Efficacy and effectiveness of screen and treat policies in prevention of type 2 diabetes: systematic review and meta-analysis of screening tests and interventions

Eleanor Barry, Samantha Roberts, Jason Oke, Shanti Vijayaraghavan, Rebecca Normansell, Trisha Greenhalgh

ABSTRACT
OBJECTIVES
To assess diagnostic accuracy of screening tests for pre-diabetes and efficacy of interventions (lifestyle or metformin) in preventing onset of type 2 diabetes in people with pre-diabetes.

DESIGN
Systematic review and meta-analysis.

DATA SOURCES AND METHOD
Medline, PreMedline, and Embase. Study protocols and seminal papers were citation-tracked in Google Scholar to identify definitive trials and additional publications. Data on study design, methods, and findings were extracted onto Excel spreadsheets; a 20% sample was checked by a second researcher. Data extracted for screening tests included diagnostic accuracy and population prevalence. Two meta-analyses were performed, one summarising accuracy of screening tests (with the oral glucose tolerance test as the standard) for identification of pre-diabetes, and the other assessing relative risk of progression to type 2 diabetes after either lifestyle intervention or treatment with metformin.

ELIGIBILITY CRITERIA
Empirical studies evaluating accuracy of tests for identification of pre-diabetes. Interventions (randomised trials and interventional studies) with a control group in people identified through screening. No language restrictions.

RESULTS
2874 titles were scanned and 148 papers (covering 138 studies) reviewed in full. The final analysis included 49 studies of screening tests (five of which were prevalence studies) and 50 intervention trials. HbA₁c had a mean sensitivity of 0.49 (95% confidence interval 0.40 to 0.58) and specificity of 0.79 (0.73 to 0.84), for identification of pre-diabetes, though different studies used different cut-off values. Fasting plasma glucose had a mean sensitivity of 0.25 (0.19 to 0.32) and specificity of 0.94 (0.92 to 0.96). Different measures of glycaemic abnormality identified different subpopulations (for example, 47% of people with abnormal HbA₁c had no other glycaemic abnormality). Lifestyle interventions were associated with a 36% (28% to 43%) reduction in relative risk of type 2 diabetes over six months to six years, attenuating to 20% (8% to 31%) at follow-up in the period after the trials.

CONCLUSIONS
HbA₁c is neither sensitive nor specific for detecting pre-diabetes; fasting glucose is specific but not sensitive. Interventions in people classified through screening as having pre-diabetes have some efficacy in preventing or delaying onset of type 2 diabetes in trial populations. As screening is inaccurate, many people will receive an incorrect diagnosis and be referred on for interventions while others will be falsely reassured and not offered the intervention. These findings suggest that “screen and treat” policies alone are unlikely to have substantial impact on the worsening epidemic of type 2 diabetes.

REGISTRATION
PROSPERO (No CRD42016042920).

Introduction
The prevalence of type 2 diabetes is rising globally; 422 million adults are living with diabetes, and the number expected to die from its complications is predicted to double between 2005 and 2030. In the United Kingdom about 3.2 million people have type 2 diabetes, and by 2025 it is predicted that this will increase to five million. This places considerable financial burden on the National Health Service (NHS). The healthcare cost of diabetes is estimated to be £23.7bn ($30.2bn, €28.2bn), a figure expected to rise to £39.8bn by 2035-36. Preventing or delaying type 2 diabetes has become an international priority.

There are two approaches to prevention: screen and treat, in which a subpopulation is identified as “high risk” and offered individual intervention, and a population-wide approach, in which everyone is targeted via public health policies on environmental moderators (sociocultural influences, socioeconomic influences, transport, green spaces). Finland is taking a multi-level
Definition of terms

Oral glucose tolerance test
- Two part blood test
- Part one: fasting plasma glucose (FPG). Blood test after overnight fast. If result is abnormal, diagnosis is impaired fasting glucose (IFG)
- Part two: 2 hour glucose tolerance test (2hGTT). Blood test two hours after ingestion of sugary drink. If result is abnormal, diagnosis is impaired glucose tolerance (IGT)
- Both tests can be performed independently of each other

HbA1c
- Measurement of glycated haemoglobin, which reflects glucose concentration over two to three months. Accuracy impaired by haemoglobinopathies

Pre-diabetes
- Arbitrary category to encompass either IFG or IGT or abnormal HbA1c

American Diabetes Association (ADA) diagnostic criteria
- Impaired fasting glucose 5.6-6.9 mmol/L
- Impaired glucose tolerance 7-11.1 mmol/L
- HbA1c “at risk” range 39-47 mmol/mol (5.7-6.4%)

