Accuracy of breath test for diabetes mellitus diagnosis: a systematic review and meta-analysis

Wenting Wang, Wenzhao Zhou, Sheng Wang, Jinyu Huang, Yanna Le, Shijiao Nie, Weijue Wang, Qing Guo

ABSTRACT
The review aimed to investigate the accuracy of breath tests in the diagnosis of diabetes mellitus, identify exhaled volatile organic compounds with the most evidence as potential biomarkers, and summarize prospects and challenges in diagnostic breath tests. Databases including Medline, PubMed, EMBASE, Cochrane Library and Science Citation Index Expanded were searched. Human studies describing diagnostic breath analysis with more than 10 subjects as controls and patients were included. Population demographics, breath test conditions, biomarkers, analytical techniques and diagnostic accuracy were extracted. Quality assessment was performed with the Standards for Reporting Diagnostic Accuracy and a modified QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2). Forty-four research with 2699 patients with diabetes were included for qualitative data analysis and 14 eligible studies were used for meta-analysis. Pooled analysis of type 2 diabetes breath test exhibited sensitivity of 91.8% (95% CI 83.6% to 96.1%), specificity of 92.1% (95% CI 88.4% to 94.7%) and area under the curve (AUC) of 0.96 (95% CI 0.94 to 0.97). Isotopic carbon dioxide (CO₂) showed the best diagnostic accuracy with pooled sensitivity of 0.949 (95% CI 0.870 to 0.981), specificity of 0.946 (95% CI 0.891 to 0.975) and AUC of 0.98 (95% CI 0.97 to 0.99). As the most widely reported biomarker, acetone showed moderate diagnostic accuracy with pooled sensitivity of 0.638 (95% CI 0.511 to 0.748), specificity of 0.801 (95% CI 0.691 to 0.878) and AUC of 0.79 (95% CI 0.75 to 0.82). Our results indicate that breath test is a promising approach with acceptable diagnostic accuracy for diabetes mellitus and isotopic CO₂ is the optimal breath biomarker. Even so, further validation and standardization in subject control, breath sampling and analysis are still required.

INTRODUCTION
Diabetes mellitus is a common metabolic disease with pathologically high blood glucose levels, causing damage to various organs and nerves. According to the latest International Diabetes Federation Diabetes Atlas, about 463 million people suffer from diabetes mellitus all over the world; however, more than half of them are undiagnosed and unaware of their status. Currently, the criteria for diagnosis of diabetes mellitus, including oral glucose tolerance test, fasting plasma glucose (FPG) test and glycosylated hemoglobin A1c (HbA1c) test, are all invasive blood-based assays, which limits screening for the disease.

In human exhaled breath, a wide variety of volatile organic compounds (VOCs) are observed and associated with health conditions. Analysis of breath VOCs provides a non-invasive approach to diagnosis of some diseases and monitoring of physiological effects and therapeutic efficacy, which enhanced the acceptability of patients for disease screening. To date, breath tests such as 13C urea test (for Helicobacter pylori), hydrogen-methane test (for gastrointestinal diseases), exhaled nitric oxide test (for asthma), Heartbreath test (for heart transplant rejection) and breath carbon monoxide test (for neonatal jaundice) have been applied in clinical practice. The association of diabetes mellitus and breath VOCs has been observed since the 1940s and is being reported by many studies up to now. Despite numerous studies, the applicability and diagnostic accuracy of breath tests for diabetes mellitus remain controversial.

In this work, we systematically reviewed studies that described the use of breath tests for diagnosis of diabetes mellitus, summarized exhaled characteristic VOCs in diabetes mellitus and assessed their diagnostic accuracy. The aim of this review is to identify breath VOCs with the most evidence as potential biomarkers in breath tests and investigate the accuracy of these VOCs in the diagnosis of diabetes mellitus.

METHODS
Search strategy and selection criteria
This systematic review was conducted in accordance with the guidelines of the Preferred Reporting Items for Systematic
Emerging technologies, pharmacology and therapeutics

Review and Meta-Analysis of Diagnostic Test Accuracy studies. The review was registered in PROSPERO (International Prospective Register of Systematic Reviews; registration ID: CRD42020222249).

The search was performed in the Medline, PubMed, Cochrane Library, EMBASE and Science Citation Index Expanded databases dated to November 23, 2020. Keywords including diabetes, diabetes mellitus, breath, exhaled, expired gas and expired air were used for searching. All search results including titles and abstracts were checked by two reviewers independently (WZ, WtW). The detailed search strategy is provided in online supplemental table S1.

Studies were included according to the following selection criteria: (1) human studies describing breath tests for diabetes mellitus diagnosis or discriminant; (2) all participating patients with diabetes mellitus were diagnosed by gold standard assay; and (3) at least two cohorts including patients and controls with more than 10 subjects in each group were studied and compared. Moreover, studies were included only if the full text was available in English language. We excluded studies meeting the following criteria: (1) review articles, conference abstracts, comments, case reports, viewpoints and editorials; (2) studies focusing on other diabetic complications (eg, diabetic nephropathy or ketoacidosis) rather than the diabetes mellitus itself; and (3) analyte was exhaled breath condensate.

Data extraction and quality assessment

Data extraction and assessment of research quality were performed by two reviewers (WZ, WtW) independently and judged by the third reviewer (JH) for controversial studies. Information including authors, country, year of publication, population characteristics (diabetes type, number of patients and controls, age, gender), analytical method, breath biomarkers and outcomes was collected. Different breath biomarkers in one study were all listed. Some data were calculated, converted or corrected based on the original data provided in the studies. The Standards for Reporting Diagnostic Accuracy (STARD) and a modified QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) appropriate for phase I studies on biomarker discovery were used to assess the quality of included studies and risk of bias. This modified QUADAS-2 referenced to Hanna et al’s study, and a signaling question was further modified to adapt to diabetes mellitus.

Statistical analysis

To estimate the diagnostic accuracy of the included breath tests, a random-effect meta-analysis was performed to generate pooled sensitivity, specificity, diagnostic OR (DOR), positive likelihood ratio (PLR), negative likelihood ratio (NLR) and area under the summary receiver operating characteristic (ROC) curve. Subgroup analysis by diabetes type and biomarkers was also conducted. Breath biomarkers that were reported at least over two times were analyzed, and pooled sensitivity, specificity and 95% CIs were assessed. The significance of heterogeneity was estimated using $I^2$ ($I^2 >50\%$ or $p<0.05$). Leave-one-out sensitivity analysis was also conducted to identify heterogeneity. Publication bias was assessed using Begg’s and Egger’s tests. Diagnostic threshold was estimated via Spearman’s correlation test. All data were analyzed in STATA V.12.0 software (Midas module).

RESULTS

Study selection and characteristics

As depicted in figure 1, a total of 6229 studies were identified from the search of databases and other sources and 2648 duplicates were first removed. Then, 1394 records were excluded by type of literature. After screening by title and abstract, 2015 studies were excluded. Afterwards, 172 full-text articles were further screened, and 44 studies were included for qualitative data analysis and 14 eligible studies were used for meta-analysis.

The characteristics of the included studies are presented in online supplemental table S2. A total of 2699 patients with diabetes mellitus, including 265 cases of type 1 diabetes (T1D), 1376 cases of type 2 diabetes (T2D) and pre-diabetes (PD), 49 cases of chemical diabetes (abolished), and 1009 cases of patients with indiscriminate diabetes, from 14 countries were studied. Overall, participants were in the 4–91 age range and the subjects in 14 eligible studies for meta-analysis were all adults. Detailed information on the population, including body mass index, FPG and HbA1c, is presented in online supplemental table S3. Among the 44 studies, 16 independent compounds were reported for diabetes mellitus diagnosis, and 4 of these biomarkers, namely acetone (n=19), isotopic carbon dioxide (CO2) (n=11), isopropanol (n=5) and dimethyl sulfide (n=2), were reported more than twice. Note that isotopic CO2 is an exogenous metabolite which derives from the ingested 13C-glucose or normal glucose. Majority of these breath VOCs were increased in exhaled breath, except 13CO2 and m-xylene which were decreased. Moreover, acetone was also decreased in the breath of patients with diabetes after dialysis. Analytical methods used for diabetic breath analysis within these studies involved spectroscopic, chromatographic, mass spectral and sensor-based methods. Sensor (n=14) was the most frequently used technique for diabetic breath tests, and gas chromatography-mass spectrometry (GC-MS) (n=7) was the most commonly used method, with qualitative ability to identify diabetic biomarkers in exhaled breath.

