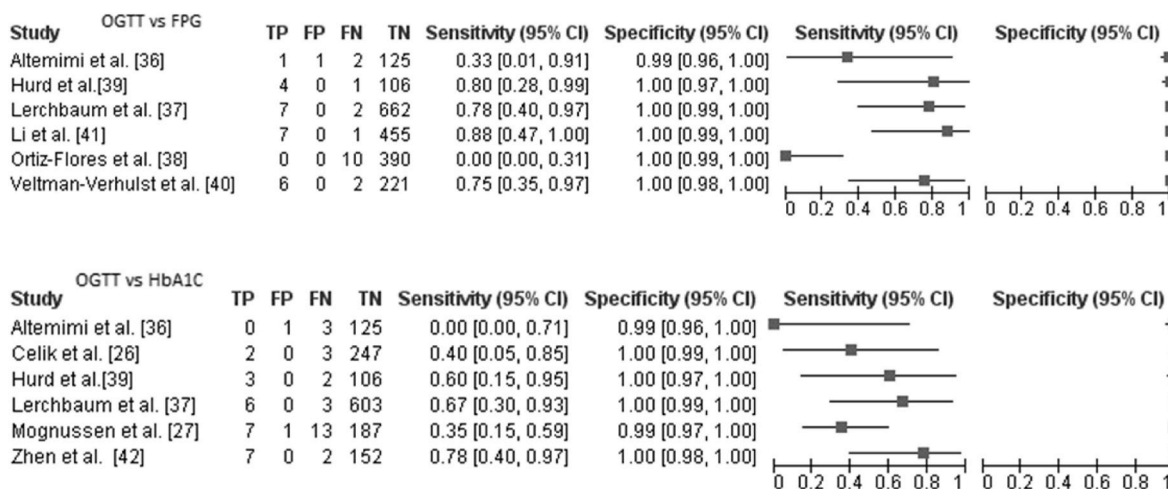

Fig. 2. Methodological quality summary table and graph using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2), illustrating the risk of bias assessment on the left and applicability concerns on the right.

Study	OGTT vs FPG	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Altemimi et al. [36]		1	1	2	125	0.33 [0.01, 0.91]	0.99 [0.96, 1.00]		
Hurd et al.[39]		4	0	1	106	0.80 [0.28, 0.99]	1.00 [0.97, 1.00]		
Lerchbaum et al. [37]		7	0	2	662	0.78 [0.40, 0.97]	1.00 [0.99, 1.00]		
Li et al. [41]		7	0	1	455	0.88 [0.47, 1.00]	1.00 [0.99, 1.00]		
Ortiz-Flores et al. [38]		0	0	10	390	0.00 [0.00, 0.31]	1.00 [0.99, 1.00]		
Veltman-Verhulst et al. [40]		6	0	2	221	0.75 [0.35, 0.97]	1.00 [0.98, 1.00]		

Study	OGTT vs HbA1C	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Altemimi et al. [36]		0	1	3	125	0.00 [0.00, 0.71]	0.99 [0.96, 1.00]		
Celik et al. [26]		2	0	3	247	0.40 [0.05, 0.85]	1.00 [0.99, 1.00]		
Hurd et al.[39]		3	0	2	106	0.60 [0.15, 0.95]	1.00 [0.97, 1.00]		
Lerchbaum et al. [37]		6	0	3	603	0.67 [0.30, 0.93]	1.00 [0.99, 1.00]		
Mognussen et al. [27]		7	1	13	187	0.35 [0.15, 0.59]	0.99 [0.97, 1.00]		
Zhen et al. [42]		7	0	2	152	0.78 [0.40, 0.97]	1.00 [0.98, 1.00]		

Fig. 3. Forest plot of sensitivity, specificity, and heterogeneity of diagnosis of type 2 diabetes using haemoglobin A1C versus oral glucose tolerance test. Abbreviations: FPG, fasting plasma glucose; HbA1c, haemoglobin A1c; OGTT, oral glucose tolerance test; TP, True positive, FP, false positive; FN, false negative; TN, true negative; CI, confidence interval.

Overall, sensitivity varied widely across studies, whereas there was less variation in specificity, with similar ranges for the latter, as shown in Table 2. The sROC plots for HbA1c and FPG are presented in Fig. 4.

3.3.1. Diagnostic accuracy of HbA1c against standard OGTT

Six studies comprising 1473 women with PCOS assessed the diagnostic accuracy of HbA1c against the OGTT with an average T2D prevalence of 3.33%. For diagnostic accuracy metrics, sensitivity ranged from 0 to 77.78%, specificity from 99.21% to 100%, and accuracy from 93.27% to 99.51%. Against the OGTT, a HbA1c $\geq$ 6.5% had a pooled sensitivity of 50.00% (95% confidence interval [CI] 35.53–64.47), specificity of 99.86% (95%CI: 99.49–99.98), positive and negative predictive values of 92.59% (95%CI: 75.27–98.09) and 98.27% (95%CI: 97.73–98.68), respectively, accuracy of 98.17% (95%CI: 97.34–98.79), kappa(k) of 0.64 (95%CI: 0.51–0.77), positive likelihood ratio of 355.50 (95%CI: 86.58–1459.64) and negative likelihood ratio of 0.50 (95%CI: 0.38–0.66) (Table 2).