WHO diagnostic criteria
- Impaired fasting glucose 6.0-6.9 mmol/L
- Impaired glucose tolerance 7-11.1 mmol/L
- HbA1c “at risk” range 42-47 mmol/mol (6.0-6.4%)

International Expert Committee (IEC) diagnostic criteria
- HbA1c “at risk” range 42-47 mmol/mol (6.0-6.4%)

approach to prevention by using both strategies. In contrast, the UK’s National Diabetes Prevention Programme follows Australia and the United States in placing the emphasis on a screen and treat approach. There is international inconsistency on how to identify individuals at high risk of diabetes, to the extent that “a transatlantic trip may cure or cause diabetes simply as a result of small but important differences in diagnostic criteria.” In the US, the American Diabetes Association criteria recommend a diagnosis of pre-diabetes in people with a fasting plasma glucose concentration of 5.6-6.9 mmol/L or HbA1c of 39-47 mmol/mol (5.7-6.4%). WHO (World Health Organization) and the International Expert Committee recommend a fasting plasma glucose cut off of 6.0-6.9 mmol/L and HbA1c of 42-47 mmol/mol (6.0-6.4%). The term pre-diabetes is used to encapsulate these ranges and implies that if individuals do not take action they will develop diabetes (though in reality this is not always the case). Since the recognition of pre-disease states (impaired glucose tolerance, impaired fasting glucose, and raised HbA1c), trials of lifestyle interventions have been associated with reduced or delayed onset of type 2 diabetes. Studies of screening and intervention programmes in real world settings, however, are sparse. Women with a history of gestational diabetes have a sevenfold risk of developing diabetes postpartum. These women might not be captured by the pre-diabetes umbrella term because many have normal glycaemic markers at the six week postpartum review and then fail to attend for annual review thereafter. Gestational diabetes is common in certain minority ethnic groups, and in deprived multi-ethnic areas a history of this condition could identify a considerable proportion of individuals who could benefit from preventive interventions.

We sought to inform national and local policymaking on prevention of type 2 diabetes by asking two questions. Which (if any) screening test should be used to identify people at risk of developing type 2 diabetes? What is the efficacy of preventive interventions (lifestyle and/or metformin) in those identified as high risk by screening?

Methods

Search strategy
We sought to identify all diagnostic accuracy and prevalence studies focusing on laboratory assessed HbA1c and fasting plasma glucose (as recommended by the UK NICE (National Institute for Health and Care Excellence) as screening tools. Capillary glucose and HbA1c point of care testing were excluded because of the lower reliability of these tests. For intervention studies we included trials whose participants were aged ≥18 and had been identified as being in one of the “at risk” groups (impaired glucose tolerance, impaired fasting glucose, raised HbA1c, or a history of gestational diabetes). We studied two kinds of intervention: lifestyle programmes and metformin, compared with a control, in any setting, and that included weight change, change in glycaemic index, or incidence of diabetes as an outcome measure. Animal studies, molecular biology studies, studies related to children, surgical interventions, and interventions related to drugs other than metformin were excluded.

The study was undertaken from December 2014 to June 2016. It was commissioned by policymakers in a London borough with high prevalence of type 2 diabetes, and concerns about applicability to a real world setting helped shape the review questions. With assistance from a specialist librarian, three searches were undertaken: one for screening tests for pre-diabetes, another for intervention trials, and a third to identify studies relating to the prevention of type 2 diabetes in women with a history of gestational diabetes. Appendix 1 shows the full search strategy. Search terms (MESH and free text) included test, screening, pre-diabetes, impaired glucose tolerance, impaired fasting glucose, gestational diabetes, post-partum, ethnic groups, metformin, and lifestyle. EB manually extracted relevant titles from this dataset and retrieved abstracts to identify papers for full review. SR checked a random sample of 750 abstracts (20%). Disagreements were resolved by discussion. Bilingual colleagues translated non-English papers and extracted data with guidance from the research team.

Diagnostic accuracy meta-analysis
Diagnostic accuracy studies were tabulated by index test, screening, pre-diabetes, impaired glucose tolerance, impaired fasting glucose, gestational diabetes, post-partum, ethnic groups, metformin, and lifestyle. EB manually extracted relevant titles from this dataset and retrieved abstracts to identify papers for full review. SR checked a random sample of 750 abstracts (20%). Disagreements were resolved by discussion. Bilingual colleagues translated non-English papers and extracted data with guidance from the research team.
an oral glucose tolerance test was the reference standard. We presented these data separately for studies using the WHO criteria and studies not using these criteria (notably, some studies used the more stringent American Diabetes Association criteria to define pre-diabetes). We also pooled studies with the fasting plasma glucose as the index test and 2 hour glucose tolerance test as the reference test. Again we examined the data as a whole as well as separately by diagnostic criteria.