Quality assessment

The results of quality assessment using QUADAS-2 are presented in online supplemental figures S1 and S2. Detailed information on the modified QUADAS-2 is provided in online supplemental table S4. The STARD scores of each study are listed in online supplemental table S5; a mean value of 11.3 and SD of 5.0 were obtained.
**Data analysis**

To assess the overall diagnostic accuracy of breath test for diabetes mellitus, 14 studies were included in the meta-analysis and the highest diagnostic accuracy in each study was adopted. Pooled PLR of 11.531 (95% CI 7.165 to 18.558), NLR of 0.101 (95% CI 0.056 to 0.183), DOR of 114.333 (95% CI 42.083 to 310.626), sensitivity of 90.7% (95% CI 83.8% to 94.8%) and specificity of 92.1% (95% CI 87.9% to 95.0%) were obtained. Summary ROC analysis was performed with AUC of 0.97 (95% CI 0.95 to 0.98) (online supplemental figure S3). Confidence results showed the potential of breath tests in diabetes mellitus diagnosis. Nevertheless, substantial heterogeneity was also observed. Therefore, these studies were further grouped according to type of diabetes and breath biomarkers.

T2D was the most widely investigated in these studies; 10 studies including T2D and PD (the precursor of T2D) were analyzed. Yan’s study was also included since only 1 case of T1D was mixed with 86 T2D samples. The diagnostic accuracy for T2D was elevated and the specificity showed no significant heterogeneity; however, heterogeneity in sensitivity remained substantial. Leave-one-out sensitivity analysis was carried out and three studies (Li et al., Yatscoff et al. and Zhou et al.) which mainly contributed to heterogeneity were excluded. After removing these three studies, heterogeneity was remarkably reduced and no significant changes were observed in sensitivity and specificity before and after exclusion. As shown in figure 2, the diagnostic accuracy for T2D was 11.591 (95% CI 7.579 to 17.725) for PLR, 0.089 (95% CI 0.042 to 0.187) for NLR, 130.461 (95% CI 45.054 to 377.770) for DOR, 91.8% (95% CI 83.6% to 96.1%) for sensitivity, 92.1% (95% CI 88.4% to 94.7%) for specificity and 0.96 (95% CI 0.94 to 0.97) for AUC.

According to group of biomarkers, isotopic CO₂ including ¹³CO₂ and C₁⁸O₂ were the most frequently used in diabetic breath tests. Among the seven studies using isotopic CO₂ as biomarker, heterogeneity in sensitivity was observed. By leave-one-out analysis, Yatscoff et al.’s study was excluded and heterogeneity was reduced to an acceptable level (I² <50%, p>0.05) and no significant changes in diagnostic accuracy were observed. As shown in figure 3A, isotopic CO₂ had PLR of 17.716 (95% CI 3

---

**Figure 1** Flow diagram of study search and selection process. SCIE, Science Citation Index Expanded.

---

**6225 records identified from database search**  
PubMed(1950), Embase(2117), Ovid MEDLINE(832), Cochrane Library(361), SCIE(1859)

**Additional 4 records identified**  
Citations(2), Arxiv(1), Patent(1)

**2648 duplication excluded**

**3581 Records after duplicates removed**

**1394 Records excluded**  
Conference abstracts(1219), reviews(112), case reports(56), Editorial(4), comments (2), letter(1)

**2015 records excluded**  
not relevant topics(1881), Other diabetic related complications (80), animal experiments(40), exhaled breath condensate(14)

**172 Full-text articles assessed for eligibility**

**128 records excluded**  
no discriminant results(93), patients or controls less than ten subjects (27), actual participants and grouping unclear(2), questionable studies with self-contradictory results(4), full-text not available in English(1), redundant publications(1)

**44 studies included in qualitative synthesis**

**30 records excluded**  
No diagnostic accuracy reported(25), Same participants may be included (5)

**14 studies included in quantitative synthesis**
Figure 2  Forest plot of (A) sensitivity, (B) specificity and (C) SROC for the type 2 diabetes subgroup. Reference details provided in online supplemental file 1. AUC, area under the curve; SENS, sensitivity; SPEC, specificity; SROC, summary receiver operating characteristic curve.
Emerging technologies, pharmacology and therapeutics

8.376 to 37.473), NLR of 0.054 (95% CI 0.020 to 0.144), DOR of 328.275 (95% CI 77.049 to 1398.643), sensitivity of 0.949 (95% CI 0.870 to 0.981), specificity of 0.946 (95% CI 0.891 to 0.975) and AUC of 0.98 (95% CI 0.97 to 0.99). Acetone was also commonly used as diabetic breath biomarker. The included five studies exhibited heterogeneity in diagnostic sensitivity and specificity. Through the leave-one-out approach, heterogeneity was reduced after exclusion of Zhou et al’s study; however, sensitivity and specificity also changed obviously. In this study, patients and health volunteers were recruited from different provinces of China (patients with diabetes were from Jilin Province and volunteers were from Sichuan Province), with visible differences in location, climate, environment and diet, which have an impact on VOC levels in exhaled breath. These differences between patients and health volunteers may amplify the distinction of breath VOCs in the two groups, which is likely to increase discriminant accuracy. Thus, this study was excluded to obtain a more reliable result. Acetone had pooled PLR of 3.199 (95% CI 2.152 to 4.756), NLR of 0.453 (95% CI 0.341 to 0.601), DOR of 7.069 (95% CI 4.191 to 11.922), sensitivity of 0.638 (95% CI 0.511 to 0.748), specificity of 0.801 (95% CI 0.691 to 0.878) and AUC of 0.79 (95% CI 0.75 to 0.82), respectively (figure 3B).

Publication bias
The Begg’s and Egger’s tests were applied to assess publication bias as presented in online supplemental figure S4. No publication bias was suggested in the T2D group, isotopic CO₂ group and acetone group (Egger’s test p=0.931, p=0.300 and p=0.888).

Figure 3  Forest plot of sensitivity, specificity and SROC for (A) isotopic carbon dioxide and (B) acetone. Reference details provided in online supplemental file 1. AUC, area under the curve; SENS, sensitivity; SPEC, specificity; SROC, summary receiver operating characteristic curve.
DISCUSSION

According to the overall diagnostic accuracy results, breath test is a promising approach to non-invasive diagnosis of diabetes mellitus with prominent performance, although most studies showed risk of bias, which may overestimate diagnostic accuracy. By subgroup analysis, it was found that breath test is suitable for T2D diagnosis and isotopic CO₂ was the most discriminant breath biomarker. Although the T1D breath analysis has been investigated by a number of studies, only one study reported the diagnostic accuracy, which is not sufficient to draw a conclusion.

CO₂ is a common component in human exhaled breath, which is mainly from oxidation of glucose. The glucose is first converted to pyruvate via the glycolysis process and then oxidized by O₂ to generate CO₂. During this process, ATP is synthesized to produce essential energy. As a metabolic disease, diabetes mellitus may change the production of cellular energy and thereby alter the breath CO₂ sequentially. By ingestion of 13 C-glucose, 13 CO₂ was observed and presented reduced concentration in diabetic breath. By contrast, C¹⁸ O₂ was increased in diabetic breath with ingestion of normal glucose. This phenomenon is mainly due to the different source of isotope. The 13 C isotope in 13 CO₂ was directly from the metabolism of 13 C-glucose, but the 18 O isotope in C¹⁸ O₂ was converted from H₂¹⁸ O in the human body and catalyzed by the carbonic anhydrase. 16 For diabetes mellitus diagnosis, evaluation indexes such as the delta over baseline (ΔDOB), 14 insulin sensitivity index (ISI), 17 and changes in carbonic anhydrase activities (ΔCA) 18 were proposed to estimate isotopic CO₂ changes over time.

Acetone is an attractive breath biomarker which has been reported to correlate with various diseases. 19 This metabolite can be derived from decarboxylation of acetoacetate and dehydrogenation of isopropanol. 20 The relationship between diabetes mellitus and breath acetone has been investigated since the 1940s. 9 However, the evidence of its validity in diabetes mellitus diagnosis is still limited. Pooled analysis showed acetone to have moderate diagnostic accuracy when used as an independent biomarker and its sensitivity is especially limited. In addition, during literature screening, we found that a number of research claimed sensors to have potential in the diagnosis of diabetes mellitus by directly citing an acetone concentration of 1.8 parts per million (ppm) as the criterion. 21–25 However, their citations seemed unreliable to support this conclusion. Breath acetone concentrations reported in the included studies are listed in Table 1, and all units were converted to ppm. As can be seen, breath acetone is significantly influenced by various factors, such as gender, age, diabetes type, diet, exercise and drug treatment. Based on available studies, the threshold of breath acetone for diabetes mellitus diagnosis is still inconclusive. It is inappropriate to cite the 1.8 ppm acetone as criteria.