3.3.2. Diagnostic accuracy metrics of FPG against standard OGTT

A total of six studies with 2007 women with PCOS assessed the diagnostic accuracy of FPG against the OGTT, with an average T2D prevalence of 2.34%. For diagnostic accuracy metrics of FPG against OGTT, sensitivity ranged from 0 to 87.50%, specificity from 99.21% to 100%, and accuracy from 97.50% to 99.78%. Against the OGTT, a FPG $\geq$7.0 mmol/l had a pooled sensitivity of 58.14% (95% CI: 42.13–72.99), specificity of 92.59% (95%CI: 75.35–98.08), positive and negative predictive values of 92.59% (95%CI: 75.35–98.08) and 99.09% (95%CI: 98.71–99.36), respectively, accuracy of 99.00% (95%CI: 98.46–99.39), kappa (k) of 0.71 (95%CI: 0.59–0.83) and positive and negative likelihood ratios of 570.06 (95%CI: 139.42–2330.85) and 0.42 (95%CI: 0.29–0.60), respectively (Table 2).

3.4. Qualitative findings

All retrieved articles in this systematic review recommend the use of OGTT for diagnosing T2D in women with PCOS [26,27,36–42], with none supporting the use of FPG or HbA1c for screening of T2D in this population. For instance, a recent study with premenopausal women with PCOS based in Iraq by Altemimi et al. 2021 [36] recommended that glycaemic disorders in PCOS, including T2D, be screened by 2-h OGTT, irrespective of risk factors such as increased BMI or family history of T2D. The remaining studies reported that HbA1c was an ineffective diagnostic marker [27,36,37] and that FPG was not sufficiently accurate in predicting T2D in women with PCOS [37,38].

3.5. The Grading of Recommendations assessment, development, and evaluation (GRADE) assessment

The GRADE principal domains of risk of bias, inconsistency, indirectness, and imprecision were adopted to assess the strength of the body of evidence for the diagnostic test comparisons included in this systematic review (HbA1c vs OGTT and FPG vs OGTT; Table 3) 43,44.

Risk of bias scoring was incorporated into GRADE from the QUADAS-2 assessment. Here, the evidence was downgraded once for serious risk of bias, due to most of the included studies having a high risk of bias, as outlined above (see section 3.2.). There were no serious inconsistencies, given that findings were largely in congruence, with no statistical heterogeneity or unexplained variability. Considering the similarity of the populations, intervention tests (index tests), comparison tests (alternative index tests), and outcomes in the body of evidence for the question at hand (diagnosis of T2D), there was no serious indirectness in the evidence. However, the evidence was downgraded once for imprecision due to the wide confidence intervals for sensitivity, despite narrow confidence intervals for the other test performance metrics. Overall, the GRADE certainty of evidence across all outcomes was deemed very low (Table 3).

4. Discussion

In women with PCOS, as with general populations, the optimal tool for screening and diagnosis of T2D remains contested. To our knowledge, this is the first systematic review and meta-analysis of diagnostic test accuracy for T2D in PCOS. By synthesising the available data from primary studies, we examined the pooled diagnostic performance and accuracy of index tests for T2D (HbA1c and FPG) against the reference gold-standard (OGTT) in women with PCOS. We found that the OGTT remains the most optimal screening method for detecting T2D in women with PCOS, compared with both FPG and HbA1c. Although FPG was marginally more sensitive than HbA1c for diagnosing T2D (58.1% vs 50.0%) at the standard cut-off point, our pooled estimate and qualitative evidence found that single use of either HbA1c or FPG performed poorly for diagnosing T2D compared to the OGTT in PCOS.

According to our findings, using FPG levels $\geq$7.0 mmol/l in women with PCOS had a pooled sensitivity of 58.14% for diagnosing T2D, suggesting that FPG is a relatively poor method for accurately identifying T2D in PCOS. This finding is supported by another study indicating that the odds of diagnosing T2D in PCOS using OGTT are three times higher than when relying solely on fasting values [45]. Similarly, it has been reported that FPG underestimates the prevalence of T2D in women

Table 2
Individual and pooled summary of diagnostic accuracy metrics.