We undertook a bivariate diagnostic random effects meta-analysis to pool study level estimates of diagnostic accuracy using the reitsma function from the R package mada. In each case, we reported the pooled sensitivity, false positive rate, false negative rate, and 95% confidence intervals. We plotted the bivariate summary receiver operating curve (sROC) over points representing study estimates of sensitivity and false positive rate, weighted by study size, and summarised the discriminative ability of each test using the area under the ROC (receiver operating characteristics) curve (AUROC) and the partial AUROC (which restricts the area to the observed false positive rates). Statistical heterogeneity was described with I² statistic for bivariate meta-analysis.

Defining at risk population
To compare differences in the at risk population identified by each test, we undertook a prevalence analysis. Using eulerAPE v3 we analysed raw data from prevalence studies to assess the degree of overlap in the population identified as abnormal by each test. This analysis highlights the differing number of people eligible for interventions, depending on which test and criteria are used. We created Venn diagrams with the area of each ellipse proportional to the prevalence.

**Intervention trial review and meta-analysis**
Data extracted into Excel files from intervention trials included participants’ demographics, type of intervention, intervention length, and primary and secondary outcomes. A second Excel sheet was used to tabulate results, including a clinically significant reduction in BMI (1 kg/m²) or weight (2 kg), clinically significant improvement in glycaemic markers (nornoglycaemia, or reduction in fasting plasma glucose by 0.5 mmol/l, 2 hour glucose tolerance by >1 mmol/L, HbA₁c to <42 mmol/mol), differences in incidence of diabetes between groups and whether this was significant. We included in the meta-analysis any trial that collected data on incidence of diabetes. Data were extracted directly from the publications and processed with RevMan software. Because of the heterogeneity of the data we used a random effects model to create forest plots showing relative risk of developing type 2 diabetes after lifestyle interventions and metformin compared with usual care or no additional intervention.

Assessment of study quality, applicability, and bias
To assess the quality and applicability of the test papers we used the validated QUADAS-2 tool, designed for the evaluation of diagnostic accuracy papers. After the refinement steps as recommended by the creators, two authors (EB and SR) piloted, adapted, and refined the tool before it was applied to all the papers used in the meta-analysis (see appendix 2). The limitations of the intervention trials were assessed with the Cochrane risk of bias tool and the CONSORT checklist. One author (RN) used the GRADE principles to assess the overall quality of the evidence at outcome level. An additional assessment was conducted to examine the extent to which participants were involved in the design of the intervention, if feedback was sought, if non-enrolment reasons were given, and if interventions could be adapted to meet the individual’s needs.

**Patient involvement**
The review was conceptualised by a patient participation group led by the project lead (SV). Patients and clinicians raised questions on how best to identify those at risk of diabetes and explore how the Clinical Commissioning Group can support people in Newham to minimise their risk. In this way, patient and citizen involvement shaped the research question and methods of this review. The authors attended regular project meetings, reporting back the results of the review to the rest of the team, which included GP leads from the practices piloting interventions as well as the area lead for diabetes.

**Results**
**Search results**
Figure 1 shows the review flowchart. We fully reviewed 148 publications (83 relating to diagnostic accuracy testing and 65 relating to intervention trials). Data from...
46 papers were extracted and used to construct the diagnostic accuracy meta-analysis. We reviewed 50 unique intervention trials in full as well as publications related to these (protocol designs, subanalyses).

Diagnostic accuracy of tests for pre-diabetes

Table 1 lists the studies included in the diagnostic accuracy meta-analysis, with country of origin, population demographics, and QUADAS-2 assessment for bias and applicability. Figures 2 and 3 show the ROC curves constructed from data extracted from these trials. The pooled sensitivity of HbA₁c at identifying abnormalities as defined by the oral glucose tolerance test was 0.49 (95% confidence interval 0.40 to 0.58); its specificity was 0.79 (0.73 to 0.84). Data were extracted from studies with both WHO and American Diabetes Association criteria, as well as studies that determined the optimal diagnostic cut-offs using the optimal sensitivity and specificity assessed from their own populations. AUROC are used to estimate the overall diagnostic accuracy of a test with a value of 1 equating to the perfect test. The calculated AUROC of the HbA₁c was 0.71. A low sensitivity, however, leads to a high number of false negative results (that is, people incorrectly identified as not having diabetes). When this is taken into account (with the partial AUROC calculation) the accuracy reduces to 0.59. A subanalysis with the International Expert Committee/WHO criteria for HbA₁c did not alter the sensitivity of this test.