Breath acetone presented significantly higher concentration in both T1D (WMD=1.374, 95% CI 0.986 to 1.762, z=6.94, p=0.000) and T2D (WMD=0.845, 95% CI 0.605 to 1.085, z=6.91, p=0.000) groups, while there was no significant difference between T1D and T2D (p=0.162±0.05). Detailed information is depicted in online supplemental figure S5.

Isopropanol was used as potential biomarker in diabetic breath analysis in five studies and exhibited higher levels in the breath of patients with diabetes. This compound is mainly metabolized from propanoates in the human body. It is also a substrate of acetone synthesis from enzyme isopropanol dehydrogenase catalysis. 20 Recently, research showed that alcohol dehydrogenase in the liver is capable of reversely converting acetone to isopropanol in some abnormal conditions. 5 Based on available research, this compound showed moderate diagnostic accuracy with higher specificity than sensitivity. Unfortunately, only two studies provided the diagnostic accuracy of isopropanol. It does not make much sense to pool the data for further analysis.

Dimethyl sulfide was also reported twice. It is considered as a metabolite of microbial activities in the human body. 28 Variations of this compound in diabetic breath may be due to alteration in the gut microbiota induced by insulin deficiency. Hence, this component may not be suitable as a stable biomarker.

In terms of methods, sensor is the most frequently used technique for diabetic breath analysis. It is a fast, convenient and cost-effective approach for VOC detection. However, this technique is not capable of identifying the exhaled components to pick out characteristic biomarkers. Therefore, the principle of sensor array for diabetes mellitus diagnosis is difficult to interpret. GC-MS is another common analytical technique which enables qualitative and quantitative analyses of breath VOCs. This is a costly laboratory instrument with time-consuming analytical process. In breath analysis, spectroscopic
### Table 1  Breath acetone concentrations in the included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Physiological conditions</th>
<th>Acetone concentrations (mean±SD)/ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kistenev et al, 2020</td>
<td>Indiscriminate diabetes mellitus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overnight fasting</td>
<td>27.00±6.85</td>
</tr>
<tr>
<td></td>
<td>No control</td>
<td>8.89±0.83</td>
</tr>
<tr>
<td></td>
<td>Overnight fasting</td>
<td>8.77±13.49</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>9.30±2.21</td>
</tr>
<tr>
<td>Trefz et al, 2019</td>
<td>T1D</td>
<td>0.24 (0.20–0.26)</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.23 (0.19–0.31)</td>
</tr>
<tr>
<td>Yu et al, 2019</td>
<td>Indiscriminate diabetes mellitus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overnight fasting</td>
<td>515.97±272.28</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>225.93±171.09</td>
</tr>
<tr>
<td>Tyas et al, 2018</td>
<td>T2D</td>
<td>1.18±0.03</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.62±0.03</td>
</tr>
<tr>
<td>Li et al, 2018</td>
<td>T2D</td>
<td>2.20±2.378</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.78±0.35</td>
</tr>
<tr>
<td>Chien et al, 2017</td>
<td>T1D</td>
<td>1.64±0.85</td>
</tr>
<tr>
<td></td>
<td>T2D</td>
<td>1.12±0.60</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.75±0.43</td>
</tr>
<tr>
<td>Polanowska et al, 2017</td>
<td>T2D (+chronic kidney disease)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overnight fasting</td>
<td>5.963 (0.644–21.543)</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>3.154 (1.092–7.194)</td>
</tr>
<tr>
<td>Jiang, 2016</td>
<td>T1D</td>
<td>6.9±21.7</td>
</tr>
<tr>
<td></td>
<td>2 hours post breakfast</td>
<td>6.3±19.6</td>
</tr>
<tr>
<td></td>
<td>2 hours post lunch</td>
<td>4.0±10.3</td>
</tr>
<tr>
<td></td>
<td>2 hours post dinner</td>
<td>1.5±0.7</td>
</tr>
<tr>
<td></td>
<td>T2D</td>
<td>1.7±0.7</td>
</tr>
<tr>
<td></td>
<td>2 hours post breakfast</td>
<td>1.5±1.1</td>
</tr>
<tr>
<td></td>
<td>2 hours post lunch</td>
<td>1.4±1.0</td>
</tr>
<tr>
<td></td>
<td>2 hours post dinner</td>
<td>1.5±1.3</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>1.3±0.3</td>
</tr>
<tr>
<td></td>
<td>2 hours post breakfast</td>
<td>0.9±0.5</td>
</tr>
<tr>
<td></td>
<td>2 hours post lunch</td>
<td>1.0±0.6</td>
</tr>
<tr>
<td></td>
<td>2 hours post dinner</td>
<td>1.1±0.4</td>
</tr>
<tr>
<td>Sun et al, 2015</td>
<td>T2D</td>
<td>1.7±2.1</td>
</tr>
<tr>
<td></td>
<td>2 hours post breakfast</td>
<td>1.4±1.2</td>
</tr>
<tr>
<td></td>
<td>2 hours post lunch</td>
<td>1.3±1.1</td>
</tr>
<tr>
<td></td>
<td>2 hours post dinner</td>
<td>1.4±1.2</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>1.2±0.4</td>
</tr>
<tr>
<td></td>
<td>2 hours post breakfast</td>
<td>0.8±0.7</td>
</tr>
<tr>
<td></td>
<td>2 hours post lunch</td>
<td>0.8±0.5</td>
</tr>
<tr>
<td></td>
<td>2 hours post dinner</td>
<td>0.9±0.5</td>
</tr>
<tr>
<td>Rydosz, 2014</td>
<td>T1D</td>
<td>2.08±0.47</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.63±0.12</td>
</tr>
<tr>
<td>Zhou et al, 2014</td>
<td>T2D</td>
<td>2.167±1.155</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.488±0.166</td>
</tr>
</tbody>
</table>

Continued
Emerging technologies, pharmacology and therapeutics

Methods were usually applied for detection of specific compounds such as CO$_2$ and acetone. In addition, as an online mass spectra method, proton transfer reaction-mass spectrometry was also used for diabetic breath analysis. This method is quick and sensitive to VOC, with strong proton affinity; however, its applicability is relatively limited and inadequate in qualitative analysis.

Prospects and challenges

Despite promising results shown by diabetic breath tests, there are still some problems to be solved before these can be applied to clinical practice. First, most available studies were initial open trials, which are helpful in identifying diabetic biomarkers, but increased the risk of bias. More blind tests should be carried out in future work for further validation.

Second, standardization of diabetic breath analysis including subject control, breath sampling and detection is essential, which have significant impact on the test results. Since breath VOCs can be influenced by various factors, subject control is the first step before breath tests. Available studies demonstrated that diet control is a crucial condition, and overnight fasting was the most widely used. It is also a standard condition for FPG test. Potential biomarkers such as acetone and CO$_2$ were significantly affected by food intake.$^{29,30}$ Overnight fasting effectively reduced interference from food. Accordingly, diet control is a requisite for diabetic breath tests unless the adopted biomarkers can be proven to be not affected by food ingestion. Other control conditions such as restraining physical exercise and conducting gas exchange before the test have also been applied in some breath analyses.$^{13}$ The specific control conditions should consider the metabolic properties and influencing factors of the biomarker used. Breath sampling is the second important step which is also relevant to detection technique. Conventional analytical techniques such as GC-MS and gas chromatography-flame ionization detector are typical offline methods which require collection of breath in a container, such as a sampling bag or a sampling bottle. Breath collection for offline analysis should ensure enough breath is filled into the container. Moreover, effective duration of the sampled breath and possible contamination from the container should be considered. Some sensors, online mass spectrometry and spectroscopic methods