Authors	Measurement method	N (sample size)	N (outcome: T2D)	Threshold cut-off	TP	FN	FP	TN	Sensitivity (95% CI)	Specificity (95% CI)	PPV (%)	NPV (%)	Accuracy (%)	kappa (k)
Altemimi et al. [36]	2-h OGTT	129	3	≥11.1 mmol/l										
	HbA1c	129	1	≥6.5%	0	3	1	125	0.00 (0.00,70.76)	99.21 (95.66,99.98)	0	97.62 (97.62,97.69)	96.90 (92.25,99.15)	−0.01 (−0.03, 0.01)
	FPG	129	2	≥7.0 mmol/l	1	2	1	125	33.33 (0.84,90.57)	99.21 (95.66,99.98)	50.00 (7.42,92.58)	98.43 (96.56,99.29)	97.67 (93.35,99.52)	0.39 (−0.16,0.94)
Lerchbaum et al. [37]	2-h OGTT	671	9	≥11.1 mmol/l										
	HbA1c	612	6	≥6.5%	6	3	0	603	66.67 (29.93,92.51)	100.00 (99.39,100.00)	100	99.50 (98.76,99.80)	99.51 (98.57,99.90)	0.80 (0.57,1.00)
	FPG	671	7	≥7.0 mmol/l	7	2	0	662	77.78 (39.99,97.19)	100.00 (99.44,100.00)	100	99.70 (98.98,99.91)	99.70 (98.93,99.96)	0.87 (0.70,1.00)
Li et al. [41]	OGTT	467	12	≥11.1 mmol/l										
	FPG	467	11	≥7 mmol/l	7	1	0	455	87.50 (47.35,99.68)	100.00 (99.19,100.00)	100	99.78 (98.64,99.96)	99.78 (98.80,99.99)	0.93 (0.80,1.00)
Mognussen et al. [27]	OGTT	208	20	>11.1 mmol/l										
	HbA1c	208	8	>6.5%	7	13	1	187	35.00 (15.39,59.22)	99.47 (97.07,99.99)	87.50 (47.55,98.18)	93.50 (91.25,95.20)	93.27 (88.96,96.27)	0.47 (0.24,0.70)
Ortiz-Flores et al. [38]	OGTT	400	10	≥11.1 mmol/l										
	FPG	400	0	≥7.0 mmol/l	0	10	0	390	0.00 (0.00,30.85)	100.00 (99.06,100.00)	–	97.50 (97.50,97.50)	97.50 (95.45,98.79)	0
Zhen et al. [42]	OGTT	161	9	≥11.1 mmol/l										
	HbA1c	161	7	≥6.5%.	7	2	0	152	77.78 (39.99,97.19)	100.00 (97.60,100.00)	100	98.70 (95.72,99.61)	98.76 (95.58,99.85)	0.87 (0.69,1.00)
Hurd et al. [39]	OGTT	111	5	≥11.1 mmol/l										
	HbA1c	111	3	≥6.5%	3	2	0	106	60.00 (14.66,94.73)	100.00 (96.58,100.00)	100	98.15 (94.77,99.36)	98.20 (93.64,99.78)	0.74 (0.40,1.00)
	FPG	111	4	≥7.0 mmol/	4	1	0	106	80.00 (28.36,99.49)	100.00 (96.58,100.00)	100	99.07 (94.84,99.84)	99.10 (95.08,99.98)	0.88 (0.66,1.00)
Celik et al. [26]	OGTT	252	5	≥11.1 mmol/l										
	HbA1c	252	2	≥6.5%	2	3	0	247	40.00 (5.27,85.34)	100.00 (98.52,100.00)	100	98.80 (97.58,99.41)	98.81 (96.56,99.75)	0.57 (0.13,1.00)
Veltman-Verhulst et al. [40]	OGTT	229	8	≥11.1 mmol/l										
	FPG	229	6	≥7.0 mmol/l	6	2	0	221	75.00 (34.91,96.81)	100 (98.34,100.00)	100	99.10 (97.08,99.73)	99.13 (96.88,99.89)	0.85 (0.65,1.00)
Overall (Pooled) FPG (6 studies)	OGTT	2007	47	≥11.1 mmol/l										
	FPG	2007	30	≥7.0 mmol/l	25	18	2	1959	58.14 (42.13,72.99)	99.90 (99.63,99.99)	92.59 (75.35,98.08)	99.09 (98.71,99.36)	99.00 (98.46,99.39)	0.71 (0.59,0.83)
HbA1c (6 studies)	OGTT	1532	51	≥11.1 mmol/l										
	HbA1c	1473	27	≥6.5%	25	26	2	1420	50.00 (35.53,64.47)	99.86 (99.49,99.98)	92.59 (75.27,98.09)	98.27 (97.73,98.68)	98.17 (97.34,98.79)	0.64 (0.51,0.77)

Abbreviations: FPG, fasting plasma glucose; HbA1c, haemoglobin A1c; IFG, impaired fasting glycaemia; IGT, impaired glucose tolerance; OGTT, oral glucose tolerance test; PCOS, polycystic ovary syndrome; T2D, type 2 diabetes; TP, True positive, FP, false positive; FN, false negative; TN, true negative; PPV, positive predictive value; NPV, negative predictive value.
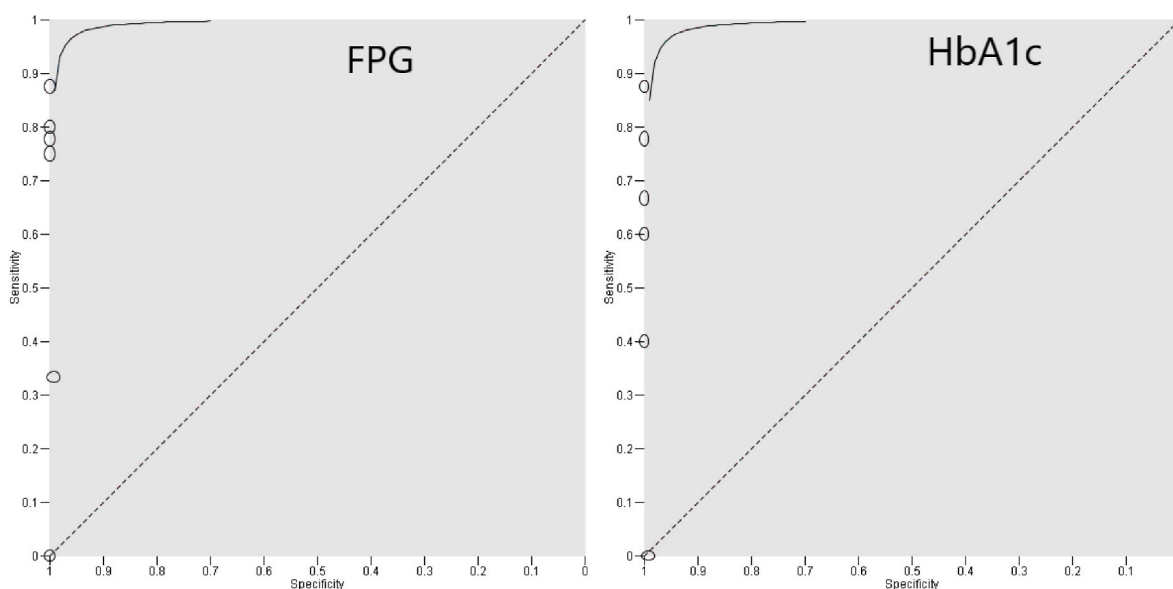
Fig. 4. Summary receiver operative characteristics (sROC) curves for haemoglobin A1C and fasting plasma glucose. Abbreviations: FPG, fasting plasma glucose; HbA1c, haemoglobin A1c.