Analysis of studies that used the fasting plasma glucose as the index test found that this test had a sensitivity of 0.25 (95% confidence interval 0.19 to 0.32) and specificity of 0.94 (0.92 to 0.96) at identifying impaired glucose tolerance. The analysis calculated an AUROC of 0.72, with a partial AUROC of 0.42. A subanlysis of studies using the criteria implemented in the UK did not change the results.

The main source of potential bias from these studies was selection bias. In many studies, the sampling strategy was unclear or participants self selected to attend for screening (for example, by answering an invitation or advertisement) rather than using a true population sample (random or consecutive). This was a particular concern in studies of follow-up after gestational diabetes, which usually defined their populations as women who had attended for the oral glucose tolerance tests, with no information on those who did not attend. Most diagnostic accuracy studies scored well on the QUADAS scale for applicability, indicating that the populations of patients were similar to those tested in primary care settings and the use of diagnostic tests and their interpretation was in keeping with our review question. These analyses showed a high level of heterogeneity, indicating that the test performs differently depending on population and setting. These are important considerations in the assessment of the diagnostic accuracy of the tests with specified populations. The results of the QUADAS tool were used to undertake a sensitivity analyses. Exclusion of studies at high risk of bias and outlying studies did not significantly alter the results (tables A and B in appendix 4).

Agreement between different diagnostic tests for pre-diabetes

Only five studies (table 2) gave a comparison of prevalence of pre-diabetes for all three tests (HbA₁c, fasting plasma glucose, 2 hour glucose tolerance test). With current International Expert Committee and WHO guidelines, 27% of the populations studied were identified as having “pre-diabetes” by one of the tests (of whom 48% had a raised HbA₁c alone, fig 4); if American Diabetes Association criteria for the HbA1c is applied to the same cohort, this figure was 49% (of whom 71% had a raised HbA₁c alone, see fig F in appendix 4). There was low agreement between the three tests on which individuals were classified as having pre-diabetes. Figure 4 illustrates this limited overlap. Substitution of the American Diabetes Association criteria for both the oral glucose tolerance test and HbA₁c increased the degree of overlap between the test results, but this doubled the estimated prevalence of pre-diabetes (fig 5).

Interventions to prevent diabetes in screen detected pre-diabetes

Fifty trials met our eligibility criteria (10-78-126 (tables A-D in appendix 3 summarise the methods and results of these studies). Only 25 of the trials (21 of lifestyle interventions alone, two of metformin alone, and two assessed both) had the necessary information available to be included in the meta-analysis. All trials were performed in adults identified as at risk of developing diabetes defined by the oral glucose tolerance test or had a history of gestational diabetes. There was heterogeneity in the number of participants in each trial (ranging from hundreds to thousands), length of interventions (four weeks to six years), intensity of intervention (frequency of contacts), and delivery method. Agreement between raters on data extraction was 100%, with the exception of a single paper in which the authors did not distinguish between primary and secondary outcomes. Of 49 trials, 19 used the development of diabetes as a primary outcome measure. Some trials had begun with this outcome but during the trial substituted it for weight reduction and/or change in glycaemic markers because of low recruitment.79 Many studies showed differences in weight and change in glycaemic markers between groups that were statistically but not clinically significant. At the end of the intervention, 20 of the 49 trials showed a clinically significant reduction in weight between the groups, 15 showed a clinically significant improvement in glycaemic markers, and 23 showed some difference in favour of the intervention arm in the number of people developing diabetes, but this difference was significant only in seven of those trials (tables C and D in appendix 3).

Meta-analysis (fig 6) showed that lifestyle interventions reduced the relative risk of developing diabetes by 31% (95% confidence interval 15% to 44%) if the intervention lasted six months to two years. This translates to 69 (95% confidence interval 56 to 85) out of 1000 people in the lifestyle intervention group developing diabetes compared with 100 out of 1000 without the intervention, or a number needed to treat (NNT) of 33
Table 1 | Details of diagnostic accuracy data for detection of pre-diabetes and QUADAS analysis in studies included in review