<table>
<thead>
<tr>
<th>Study</th>
<th>Physiological conditions</th>
<th>Acetone concentrations (mean±SD)/ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang, 2010</td>
<td>T1D Age ≤15</td>
<td>2.46±1.63 2.03±1.43</td>
</tr>
<tr>
<td></td>
<td>Age ≥16</td>
<td>1.75±1.07</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>2.30±1.39</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1.87±1.47</td>
</tr>
<tr>
<td></td>
<td>T2D</td>
<td>2.05±2.31</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.48±0.02</td>
</tr>
<tr>
<td>Deng et al, 2004</td>
<td>T2D</td>
<td>2.35±0.56</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.51±0.18</td>
</tr>
<tr>
<td>Tassopoulos et al, 1969</td>
<td>Indiscriminate diabetes mellitus</td>
<td>Overnight fasting No treatment Oral hypoglycemic agent Sulfonylureas Sulfonylureas+phenformin Insulin-treated Insulin-treated</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.40</td>
</tr>
<tr>
<td>Rooth, 1966</td>
<td>T1D</td>
<td>1.71±1.91</td>
</tr>
<tr>
<td></td>
<td>T2D</td>
<td>0.66±0.09</td>
</tr>
<tr>
<td></td>
<td>PD</td>
<td>0.31±0.03</td>
</tr>
<tr>
<td></td>
<td>CD</td>
<td>0.45±0.06</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.42±0.34</td>
</tr>
<tr>
<td>Stewart et al, 1964</td>
<td>T1D</td>
<td>Overnight fasting</td>
</tr>
<tr>
<td></td>
<td>T2D</td>
<td>1.74±5.25</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>0.42±0.19</td>
</tr>
</tbody>
</table>

Studies in bold font were used for pooled analysis.
Reference details provided in online supplemental file 1.
CD, chemical diabetes; ND, non-diabetes; PD, pre-diabetes; ppm, parts per million; T1D, type 1 diabetes; T2D, type 2 diabetes.
allowed direct analysis of exhaled breath in real time. For online breath sampling, the core is standardized
of breath exhalation. Breath exhalation should be carried out in a standard way to guarantee comparabil-
ity of test results. Thus, training of subjects seems necessary before an online breath test. For breath
detection, analytical methods ought to be further optimized and tested to obtain reliable stability and
reproducibility for clinical practice.

CONCLUSION
In this systematic review and meta-analysis, breath tests for diabetes mellitus diagnosis were investigated and the diagnostic accuracy of breath biomarkers was further estimated. The applicability of breath test for T2D diagnosis was demonstrated with great diagnostic accuracy. Besides, among the included biomarkers, isotopic CO₂ exhibited optimal sensitivity and specific-
ity, while the diagnostic accuracy of acetone and isopropanol was relatively moderate. Our results
suggest that breath test is a promising approach to non-invasive diagnosis of diabetes mellitus and is
especially appropriate for large-scale preliminary screening. Even so, before clinical practice, there is
still a lot of work to do with standardization of breath tests, including subject control, breath sampling and analyzing.

Author affiliations
1Affiliated Hangzhou First People’s Hospital Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
2Department of Biology and Chemistry, Zhejiang Institute of Metrology, Hangzhou, China
3Department of Medicine, Hangzhou Normal University, Hangzhou, Zhejiang, China
4Hangzhou Medical Association, Hangzhou, China
5School of Humanities and Management, Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China

Contributors WZ and WT were involved in the study design, data extraction, data analysis and drafting of the manuscript. JH and GQ were responsible for data
screening and results checking. SW, SN and WJW contributed to the statistics. All authors were involved in data interpretation and manuscript revision.

Funding This study was funded by Zhejiang Provincial Natural Science Foundation of China (no. LQ20B070001), National Natural Science Foundation of China (no. 71774147), Hangzhou Agricultural and Social Development Research Project (no. 202012038178) and Hangzhou Soft Science Project(no. 20200834M23).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplemental information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

REFERENCES
3 Kaur G, Lakshmi PVM, Rastogi A, et al. Diagnostic accuracy of tests for type 2 diabetes and prediabetes: a systematic review and meta-
Emerging technologies, pharmacology and therapeutics