Table 3
Evidence quality based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework.

Number of studies	Design	Risk of bias	Consistency	Directness	Imprecision	GRADE Certainty of the evidence
FPG against OGTT						
6	Cross-sectional and retrospective observational studies	Serious[a]	No serious inconsistency	No serious indirectness	Serious[b]	⊕◯◯◯ VERY LOW
HbA1c against OGTT						
6	Cross-sectional and retrospective observational studies	Serious[a]	No serious inconsistency	No serious indirectness	Serious[b]	⊕◯◯◯ VERY LOW

Abbreviations: FPG, fasting plasma glucose; HbA1c, haemoglobin A1c; OGTT, oral glucose tolerance test.
[a] Downgraded once for serious risk of bias due to most included studies having a high risk of bias.
[b] Downgraded once for serious imprecision due to wide confidence intervals for sensitivity.

with PCOS by >50% [24]. Hence, the use of FPG alone is not adequately discriminative for the diagnosis of T2D in women with PCOS.

Despite being inferior to the OGTT, our meta-analysis revealed that FPG ≥7.0 mmol/l is a more sensitive diagnostic tool for detecting T2D in women with PCOS, compared with a HbA1c ≥ 6.5%. This is consistent with the results of a previous meta-analysis conducted on the diagnostic accuracy tests of T2D in the general population, which concluded that FPG is more strongly correlated with current and future diabetes than HbA1c [46].

In comparing HbA1c with the OGTT, a pooled sensitivity of 50.0% was observed for HbA1c ≥ 6.5%, suggesting that this test would fail to detect T2D in 50% of individuals with the condition. Despite the ADA recommendation of HbA1c evaluation in low-risk patients [25], the clinical benefit of HbA1c for diagnosing T2D in PCOS is consistently reported to be relatively low, as highlighted here and elsewhere [26,27]. Nevertheless, HbA1c testing, requiring a single, non-fasting blood collection with less day-to-day variability might offer advantages that could improve adherence, particularly in high-risk or resource-constrained settings characterised by low health literacy or poor response rates.

Overall, both FPG and HbA1c currently appear to be ineffective screening or diagnostic tools for detecting T2D in PCOS. Mechanistically, it is likely that defective insulin secretion is present in PCOS [47], which may explain why postprandial glucose levels rather than fasting concentrations are more indicative of impaired glucose metabolism and T2D risk in this population. Indeed, several studies (as reviewed in Tomlinson et al. [21]) have demonstrated that women with PCOS have

normal FPG, but elevated postprandial glucose. Moreover, HbA1c has poor sensitivity for detecting T2D risk in women both with PCOS [26] and in general populations [48], as well as in women with previous GDM [49], suggesting that HbA1c likely reflects different facets of glucose metabolism and that the respective contributions of glucose alterations to HbA1c remain unclear [50].

It should be noted that high variability in the sensitivity of FPG and HbA1c was observed across the included studies. Beyond the role of chance, factors such as study design and bias, as well as the varied prevalence of the target condition, may cause variations in test accuracy measures. Although some literature challenges the impact of disease prevalence on the sensitivity and specificity of diagnostic tests [51], others suggest that prevalence differences due to clinical heterogeneity and/or artefactual differences, can contribute to variations in test accuracy [52–54]. Clinical heterogeneity refers to factors such as spectrum effects (differences in symptoms and severity), referral filters (differences in populations, settings or prior testing), or reader expectation (altered clinical thresholds due to prevalence expectations). For instance, populations with a high disease prevalence may include individuals with greater disease severity and, as a result, the test will perform better when applied to these populations [55,56]. Artefactual variability refers to prevalence differences which arise from study design and conduct, including via biased sampling methods and/or misclassification of the reference standard [52–54]. The range of T2D prevalence across the included studies (e.g. ranging from 1.9% to 10% prevalence by OGTT), whether caused by clinical heterogeneity, artefactual differences, or both, could therefore underpin the observed variability in

sensitivity. Similarly, variations in T2D risk factors including the mean/median BMI and age of study participants may also contribute to the wide sensitivity range observed and should be considered in the interpretation of these results.

As there are no previous systematic reviews examining the accuracy of T2D diagnostic tests among the PCOS population, our findings bear direct clinical relevance in that they provide much needed guidance for clinical decision-making in this context. Indeed, these findings have been used by the expert guideline development group, including consumers, to directly formulate clinical practice recommendations in the 2023 International PCOS Guideline [28]. Our results were considered alongside factors such as acceptability, cost, implementation and convenience of the available diagnostic tests, while recognising the inherently higher risk of T2D in PCOS, particularly during reproductive years when the risk of undiagnosed T2D in pregnancy is high. Based on our findings of the potential for misdiagnosis using HbA1c, and to a lesser extent using FPG, the current 2023 guideline now recommends that the OGTT is the most accurate and appropriate test for screening and diagnosis of T2D in PCOS, with the frequency of tests decided based on the presence of other risk factors [28]. These findings enable both researchers and clinicians to better understand the limitations around available testing methods, and to make informed, evidence-based decisions regarding their use in clinical practice.

4.1. Limitations

Some limitations should be noted. First, the review is limited by the number and quality of the existing literature. We incorporated nine studies from eight countries, of which almost all presented concerns regarding external validity. Nearly all the studies included were at either a high or unclear risk of bias in at least one domain, underscoring the importance of further, high-quality research in this field, including in large populations with diverse risk profiles. Subgroup analysis was not possible due to limited data and the lack of stratification by risk factors such as BMI and age. This is an important consideration for future studies in order to establish differences in diagnostic test accuracy among population subgroups. We also could not assess the combination of HbA1c and FPG against the OGTT due to missing data, and this is a key question warranting further study. Notwithstanding these limitations, this is the first published review to assess and compare T2D diagnostic tests in PCOS. We used a comprehensive search strategy with multiple databases and an internationally endorsed methodology that was developed in consultation with global leaders in the field and experts in evidence synthesis methods. Further, we used validated tools for the assessment of methodological quality, including the GRADE assessment tool, and our findings directly informed the current update of the International Evidence-based Guideline for the Assessment and Management of PCOS [28].

5. Conclusions

The OGTT is the most optimal screening method for detecting T2D in women with PCOS, compared with both FPG and HbA1c. If other tests are to be used, FPG is more sensitive than HbA1c for this purpose at the standard cut-off point. Concerns with the quality of the available literature are acknowledged and we highlight the need for further exploration of the potential utility of these tests, particularly in large cohorts with stratification by risk profiles to identify potential subgroups of interest. Collectively, based on current evidence and considerations of costs, implementation, convenience and the major implications of dysglycaemia, the International PCOS Guideline [28] recommends that the OGTT is the most accurate and appropriate test for T2D screening and diagnosis in PCOS, with frequency of tests based on the presence of other risk factors. Further studies to assess the differential diagnostic performance of T2D tests across different BMI and age groups, and to assess the combined accuracy of HbA1c and FPG against the OGTT are warranted.

Author contributions

YB conducted the database searches, title, abstract and full-text screening, quality appraisal, and wrote the first draft of the manuscript. RA performed title, abstract, and full-text screening and risk of bias assessments. JE and LM reviewed and edited the manuscript. CTT co-leads the evidence team for the PCOS guidelines and reviewed and edited the manuscript. RCWM served as the primary contact for this research question during guideline development, overseeing the evidence collation and synthesis, providing content expertise and supervision, and reviewing the manuscript. AJ and JL were the co-chair and chair, respectively, for the guideline development group, reviewed the evidence, and edited the manuscript. HT led the International PCOS guidelines, developed the PICO and guideline methods and reviewed and edited the manuscript. AM co-leads the evidence team for the PCOS guidelines and supervised the review process, contributed to determining the scope and direction of the research, provided methodological expertise, and reviewed and edited the manuscript. All authors approved the final version for publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Aya Mousa and Helena Teede are financially supported by fellowships from the Australian National Health and Medical Research Council (NHMRC). Joop Laven received grants from Ferring, Merck, Roche Diagnostics and AnshLabs and received consultancy honoraria from Ferring, Gedeon Richter and Roche Diagnostics. All other 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dsx.2024.102970.

References

[1] Barthelmess EK, Naz RK. Polycystic ovary syndrome: current status and future perspective. Front Biosci Elite Ed 2014 Jan 1;6(1):104–19. https://doi.org/10.2741/e695. PMID 24389146.

[2] Azziz R, Woods KS, Reyna R, Key TJ, Knochenhauer ES, Yildiz BO. The prevalence and features of the polycystic ovary syndrome in an unselected population. J Clin Endocrinol Metab 2004 Jun;89(6):2745–9. https://doi.org/10.1210/jc.2003-032046. PMID 15181052.

[3] Deswal R, Narwal V, Dang A, Pundir CS. The prevalence of polycystic ovary syndrome: a brief systematic review. J Hum Reprod Sci 2020;13(4):261–71. https://doi.org/10.4103/jhrs.JHRS_95_18. PMID 33627974.

[4] Conway G, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Franks S, Gambineri A, et al. The polycystic ovary syndrome: a position statement from the European Society of Endocrinology. Eur J Endocrinol 2014 Oct 1;171(4):P1–29. https://doi.org/10.1530/EJE-14-0253. PMID 24849517.