<table>
<thead>
<tr>
<th>Database</th>
<th>Search strategy</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pubmed</strong></td>
<td><strong>#1</strong> (&quot;diabetes mellitus&quot;) [MeSH Terms] OR &quot;diabetes mellitus&quot; [All Fields] OR &quot;diabetes&quot; [All Fields]</td>
<td>742994</td>
</tr>
<tr>
<td></td>
<td><strong>#3</strong> (<strong>#1</strong> AND <strong>#2</strong>)</td>
<td>1427</td>
</tr>
<tr>
<td></td>
<td><strong>#4</strong> limit #3 to humans</td>
<td>1143</td>
</tr>
<tr>
<td></td>
<td><strong>#5</strong> limit #4 to English language</td>
<td>1056</td>
</tr>
<tr>
<td><strong>Embase</strong></td>
<td><strong>#1</strong> 'diabetes':ab,ti OR 'diabetes mellitus':ab,ti</td>
<td>817631</td>
</tr>
<tr>
<td></td>
<td><strong>#2</strong> 'breath':ab,ti OR 'exhalation':ab,ti OR 'exhaled':ab,ti OR 'expired air':ab,ti</td>
<td>89470</td>
</tr>
<tr>
<td></td>
<td><strong>#3</strong> <strong>#1</strong> AND <strong>#2</strong></td>
<td>2398</td>
</tr>
<tr>
<td></td>
<td><strong>#4</strong> limit #3 to humans</td>
<td>2200</td>
</tr>
<tr>
<td></td>
<td><strong>#5</strong> limit #4 to English language</td>
<td>2117</td>
</tr>
<tr>
<td><strong>Ovid MEDLINE</strong></td>
<td><strong>#1</strong> exp diabetes mellitus/</td>
<td>431737</td>
</tr>
<tr>
<td></td>
<td><strong>#2</strong> (diabetes or &quot;diabetes mellitus&quot;).ti,ab,</td>
<td>530686</td>
</tr>
<tr>
<td></td>
<td><strong>#3</strong> (<strong>#1</strong> or <strong>#2</strong>)</td>
<td>660386</td>
</tr>
<tr>
<td></td>
<td><strong>#4</strong> exp breath test/</td>
<td>15065</td>
</tr>
<tr>
<td></td>
<td><strong>#5</strong> (breath or exhaled or exhalation or (expired adj2 gas) or (expired adj2 air)).ti,ab,</td>
<td>55181</td>
</tr>
<tr>
<td></td>
<td><strong>#6</strong> (<strong>#4</strong> or <strong>#5</strong>)</td>
<td>58517</td>
</tr>
<tr>
<td></td>
<td><strong>#7</strong> (<strong>#3</strong> AND <strong>#6</strong>)</td>
<td>1132</td>
</tr>
<tr>
<td></td>
<td><strong>#8</strong> limit #7 to humans</td>
<td>920</td>
</tr>
<tr>
<td></td>
<td><strong>#9</strong> limit #8 to English language</td>
<td>832</td>
</tr>
<tr>
<td><strong>Cochrane Library</strong></td>
<td><strong>#1</strong> MeSH descriptor: [Diabetes Mellitus] explode all trees</td>
<td>31368</td>
</tr>
<tr>
<td></td>
<td><strong>#2</strong> (diabetes):ti,ab,kw OR (diabetes mellitus):ti,ab,kw</td>
<td>80884</td>
</tr>
<tr>
<td></td>
<td><strong>#3</strong> (<strong>#1</strong> OR <strong>#2</strong>)</td>
<td>83304</td>
</tr>
<tr>
<td></td>
<td><strong>#4</strong> MeSH descriptor: [Breath Tests] explode all trees</td>
<td>1535</td>
</tr>
<tr>
<td></td>
<td><strong>#5</strong> (breath):ti,ab,kw OR (exhaled):ti,ab,kw OR (exhalation):ti,ab,kw OR (expired gas):ti,ab,kw OR (expired air):ti,ab,kw</td>
<td>12706</td>
</tr>
<tr>
<td></td>
<td><strong>#6</strong> (<strong>#4</strong> OR <strong>#5</strong>)</td>
<td>12706</td>
</tr>
<tr>
<td></td>
<td><strong>#7</strong> (<strong>#3</strong> AND <strong>#6</strong>)</td>
<td>361</td>
</tr>
<tr>
<td><strong>Science Citation Index Expanded</strong></td>
<td><strong>#1</strong> TOPIC: (diabetes) OR TOPIC: (diabetes mellitus)</td>
<td>808,330</td>
</tr>
<tr>
<td></td>
<td><strong>#2</strong> TOPIC: (breath) OR TOPIC: (exhaled) OR TOPIC: (exhalation) OR TOPIC: (expired gas) OR TOPIC: (expired air)</td>
<td>91,049</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>#3</td>
<td>(#1 AND #2)</td>
<td>1918</td>
</tr>
<tr>
<td>#4</td>
<td>limit #3 to English language</td>
<td>1859</td>
</tr>
</tbody>
</table>
**Table S2.** Characteristics of included studies on diabetic breath analysis.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Country</th>
<th>Diabetes type</th>
<th>Number of patients/controls</th>
<th>Breath condition control</th>
<th>Analytical techniques</th>
<th>Breath biomarkers</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalidoss et al.</td>
<td>2020</td>
<td>India</td>
<td>Indiscriminate</td>
<td>30/30</td>
<td>Fasting/2h post prandial</td>
<td>Sensor</td>
<td>Breathprint (Acetone(+))</td>
<td>66.66%</td>
<td>70.00%</td>
<td>68.33%</td>
<td>-</td>
</tr>
<tr>
<td>Kistenev et al.</td>
<td>2020</td>
<td>Russia</td>
<td>Indiscriminate</td>
<td>11/11</td>
<td>Fasting</td>
<td>Terahertz spectroscopy</td>
<td>Acetone(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tazi et al.</td>
<td>2019</td>
<td>Indonesia</td>
<td>Indiscriminate</td>
<td>25/25</td>
<td>2h fasting</td>
<td>Sensor</td>
<td>Breathprint</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>K Kalidoss et al</td>
<td>2019</td>
<td>India</td>
<td>Indiscriminate</td>
<td>17/13</td>
<td>Fasting/2h post prandial</td>
<td>Sensor</td>
<td>Breathprint</td>
<td>70.05%</td>
<td>69.23%</td>
<td>69.64%</td>
<td>-</td>
</tr>
<tr>
<td>Trefz et al.</td>
<td>2019</td>
<td>Germany</td>
<td>T1D</td>
<td>53/60</td>
<td>-</td>
<td>PTR-MS</td>
<td>Ethanol(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Isopropanol(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dimethyl-sulfide(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Isoprene(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pentanal(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nevezhin et al.</td>
<td>2019</td>
<td>Russia</td>
<td>T2D</td>
<td>22/14</td>
<td>No control</td>
<td>laser optical-acoustic IR spectroscopy</td>
<td>Breathprint</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yu et al.</td>
<td>2019</td>
<td>China</td>
<td>Indiscriminate</td>
<td>16/16</td>
<td>Overnight fasting</td>
<td>GC-FID</td>
<td>Acetone(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Isopropanol(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Som et al.</td>
<td>2018</td>
<td>India</td>
<td>T2D(+HP)</td>
<td>37/51</td>
<td>Overnight fasting/(^{13})C-glucose ingestion</td>
<td>IR</td>
<td>(^{13})CO(_2) ((-))</td>
<td>100.00%</td>
<td>96.90%</td>
<td>98.27%</td>
<td>0.993</td>
</tr>
<tr>
<td>Jeon et al.</td>
<td>2018</td>
<td>Korea</td>
<td>T2D</td>
<td>12/13</td>
<td>2h</td>
<td>Sensor</td>
<td>Breathprint</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Study</td>
<td>Year</td>
<td>Country</td>
<td>Condition</td>
<td>No.</td>
<td>Baseline Measurements</td>
<td>Methodology</td>
<td>Acetone (%)</td>
<td>Isopropanol (%)</td>
<td>Acetone + Isopropanol (%)</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>------</td>
<td>---------</td>
<td>------------</td>
<td>-----</td>
<td>------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>---------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Tyas et al.</td>
<td>2018</td>
<td>Indonesia</td>
<td>T2D</td>
<td>31/31</td>
<td>-</td>
<td>Laser photoacoustic spectrometer</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Takemoto et al.</td>
<td>2018</td>
<td>Japan</td>
<td>T2D</td>
<td>20/16</td>
<td>Overnight fasting/</td>
<td>ND-IR</td>
<td>13CO₂⁺ (-)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>¹³C-glucose ingestion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yucel et al.</td>
<td>2018</td>
<td>Turkey</td>
<td>PD</td>
<td>10/10</td>
<td>-</td>
<td>Sensor</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ghosh et al.</td>
<td>2017</td>
<td>India</td>
<td>T2D</td>
<td>23/28</td>
<td>Overnight fasting/</td>
<td>Laser-based integrated cavity output spectroscopy (ICOS)</td>
<td>13CO₂⁺ (-)</td>
<td>95.7%</td>
<td>92.10%</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>¹³C-glucose ingestion</td>
<td></td>
<td>C¹⁸O₂⁺ (+)</td>
<td>89.3% (95.8%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chien et al.</td>
<td>2017</td>
<td>Japan</td>
<td>T1D (n=4)</td>
<td>25/55</td>
<td>No smoking in 1h;</td>
<td>Sensor</td>
<td>Acetone (+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>T2D (n=21)</td>
<td></td>
<td>No alcohol drinking in 24h; 15min rest for gas exchange</td>
<td></td>
<td>Isopropanol (+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Polanowska et al.</td>
<td>2017</td>
<td>Poland</td>
<td>T2D (+chronic kidney disease)</td>
<td>10/10</td>
<td>-</td>
<td>GC-MS</td>
<td>Trimethylamine (+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Metanethiol (+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dimethyl-sulfide (+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Li W et al.</td>
<td>2017</td>
<td>China</td>
<td>T2D</td>
<td>85/56</td>
<td>Overnight fasting/</td>
<td>GC-MS</td>
<td>Acetone</td>
<td>62.40%</td>
<td>70.9%</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Exercise restrained/</td>
<td></td>
<td>Isopropanol</td>
<td>75.30%</td>
<td>79.4%</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acetone + Isopropanol</td>
<td>71.80%</td>
<td>87.50%</td>
<td>78.0%</td>
<td>0.86</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Country</td>
<td>Group</td>
<td>Total</td>
<td>Test Type</td>
<td>Indicator</td>
<td>T1D</td>
<td>T2D</td>
<td>p Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>---------</td>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
<td>----------</td>
<td>-----</td>
<td>-----</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jiang et al.</td>
<td>2016</td>
<td>China</td>
<td>T1D (n=22), T2D (n=312)</td>
<td>334/52</td>
<td>Overnight fasting/postprandial</td>
<td>Cavity ringdown spectroscopy (CRDS)</td>
<td>Acetone(+)</td>
<td>-</td>
<td>-</td>
<td>74% for T1D</td>
<td></td>
</tr>
<tr>
<td>Ghosh et al.</td>
<td>2016</td>
<td>India</td>
<td>T1D (n=14), T2D (n=24)</td>
<td>38/27</td>
<td>Overnight fasting/13C-glucose ingestion</td>
<td>Laser-based ICOS</td>
<td>C(^{13})O(_2) * (+) T1D</td>
<td>92.90%</td>
<td>100%</td>
<td>96.40%</td>
<td>0.991</td>
</tr>
<tr>
<td>Sun et al.</td>
<td>2015</td>
<td>China</td>
<td>T2D</td>
<td>149/42</td>
<td>Overnight fasting/2h postprandial</td>
<td>GC-MS/CDRS</td>
<td>Acetone(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ghosh et al.</td>
<td>2015</td>
<td>India</td>
<td>PD (n=39), T2D (n=48)</td>
<td>87/32</td>
<td>Overnight fasting/glucose ingestion</td>
<td>Laser-based ICOS</td>
<td>C(^{13})O(_2) * (+)</td>
<td>96%</td>
<td>92%</td>
<td>94.2%</td>
<td>0.957</td>
</tr>
<tr>
<td>Obermeier et al.</td>
<td>2015</td>
<td>Germany</td>
<td>Indiscriminate</td>
<td>15/14</td>
<td>-</td>
<td>Sensor</td>
<td>Aldehydes(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ghosh et al.</td>
<td>2015</td>
<td>India</td>
<td>PD (n=38), T2D (n=47)</td>
<td>85/31</td>
<td>Overnight fasting/13C-glucose ingestion</td>
<td>Laser-based ICOS</td>
<td>C(^{13})O(_2) * (-)</td>
<td>92.10%</td>
<td>100%</td>
<td>95.90%</td>
<td>0.963</td>
</tr>
<tr>
<td>Rydloz</td>
<td>2014</td>
<td>Poland</td>
<td>T1D</td>
<td>14/11</td>
<td>2h after glucose ingestion</td>
<td>LTCC-MS</td>
<td>Acetone(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Yan K et al.</td>
<td>2014</td>
<td>China</td>
<td>T1D (n=1)</td>
<td>87/n&gt;25</td>
<td>2h post</td>
<td>Sensor</td>
<td>Breathprint</td>
<td>91.51%</td>
<td>90.77%</td>
<td>91.14%</td>
<td>n.a</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Year</td>
<td>Country</td>
<td>Diagnosis</td>
<td>Sample Size</td>
<td>Procedure</td>
<td>Method</td>
<td>Acetone(+)</td>
<td>Isopropanol(+)</td>
<td>Toluene(+)</td>
<td>2,3,4-Trimethylhexane(+)</td>
<td>2,6,8-Trimethyldecane(+)</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>---------</td>
<td>-----------</td>
<td>-------------</td>
<td>------------</td>
<td>--------</td>
<td>------------</td>
<td>---------------</td>
<td>------------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Zhou M et al.</td>
<td>2014</td>
<td>China</td>
<td>T2D</td>
<td>25/44</td>
<td>Overnight Fasting/ Exercise restrained</td>
<td>GC-MS</td>
<td>100%</td>
<td>97.7%</td>
<td>98.6%</td>
<td>n.a</td>
<td></td>
</tr>
<tr>
<td>Yan Y et al.</td>
<td>2014</td>
<td>China</td>
<td>T2D</td>
<td>48/39</td>
<td>-</td>
<td>GC-MS</td>
<td>45.8%</td>
<td>87.2%</td>
<td>64.4%</td>
<td>0.679</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>79.2%</td>
<td>92.3%</td>
<td>85.1%</td>
<td>0.876</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>66.7%</td>
<td>82.1%</td>
<td>73.6%</td>
<td>0.737</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89.6%</td>
<td>94.9%</td>
<td>92.0%</td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89.6%</td>
<td>94.9%</td>
<td>92.0%</td>
<td>0.949</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89.6%</td>
<td>84.6%</td>
<td>87.4%</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89.6%</td>
<td>82.1%</td>
<td>86.2%</td>
<td>0.911</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>72.9% (69.2%)</td>
<td>69.2% (72.9%)</td>
<td>71.3%</td>
<td>0.677</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghosh et al.</td>
<td>2014</td>
<td>India</td>
<td>PD (n=22)</td>
<td>47/18</td>
<td>Overnight fasting/ 13C-glucose ingestion</td>
<td>RGA-MS</td>
<td>100%</td>
<td>94.40%</td>
<td>97.5%</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T2D (n=25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goerl et al.</td>
<td>2013</td>
<td>Germany</td>
<td>T1D (+end-stage renal disease)</td>
<td>10/20</td>
<td>-</td>
<td>GC-MS</td>
<td>Acetone(-)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanaka et al.</td>
<td>2013</td>
<td>Japan</td>
<td>Indiscriminate</td>
<td>15/40</td>
<td>Overnight fasting/ 13C-glucose ingestion</td>
<td>IR</td>
<td>13CO₂ * (-)</td>
<td>100%</td>
<td>95%</td>
<td>96.4%</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsuchiya et al.</td>
<td>2012</td>
<td>Japan</td>
<td>T2D (+liver cirrhosis)</td>
<td>11/27(19 IGT)</td>
<td>Overnight fasting/</td>
<td>IR</td>
<td>13CO₂ * (-)</td>
<td>81.8%</td>
<td>85.2%</td>
<td>84.2%</td>
<td>0.88</td>
</tr>
<tr>
<td>Year</td>
<td>Country</td>
<td>Group</td>
<td>Number</td>
<td>Method</td>
<td>Test</td>
<td>Accuracy (%)</td>
<td>Specificity (%)</td>
<td>Sensitivity (%)</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>------------------</td>
<td>--------</td>
<td>-------------------</td>
<td>------</td>
<td>--------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>China</td>
<td>Indiscriminate</td>
<td>404/294</td>
<td>SensorBreathprint</td>
<td></td>
<td>91.43%</td>
<td>89.86%</td>
<td>90.65%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>China</td>
<td>Indiscriminate</td>
<td>117/108</td>
<td>SensorBreathprint</td>
<td></td>
<td>87.67%</td>
<td>86.87%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>Germany</td>
<td>T2D</td>
<td>21/26</td>
<td>PTR-MS</td>
<td>Acetone(+)</td>
<td>76.2%</td>
<td>80.8%</td>
<td>78.7%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>China</td>
<td>Indiscriminate</td>
<td>90/108</td>
<td>SensorBreathprint</td>
<td></td>
<td>93.52%</td>
<td>92.66%</td>
<td>92.93%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>China</td>
<td>T1D (n=34) T2D (n=10)</td>
<td>44/15</td>
<td>CRDS</td>
<td>Acetone(+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>China</td>
<td>T2D</td>
<td>15/15</td>
<td>GC-MS</td>
<td>Acetone(+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>Canada</td>
<td>T2D</td>
<td>52/50 (+2 undefined)</td>
<td>IR-MS</td>
<td>$^{13}$CO$_2$ * (-)</td>
<td>73.0%</td>
<td>92.0%</td>
<td>82.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>China</td>
<td>Indiscriminate</td>
<td>18/14</td>
<td>SensorBCrprint</td>
<td>CO(+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>UK</td>
<td>T1D (n=8) T2D (n=16)</td>
<td>24/37</td>
<td>Sensor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>China</td>
<td>T1D (Idiopathic)</td>
<td>18/14</td>
<td>Sensor</td>
<td>Breathprint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td>Belgium</td>
<td>T1D (n=5) CD (n=7)</td>
<td>12/11</td>
<td>MS</td>
<td>$^{13}$CO$_2$ * (-)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $^{13}$CO$_2$ * indicates the presence of $^{13}$CO$_2$ in the breathprint.