[5] Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Clin Endocrinol 2018; 89(3):251–68. https://doi.org/10.1111/cen.13795. PMID 30024653.

[6] Ovalle F, Azziz R. Insulin resistance, polycystic ovary syndrome, and type 2 diabetes mellitus. Fertil Steril 2002 Jun;77(6):1095–105. https://doi.org/10.1016/s0015-0282(02)03111-4. PMID 12057712.

[7] Stepto NK, Cassar S, Joham AE, Hutchison SK, Harrison CL, Goldstein RF, et al. Women with polycystic ovary syndrome have intrinsic insulin resistance on

euglycaemic-hyperinsulaemic clamp. Hum Reprod 2013 Mar;28(3):777–84. https://doi.org/10.1093/humrep/des463. PMID 23315061.

[8] Pelanis R, Mellembakken JR, Sundström-Poromaa I, Ravn P, Morin-Papunen L, Tapanainen JS, et al. The prevalence of Type 2 diabetes is not increased in normal-weight women with PCOS. Hum Reprod 2017 Nov 1;32(11):2279–86. https://doi.org/10.1093/humrep/dex294. PMID 29040530.

[9] Wekker V, van Dammen L, Koning A, Heida KY, Painter RC, Limpens J, et al. Long-term cardiometabolic disease risk in women with PCOS: a systematic review and meta-analysis. Hum Reprod Update 2020 Nov 1;26(6):942–60. https://doi.org/10.1093/humupd/dmaa029. PMID 32995872.

[10] Paparodis R, Evaggelos K, Bantouna D, Chourpiliadis C, Hara H, Livadas S, et al. Mon-547 post-surgically discovered differentiated thyroid microcarcinomas are more commonly found in patients eith chronic lymphocytic thyroiditis compared to those with multinodular goiter or graves' disease. J Endocr Soc 2019 Apr 15; (Suppl 1):3. https://doi.org/10.1210/js.2019-MON-547 (Supplement_1): MON.

[11] Ganie MA, Dhingra A, Nisar S, Sreenivas V, Shah ZA, Rashid A, et al. Oral glucose tolerance test significantly impacts the prevalence of abnormal glucose tolerance among Indian women with polycystic ovary syndrome: lessons from a large database of two tertiary care centers on the Indian subcontinent. Fertil Steril 2016 Jan 1;105(1):194–201.e1. https://doi.org/10.1016/j.fertnstert.2015.09.005. PMID 26407537.

[12] Seneviratne HR, Lankeshwara D, Wijeratne S, Somasunderam N, Athukorale D. Serum insulin patterns and the relationship between insulin sensitivity and glycaemic profile in women with polycystic ovary syndrome. BJOG 2009;116(13): 1722–8. https://doi.org/10.1111/j.1471-0528.2009.02360.x. PMID 19775306.

[13] Nanditha A, Ma RCW, Ramachandran A, Snehalatha C, Chan JCN, Chia KS, et al. Diabetes in Asia and the pacific: implications for the global epidemic. Diabetes Care 2016 Feb 12;39(3):472–85. https://doi.org/10.2337/dc15-1536. PMID 26908931.

[14] Liao WT, Huang JY, Lee MT, Yang YC, Wu CC. Higher risk of type 2 diabetes in young women with polycystic ovary syndrome: a 10-year retrospective cohort study. World J Diabetes 2022 Mar 15;13(3):240–50. https://doi.org/10.4239/wjd.v13.i3.240. PMID 35432752.

[15] Wild RA, Carmina E, Diamanti-Kandarakis E, Dokras A, Escobar-Morreale HF, Futterweit W, et al. Assessment of cardiovascular risk and prevention of cardiovascular disease in women with the polycystic ovary syndrome: a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome (AE-PCOS) Society. J Clin Endocrinol Metab 2010 May;95(5):2038–49. https://doi.org/10.1210/jc.2009-2724. PMID 20375205.

[16] Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. Fertil Steril 2004 Jan;81(1):19–25. https://doi.org/10.1016/j.fertnstert.2003.10.004. PMID 14711538.

[17] Andersen M, Glintborg D. Diagnosis and follow-up of type 2 diabetes in women with PCOS: a role for OGTT? Eur J Endocrinol 2018 Sep;179(3):D1–14. https://doi.org/10.1530/EJE-18-0237. PMID 29921567.

[18] Sicree RA, Zimmet PZ, Dunstan DW, Cameron AJ, Welborn TA, Shaw JE. Differences in height explain gender differences in the response to the oral glucose tolerance test- the AusDiab study. Diabet Med 2008 Mar;25(3):296–302. https://doi.org/10.1111/j.1464-5491.2007.02362.x. PMID 18307457.

[19] Sacks DB, Bruns DE, Goldstein DE, Maclaren NK, McDonald JM, Parrott M. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Clin Chem 2002 Mar;48(3):436–72. https://doi.org/10.1093/clinchem/48.3.436. PMID 11861436.

[20] Standards of medical care in diabetes. American Diabetes Association. Diabetes Care [cited Dec 29 2022]. Available from: https://diabetesjournals.org/care/article/36/Supplement_1/S11/27342/Standards-of-Medical-Care-in-Diabetes-2013; 2013.