**Notes:**
- BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s).
- Supplemented material is provided with this document.
- doi: 10.1136/bmjdc-2021-002174 e002174.
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tassopoulos et al.</td>
<td>1969</td>
<td>UK</td>
<td>Indiscriminate</td>
</tr>
<tr>
<td>Rooth, et al.</td>
<td>1966</td>
<td>Sweden</td>
<td>T1D (n=49)</td>
</tr>
<tr>
<td>Stewart et al.</td>
<td>1964</td>
<td>USA</td>
<td>T1D(n=33) T2D(n=96)</td>
</tr>
</tbody>
</table>

*Biomarkers from exogenous test meal.

*Same population may be used, these studies were excluded in meta-analysis.


Data in bold gray font were corrected according the available data provided in the literatures, and (x) were data provided in original text.
Table S3. Detailed population information of the included studies. The FPG units were all converted to mg/dL, while all HbA1c were converted to DCCT-HbA1c (%).

<table>
<thead>
<tr>
<th>Studies</th>
<th>Number of participants in each group</th>
<th>Age</th>
<th>Gender (male/female)</th>
<th>BMI</th>
<th>FPG mg/dL</th>
<th>HbA1c/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalidoss et al., 2020</td>
<td>30 DM</td>
<td>37 ± 9</td>
<td>13/17</td>
<td>25.7 ± 3.2</td>
<td>123.2 ± 21.9</td>
<td>7.00 ± 2.97</td>
</tr>
<tr>
<td></td>
<td>30 ND</td>
<td>44 ± 9</td>
<td>17/13</td>
<td>26.4 ± 3.1</td>
<td>103.4 ± 21.8</td>
<td>5.86 ± 2.98</td>
</tr>
<tr>
<td>Kistenev et al., 2020</td>
<td>11 DM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>11 ND</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kalidoss et al., 2019</td>
<td>17 DM</td>
<td>35±3</td>
<td>4/9</td>
<td>25.7±3.2</td>
<td>126.52±22.14</td>
<td>7.12±2.60</td>
</tr>
<tr>
<td></td>
<td>13 ND</td>
<td>50±11</td>
<td>11/6</td>
<td>26.4±3.1</td>
<td>96.07±22.42</td>
<td>5.70±2.89</td>
</tr>
<tr>
<td>Trefz et al., 2019</td>
<td>53 T1D</td>
<td>12.36±3.15</td>
<td>32/21</td>
<td>0.35±0.69</td>
<td>-</td>
<td>8.60±(-)</td>
</tr>
<tr>
<td></td>
<td>60 ND</td>
<td>13.67±2.80</td>
<td>28/32</td>
<td>0.06±0.82</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tazi et al., 2019</td>
<td>25 DM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nevezhin et al., 2019</td>
<td>22 T2D</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yu et al., 2019</td>
<td>16 DM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Jeon et al., 2018</td>
<td>12 T2D</td>
<td>55.45±16.86</td>
<td>6/6</td>
<td>24.0±3.7</td>
<td>-</td>
<td>7.2±1.7</td>
</tr>
<tr>
<td></td>
<td>13 ND</td>
<td>34.92±11.26</td>
<td>4/9</td>
<td>21.7±3.0</td>
<td>-</td>
<td>5.2±0.3</td>
</tr>
<tr>
<td>Som et al., 2018</td>
<td>37 T2D (+Helicobacter pylori)</td>
<td>50 ± 9.5</td>
<td>73/58</td>
<td>-</td>
<td>159±11.4</td>
<td>10.7±1.1</td>
</tr>
<tr>
<td></td>
<td>51 ND (+Helicobacter pylori)</td>
<td>46 ± 7</td>
<td>35/63</td>
<td>-</td>
<td>101±6.4</td>
<td>5.31±0.27</td>
</tr>
<tr>
<td></td>
<td>43 ND</td>
<td>47 ± 10</td>
<td>-</td>
<td>-</td>
<td>97±10.1</td>
<td>5.24±0.19</td>
</tr>
<tr>
<td>Yucel et al., 2018</td>
<td>10 PD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100-125</td>
<td>-</td>
</tr>
<tr>
<td>Takemoto et al., 2018</td>
<td>20 T2D</td>
<td>58.4±14.4</td>
<td>15/5</td>
<td>24.6±4.0</td>
<td>-</td>
<td>9.2±2.2</td>
</tr>
<tr>
<td></td>
<td>16 ND</td>
<td>21.5±1.0</td>
<td>7/9</td>
<td>20.8±2.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tyas et al., 2018</td>
<td>31 T2D</td>
<td>25-80</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Polanowska et al., 2017</td>
<td>10 T2D (+chronic kidney disease)</td>
<td>49.4±6.4</td>
<td>4/6</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>10 ND</td>
<td>38.1±2.7</td>
<td>4/6</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Chien et al., 2017</td>
<td>4 T1D</td>
<td>44.5 ± 17.2</td>
<td>9/16</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>21 T2D</td>
<td>64.5 ± 10.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>55 ND</td>
<td>30.8 ± 10.9</td>
<td>41/14</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ghosh et al., 2017</td>
<td>28 T2D</td>
<td>36.1±8.5</td>
<td>14/9</td>
<td>24.8±1.7</td>
<td>119.6±13.4</td>
<td>8.6±1.1</td>
</tr>
<tr>
<td></td>
<td>23 ND</td>
<td>33.2±10.1</td>
<td>17/11</td>
<td>23.7±1.9</td>
<td>237.2±49.2</td>
<td>5.3±0.3</td>
</tr>
<tr>
<td>Li W et al., 2017</td>
<td>85 T2D</td>
<td>20-85</td>
<td>41/44</td>
<td>-</td>
<td>270.0±121.86</td>
<td>10.56±2.60</td>
</tr>
<tr>
<td></td>
<td>56 ND</td>
<td>21-71</td>
<td>21/32</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Jiang et al., 2017</td>
<td>22 T1D/ (Helicobacter pylori)</td>
<td>28.1±13.4</td>
<td>8/14</td>
<td>20.6±2.5</td>
<td>203.3±53.9</td>
<td>n.a</td>
</tr>
<tr>
<td>Year</td>
<td>Study/Condition</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>312 T2D</td>
<td>57.0±12.9</td>
<td>178/134</td>
<td>26.5±3.8</td>
<td>152.1±41.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>52 ND</td>
<td>27.0±4.8</td>
<td>22/30</td>
<td>22.1±3.1</td>
<td>91.8±7.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14 T1D</td>
<td>21.7±13.8</td>
<td>8/6</td>
<td>19.4±2.9</td>
<td>272.6±107.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 T2D</td>
<td>36.2±11.7</td>
<td>15/9</td>
<td>24.6±2.3</td>
<td>132.0±12.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27 ND</td>
<td>31.7±8.9</td>
<td>16/11</td>
<td>23.9±2.4</td>
<td>87.7±9.8</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>149 T2D/</td>
<td>57.2±13.4</td>
<td>67/82</td>
<td>26.2±3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>42 ND</td>
<td>27.8±6.3</td>
<td>23/19</td>
<td>22.7±3.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48 T2D</td>
<td>38.6±9.7</td>
<td>19/29</td>
<td>23.8±2</td>
<td>90.1±6.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>39 PD</td>
<td>34.6±6.5</td>
<td>14/25</td>
<td>24.17±2.4</td>
<td>107.0±5.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>32 ND</td>
<td>35.9±7.1</td>
<td>13/19</td>
<td>24.28±2.8</td>
<td>140.4±7.7</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>15 DM/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14 ND</td>
<td></td>
<td>7/7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47 T2D</td>
<td>38.6±9.8</td>
<td>31/16</td>
<td>23.9±2</td>
<td>196±17.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>38 PD</td>
<td>34.5±6.6</td>
<td>26/12</td>
<td>24.21±2.5</td>
<td>107.2±9.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31 ND</td>
<td>35.8±6.8</td>
<td>18/13</td>
<td>24.43±2.7</td>
<td>90.1±12.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 T2D</td>
<td>50.0±15.4</td>
<td>15/10</td>
<td>23.36±1.2</td>
<td>221.5±80.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22 PD</td>
<td>51.45±15.3</td>
<td>12/10</td>
<td>23.83±0.7</td>
<td>110.6±10.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18 ND</td>
<td>41.7±11.5</td>
<td>11/7</td>
<td>23.37±0.8</td>
<td>98.6±11.8</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>1 T1D/6 T2D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n=25 ND</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 T2D</td>
<td>42.7±2.9</td>
<td>12/13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>44 ND</td>
<td>29.3±1.8</td>
<td>25/19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48 T2D</td>
<td>14.85</td>
<td>22/26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>39 ND</td>
<td>21.71</td>
<td>15/24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>14 T1D</td>
<td>32.09±6.01</td>
<td>6/5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11 ND</td>
<td>28.2±9.74</td>
<td>6/8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>15 DM</td>
<td></td>
<td>8/7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40 ND</td>
<td></td>
<td>19/21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>10 T1D (+end-stage renal disease)</td>
<td>73.3±6.5</td>
<td>5/5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 ND (+end-stage renal disease)</td>
<td>60.7±14.7</td>
<td>11/9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>11 DM (+liver cirrhosis)</td>
<td>67.5±10.5</td>
<td>23/15</td>
<td>23.1±3.0</td>
<td>93.1±10.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19 IGT (+liver cirrhosis)</td>
<td></td>
<td></td>
<td></td>
<td>5.1±0.4</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>8 ND (+liver cirrhosis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>404 DM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>117 DM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>108 ND</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Type</td>
<td>Age Median (Range)</td>
<td>Blood Glucose Median (Range)</td>
<td>Weight Median (Range)</td>
<td>BMI Median (Range)</td>
<td>Other Data</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>---------</td>
<td>-------------------</td>
<td>------------------------------</td>
<td>-----------------------</td>
<td>------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Guo et al., 2010</td>
<td>90 DM</td>
<td>25-77</td>
<td>49/41</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Greiter et al., 2010</td>
<td>21 T2D</td>
<td>66.1±8.3/</td>
<td>28.9±4.9</td>
<td>7.7±1.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Greiter et al., 2010</td>
<td>26 ND</td>
<td>40-59</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wang C et al., 2009</td>
<td>34 T1D</td>
<td>46.6±19.5</td>
<td>14/20</td>
<td>7.8±1.9(n=20)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wang C et al., 2009</td>
<td>10 T2D</td>
<td>54.5±12.2</td>
<td>3/6(a unknown)</td>
<td>8.8±1.9(n=3)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wang C et al., 2009</td>
<td>15 ND</td>
<td>25–45</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Deng et al., 2004</td>
<td>15 T2D</td>
<td>54.1±(-)</td>
<td>7/8</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Deng et al., 2004</td>
<td>15 ND</td>
<td>47.6±(-)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yatscoff et al., 2003</td>
<td>54 T2D</td>
<td>54±11</td>
<td>31.6±5.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yatscoff et al., 2003</td>
<td>50 ND</td>
<td>44±14</td>
<td>28.8±5.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zhang et al., 2000</td>
<td>18 DM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zhang et al., 2000</td>
<td>14 ND</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Paredi et al., 1999</td>
<td>8 T1D</td>
<td>50 ± 8</td>
<td>4/4</td>
<td>7.5±0.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Paredi et al., 1999</td>
<td>16 T2D</td>
<td>63 ± 8</td>
<td>5/11</td>
<td>4.9±0.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Paredi et al., 1999</td>
<td>37 ND</td>
<td>33 ± 3</td>
<td>20/17</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wang et al., 1997</td>
<td>18 DM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wang et al., 1997</td>
<td>14 ND</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lefebvre et al., 1974</td>
<td>5 T1D</td>
<td>22.6±9.2</td>
<td>5/0</td>
<td>19.3 ± 1.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lefebvre et al., 1974</td>
<td>7 CD</td>
<td>35.3±13.3</td>
<td>2/5</td>
<td>32.4 ± 6.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lefebvre et al., 1974</td>
<td>11 ND</td>
<td>25.6±4.8</td>
<td>9/2</td>
<td>21.7 ± 2.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tassopoulos et al., 1969</td>
<td>251 DM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tassopoulos et al., 1969</td>
<td>45 ND</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rooth et al., 1966</td>
<td>49 T1D</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rooth et al., 1966</td>
<td>20 T2D</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rooth et al., 1966</td>
<td>40 PD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rooth et al., 1966</td>
<td>42 CD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rooth et al., 1966</td>
<td>67 ND</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Stewart et al., 1964</td>
<td>33 T1D</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Stewart et al., 1964</td>
<td>96 T2D</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Stewart et al., 1964</td>
<td>40 ND</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table S4. The QUADAS-2 referenced to the Hanna, et. al.’s version.