[21] Tomlinson J, Millward A, Stenhouse E, Pinkney J. Type 2 diabetes and cardiovascular disease in polycystic ovary syndrome: what are the risks and can they be reduced? Diabet Med 2010 May;27(5):498–515. https://doi.org/10.1111/j.1464-5491.2010.02994.x. PMID 20536945.

[22] Little RR, Rohlfing CL, Tennill AL, Connolly S, Hanson S. Effects of sample storage conditions on glycated hemoglobin measurement: evaluation of five different high performance liquid chromatography methods. Diabetes Technol Therapeut 2007 Feb;9(1):36–42. https://doi.org/10.1089/dia.2006.0055. PMID 17316096.

[23] Gillett MJ. International Expert Committee report on the role of the A1c assay in the diagnosis of diabetes: diabetes Care. Clin Biochem Rev 2009;30(4):197–200. 32 (7):1327-34.

[24] Bry L, Chen PC, Sacks DB. Effects of hemoglobin variants and chemically modified derivatives on assays for glycohemoglobin. Clin Chem 2001 Feb;47(2):153–63. https://doi.org/10.1093/clinchem/47.2.153. PMID 11159762.

[25] American Diabetes Association. Standards of medical care in diabetes–2012 (Suppl 1):S11-63:S11-63 Diabetes Care 2012 Jan;35(Suppl 1). https://doi.org/10.2337/dc12-s011. PMID 22187469.

[26] Celik C, Abali R, Bastu E, Tasdemir N, Tasdemir UG, Gul A. Assessment of impaired glucose tolerance prevalence with hemoglobin A₁c and oral glucose tolerance test in 252 Turkish women with polycystic ovary syndrome: a prospective, controlled study. Hum Reprod 2013 Apr;28(4):1062–8. https://doi.org/10.1093/humrep/det002. PMID 23335611.

[27] Velling Magnussen L, Mumm H, Andersen M, Glintborg D. Hemoglobin A1c as a tool for the diagnosis of type 2 diabetes in 208 premenopausal women with polycystic ovary syndrome. Fertil Steril 2011 Nov;96(5):1275–80. https://doi.org/10.1016/j.fertnstert.2011.08.035. PMID 21982282.

[28] Teede HJ, Tay CT, Laven JJE, Dokras A, Moran LJ, Piltonen TT, et al. Recommendations from the 2023 international evidence-based guideline for the assessment and management of polycystic ovary syndrome. J Clin Endocrinol Metab 2023 Oct 1;108(10):2447–69. https://doi.org/10.1210/clinem/dgad463. PMID 37580314.

[29] McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, the Prisma-DTA Group, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA 2018 Jan 23; 319(4):388–96. https://doi.org/10.1001/jama.2017.19163. PMID 29362800.

[30] Guidelines. National Health and Medical Research Council [internet] [cited Apr 4 2023]. Available from: https://www.nhmrc.gov.au/guidelines.

[31] PCOS. Evidence-Based-Guidelines_20181009.pdf [internet] [cited Mar 8 2023]. Available from: https://www.monash.edu/__data/assets/pdf_file/0004/1412644/PCOS_Evidence-Based-Guidelines_20181009.pdf.

[32] Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009. PMID 22007046.

[33] Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. [GRADE guidelines]. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 2013 Jul;66(7):726–35. https://doi.org/10.1016/j.jclinepi.2013.02.003. PMID 23570745.

[34] American Diabetes Association. 6. Glycemic targets: *Standards of medical care in diabetes—2020*. Diabetes Care 2020 Jan 1;43(Supplement_1):S66–76.

[35] HEARTS D: diagnosis and management of type 2 diabetes [Internet] [cited Apr 19 2023]. Available from: https://www.who.int/publications-detail-redirect/who-ucn-ncd-20.1.

[36] Altemimi MT, Musa AK, Mansour AA. The Performance of glycated hemoglobin vs. oral glucose tolerance test in the Diagnosis of glycemic Disorders among Women with polycystic ovary syndrome in Southern Iraq. Indones Biomed J 2021 Jun 14; 13(2):178–85. https://doi.org/10.18585/inabj.v13i2.1431.

[37] Lerchbaum E, Schwetz V, Giuliani A, Obermayer-Pietsch B. Assessment of glucose metabolism in polycystic ovary syndrome: HbA1c or fasting glucose compared with the oral glucose tolerance test as a screening method. Hum Reprod 2013 Sep;28(9): 2537–44. https://doi.org/10.1093/humrep/det255. PMID 23756702.

[38] Ortiz-Flores AE, Luque-Ramírez M, Fernández-Durán E, Alvarez-Blasco F, Escobar-Morreale HF. Diagnosis of disorders of glucose tolerance in women with polycystic ovary syndrome (PCOS) at a tertiary care center: fasting plasma glucose or oral glucose tolerance test? Metabolism 2019 Apr;93:86–92. https://doi.org/10.1016/j.metabol.2019.01.015. PMID 30710572.

[39] Hurd WW, Abdel-Rahman MY, Ismail SA, Abdellah MA, Schmotzer CL, Sood A. Comparison of diabetes mellitus and insulin resistance screening methods for women with polycystic ovary syndrome. Fertil Steril 2011 Oct;96(4):1043–7. https://doi.org/10.1016/j.fertnstert.2011.07.002. PMID 21813121.