<table>
<thead>
<tr>
<th>Applicability concern</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient selection</strong></td>
<td>Were sampled patients representative of the intended population?</td>
</tr>
<tr>
<td></td>
<td>Did the study include both positive (benign conditions) as well as healthy controls?</td>
</tr>
<tr>
<td></td>
<td>Did the study avoid inappropriate exclusions?</td>
</tr>
<tr>
<td><strong>Index test</strong></td>
<td>Was the index test and interpretation of data performed in standardised and reproducible fashion?</td>
</tr>
<tr>
<td></td>
<td>Was validation of results performed (internal or external)?</td>
</tr>
<tr>
<td><strong>Reference standard</strong></td>
<td>Is the reference standard likely to correctly classify the target condition?</td>
</tr>
<tr>
<td><strong>Flow and timing</strong></td>
<td>Was there an appropriate interval between index test and reference standard?</td>
</tr>
<tr>
<td></td>
<td>#Was diet control performed prior to index test and reference standard?</td>
</tr>
<tr>
<td></td>
<td>Did all patients receive the same reference standard?</td>
</tr>
<tr>
<td></td>
<td>Were all patients included in the analysis?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Applicability concern</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient selection</strong></td>
<td>Are there concerns that the included patients and setting do not match the review question?</td>
</tr>
<tr>
<td><strong>Index test</strong></td>
<td>Are there concerns that the authors have not demonstrated suitable reproducibility and sensitivity of the chosen index test?</td>
</tr>
<tr>
<td><strong>Reference standard</strong></td>
<td>Are there concerns that the target condition as defined by the reference standard does not match the question?</td>
</tr>
</tbody>
</table>

# The signalling question further modified
Figure S1. The summary of risk of bias and applicability concern.
Figure S2. Risk of bias and applicability concern using QUADAS-2.
Table S5. The STARD assessment of each study.

<table>
<thead>
<tr>
<th>Studies</th>
<th>Title</th>
<th>Abs.</th>
<th>Intro.</th>
<th>Methods</th>
<th>Results</th>
<th>Discussion</th>
<th>Other information</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Kalidoss, 2020</td>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Kistenev, 2020</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Trefz, 2019</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yu, 2019</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nevezhin, 2019</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Li W, 2017</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Som, 2018</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jeon, 2018</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tyas, 2018</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yucel, 2018</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Takemoto, 2018</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chien et al., 2017</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ghosh, 2017</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Li W, 2017</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Polansowska, 2017</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jiang, 2016</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ghosh, 2016</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ghosh, 2015</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Sun, 2015</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Obermeier, 2015</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ghosh, 2015</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s).
<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yan K.</td>
<td>2014</td>
<td>1 0 1 1 1 0 0 1 1 1 1 0 0 0 1 0 1 0 0 0 0 1 1 0 1 0 0 0 0 1 1 0 1 0 0 1 1</td>
</tr>
<tr>
<td>Zhou M.</td>
<td>2014</td>
<td>1 0 1 1 1 1 0 0 1 1 1 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 1 1 0 1 0 0 0 1 1</td>
</tr>
<tr>
<td>Yan Y.</td>
<td>2014</td>
<td>1 0 1 1 1 1 1 1 0 1 1 1 0 0 1 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1 1</td>
</tr>
<tr>
<td>Rydosz,</td>
<td>2014</td>
<td>0 0 1 0 1 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0 1 1</td>
</tr>
<tr>
<td>Ghosh,</td>
<td>2014</td>
<td>0 0 1 1 1 1 0 0 0 1 1 1 1 1 1 0 1 0 0 0 0 0 1 1 0 0 0 0 1 1 0 1 0 0 1 1</td>
</tr>
<tr>
<td>Tanaka,</td>
<td>2013</td>
<td>0 0 0 1 1 1 1 0 0 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 1 0 0 1 1</td>
</tr>
<tr>
<td>Goerl,</td>
<td>2013</td>
<td>0 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 1 5</td>
</tr>
<tr>
<td>Tsuchiya,</td>
<td>2012</td>
<td>1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 0 1 0 0 0 0 0 1 1 0 1 0 1 0 0 1 0 0 0 0 0 0 1 18</td>
</tr>
<tr>
<td>Yan K.</td>
<td>2012</td>
<td>0 0 1 1 1 0 0 1 1 1 0 0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1 1 1 11</td>
</tr>
<tr>
<td>Guo,</td>
<td>2010</td>
<td>0 0 0 1 1 1 0 0 1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 1 0 0 1 0 0 1 10</td>
</tr>
<tr>
<td>Greiter,</td>
<td>2010</td>
<td>1 1 1 1 1 1 1 0 0 1 1 1 0 0 1 1 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0 1 1 1 16</td>
</tr>
<tr>
<td>Guo,</td>
<td>2010</td>
<td>0 0 1 1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 0 0 0 1 0 0 1 0 0 1 13</td>
</tr>
<tr>
<td>Wang C.</td>
<td>2009</td>
<td>0 0 1 1 1 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 1 0 0 1 0 0 1 14</td>
</tr>
<tr>
<td>Deng,</td>
<td>2004</td>
<td>0 0 1 0 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 1 1 0 1 0 1 9 1</td>
</tr>
<tr>
<td>Yatscoff,</td>
<td>2003</td>
<td>0 0 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 0 1 1 0 0 0 0 1 0 0 0 1 18</td>
</tr>
<tr>
<td>Zhang,</td>
<td>2000</td>
<td>0 0 1 1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0 1 1 1 8</td>
</tr>
<tr>
<td>Paredi,</td>
<td>1999</td>
<td>0 0 1 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 0 7 1</td>
</tr>
<tr>
<td>Wang P.</td>
<td>1997</td>
<td>0 0 1 1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 1 8</td>
</tr>
<tr>
<td>Lefebvre,</td>
<td>1974</td>
<td>0 0 0 0 0 1 1 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0 1 1 8 1</td>
</tr>
<tr>
<td>Tassopoulos,</td>
<td>1969</td>
<td>0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 4 1</td>
</tr>
<tr>
<td>Rooth,</td>
<td>1966</td>
<td>0 0 0 0 0 1 1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 1 1 8 1</td>
</tr>
<tr>
<td>Stewart,</td>
<td>1964</td>
<td>0 0 0 1 1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 1 0 0 0 1 0 0 0 9</td>
</tr>
</tbody>
</table>
Figure S3. The forest plot of (a) sensitivity, (b) specificity and (c) the summary receiver operating characteristic curve for overall breath tests.
Figure S4. The funnel plot of the publication bias. (a) T2D group, (b) isotopic CO₂, (c) acetone.
Figure S5. A random effect model was applied, and the effect sizes for the comparison of breath acetone concentrations in (a) T1D vs ND, (b) T2D vs ND, (c) T1D vs T2D were presented. The comparison of T1D vs ND, T2D vs ND, and T1D vs T2D exhibited WMD of 1.374 (95%CI 0.986-1.762, z=6.94, p=0.000), 0.845 (95%CI 0.605-1.085, z=6.91, p=0.000), and 0.725 (95%CI -0.291-1.741, z=1.40, p=0.162), respectively.
Reference


42. Tassopoulos CN, Barnett D, Fraser TR. BREATH-ACETONE AND BLOOD-SUGAR MEASUREMENTS IN DIABETES. *Lancet* 1969;1(7609):1282+-.