[40] Veltman-Verhulst SM, Goverde AJ, van Haeften TW, Fauser BCJM. Fasting glucose measurement as a potential first step screening for glucose metabolism abnormalities in women with anovulatory polycystic ovary syndrome. Hum Reprod 2013 Aug;28(8):2228–34. https://doi.org/10.1093/humrep/det226. PMID 23739218.

[41] Li HWR, Lam KSL, Tam S, Lee VCY, Yeung TWY, Cheung PT, et al. Screening for dysglycaemia by oral glucose tolerance test should be recommended in all women with polycystic ovary syndrome. Hum Reprod 2015 Sep;30(9):2178–83. https://doi.org/10.1093/humrep/dev166. PMID 26202923.

[42] Zhen Y, Yang P, Dong R, Wu Y, Sang Y, Du X, et al. Effect of HbA1c detection on the diagnostic screening for glucose metabolic disorders in polycystic ovary syndrome. Clin Exp Obstet Gynecol 2014;41(1):58–61. https://doi.org/10.12891/ceog16312014. PMID 24707685.

[43] Gopalakrishna G, Mustafa RA, Davenport C, Scholten RJPM, Hyde C, Brozek J, et al. Applying grading of recommendations assessment, development and evaluation (GRADE) to diagnostic tests was challenging but doable. J Clin Epidemiol 2014 Jul;67(7):760–8. https://doi.org/10.1016/j.jclinepi.2014.01.006. PMID 24725643.

[44] Yang B, Mustafa RA, Bossuyt PM, Brozek J, Hultcrantz M, Leeflang MMG, et al. GRADE guidance: 31. Assessing the certainty across a body of evidence for comparative test accuracy. J Clin Epidemiol 2021 Aug 1;136:146–56. https://doi.org/10.1016/j.jclinepi.2021.04.001. PMID 33864930.

[45] Legro RS. Diabetes prevalence and risk factors in polycystic ovary syndrome. Obstet Gynecol Clin N Am 2001 Mar 1;28(1):99–109. https://doi.org/10.1016/s0889-8545(05)70188-1. PMID 11293007.

[46] Kodama S, Horikawa C, Fujihara K, Hirasawa R, Yachi Y, Yoshizawa S, et al. Use of high-normal levels of haemoglobin A₁C and fasting plasma glucose for diabetes screening and for prediction: a meta-analysis: a₁C/FPG for Diabetes Screening and Prediction. Diabetes Metab Res Rev 2013 Nov;29(8):680–92. https://doi.org/10.1002/dmrr.2445. PMID 23963843.

[47] Holte J, Bergh TO, Berne CH, Berglund LA, Lithell HA. Enhanced early insulin response to glucose in relation to insulin resistance in women with polycystic ovary syndrome and normal glucose tolerance. J Clin Endocrinol Metab 1994 May 1;78 (5):1052–8.

[48] Cowie CC, Rust KF, Byrd-Holt DD, Gregg EW, Ford ES, Geiss LS, et al. Prevalence of diabetes and high risk for diabetes using A1C criteria in the U.S. population in 1988-2006. Diabetes Care 2010 Mar;33(3):562–8. https://doi.org/10.2337/dc09-1524. PMID 20067953.

[49] Picón MJ, Murri M, Muñoz A, Fernández-García JC, Gomez-Huelgas R, Tinahones FJ. Hemoglobin A1c versus oral glucose tolerance test in postpartum diabetes screening. Diabetes Care 2012 Aug;35(8):1648–53. https://doi.org/10.2337/dc11-2111. PMID 22688550.

[50] Legro RS, Castracane VD, Kauffman RP. Detecting insulin resistance in polycystic ovary syndrome: purposes and pitfalls. Obstet Gynecol Surv 2004 Feb;59(2): 141–54. https://doi.org/10.1097/01.OGX.0000109523.25076.E2. PMID 14752302.

[51] Wong HB, Lim GH. Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV. Proceedings of Singapore Healthcare. Proc Singap Healthc 2011 Dec 1;20 (4):316–8. https://doi.org/10.1177/201010581102000411.

[52] Leeflang MMG, Rutjes AWS, Reitsma JB, Hooft L, Bossuyt PMM. Variation of a test's sensitivity and specificity with disease prevalence. CMAJ Can Med Assoc J 2013 Aug 6;185(11):E537–44. https://doi.org/10.1503/cmaj.121286. PMID 23798453.

[53] Murad MH, Lin L, Chu H, Hasan B, Alsibai RA, Abbas AS, et al. The association of sensitivity and specificity with disease prevalence: analysis of 6909 studies of diagnostic test accuracy. CMAJ (Can Med Assoc J) 2023 Jul 17;195(27):E925–31. https://doi.org/10.1503/cmaj.221802. PMID 37460126.

[54] Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis – PubMed [internet] [cited Jan 29 2024]. Available from: https://pubmed.ncbi.nlm.nih.gov/18778913/.

[55] Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Ann Intern Med 2002 Oct 1;137(7):598–602. https://doi.org/10.7326/0003-4819-137-7-200210010-00011. PMID 12353947.

[56] Feinstein AR. Misguided efforts and future challenges for research on "diagnostic tests". J Epidemiol Community Health 2002;56(5):330–2. https://doi.org/10.1136/jech.56.5.330. PMID 11964422.