<table>
<thead>
<tr>
<th>Country</th>
<th>Reference test criteria</th>
<th>Population information</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
<th>Index sensitivity</th>
<th>Index specificity</th>
<th>QUADAS bias*</th>
<th>QUADAS applicability*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incani 2012</td>
<td>Italy</td>
<td>ADA</td>
<td>One risk factor</td>
<td>0.81</td>
<td>0.37</td>
<td>0.53</td>
<td>0.7</td>
<td>U</td>
</tr>
<tr>
<td>Guo 2014</td>
<td>US</td>
<td>ADA</td>
<td>National survey</td>
<td>0.71</td>
<td>0.57</td>
<td>0.35</td>
<td>0.86</td>
<td>L</td>
</tr>
<tr>
<td>Wu 2013</td>
<td>China</td>
<td>WHO</td>
<td>Rural Chinese</td>
<td>0.75</td>
<td>0.57</td>
<td>0.72</td>
<td>0.61</td>
<td>H</td>
</tr>
<tr>
<td>Lee 2013</td>
<td>Korean</td>
<td>ADA</td>
<td>Korean</td>
<td>0.73</td>
<td>0.39</td>
<td>0.49</td>
<td>0.66</td>
<td>H</td>
</tr>
<tr>
<td>Bhowmik 2012</td>
<td>India (Bangladesh)</td>
<td>WHO</td>
<td>Rural</td>
<td>0.18</td>
<td>0.96</td>
<td>0.64</td>
<td>0.73</td>
<td>L</td>
</tr>
<tr>
<td>Ma 2013</td>
<td>China</td>
<td>WHO/ADA</td>
<td>Urban</td>
<td>0.34</td>
<td>0.76</td>
<td>0.45</td>
<td>0.67</td>
<td>U</td>
</tr>
<tr>
<td>Vlaar 2013</td>
<td>Netherlands</td>
<td>WHO</td>
<td>South Asian</td>
<td>0.35</td>
<td>0.88</td>
<td>0.66</td>
<td>0.68</td>
<td>L</td>
</tr>
<tr>
<td>Bhansali 2012</td>
<td>India (Bangladesh)</td>
<td>WHO</td>
<td>Rural</td>
<td>0.43</td>
<td>0.85</td>
<td>0.36</td>
<td>0.9</td>
<td>L</td>
</tr>
<tr>
<td>Incani 2012</td>
<td>Italy</td>
<td>ADA</td>
<td>One risk factor</td>
<td>0.47</td>
<td>0.43</td>
<td>0.71</td>
<td>0.64</td>
<td>H</td>
</tr>
<tr>
<td>Pinelli 2011</td>
<td>US</td>
<td>ADA</td>
<td>Arabic ethnicity</td>
<td>0.57</td>
<td>0.55</td>
<td>0.14</td>
<td>0.91</td>
<td>L</td>
</tr>
<tr>
<td>Mostafa 2010</td>
<td>UK</td>
<td>WHO</td>
<td>23% South Asian</td>
<td>0.30</td>
<td>0.89</td>
<td>0.40</td>
<td>0.84</td>
<td>L</td>
</tr>
<tr>
<td>Wu 2013</td>
<td>China</td>
<td>WHO</td>
<td>Chinese</td>
<td>0.52</td>
<td>0.65</td>
<td>0.66</td>
<td>0.51</td>
<td>L</td>
</tr>
<tr>
<td>Lorenzo 2010</td>
<td>US</td>
<td>ADA</td>
<td>Mixed</td>
<td>0.63</td>
<td>0.51</td>
<td>0.24</td>
<td>0.85</td>
<td>U</td>
</tr>
<tr>
<td>Mohan 2010</td>
<td>South India</td>
<td>WHO &amp; ADA</td>
<td>Indian</td>
<td>0.20</td>
<td>0.93</td>
<td>0.66</td>
<td>0.62</td>
<td>L</td>
</tr>
<tr>
<td>Zhou 2009</td>
<td>China</td>
<td>WHO</td>
<td>Chinese</td>
<td>0.62</td>
<td>0.79</td>
<td>0.25</td>
<td>0.95</td>
<td>L</td>
</tr>
<tr>
<td>Tanaka 2011</td>
<td>Japan</td>
<td>WHO</td>
<td>Japanese</td>
<td>0.80</td>
<td>0.73</td>
<td>0.76</td>
<td>0.77</td>
<td>U</td>
</tr>
<tr>
<td>Yoshinaga 1996</td>
<td>Japan</td>
<td>WHO</td>
<td>Japanese</td>
<td>0.43</td>
<td>0.65</td>
<td>0.43</td>
<td>0.65</td>
<td>L</td>
</tr>
<tr>
<td>Lin 2014</td>
<td>China</td>
<td>WHO</td>
<td>Chinese</td>
<td>0.39</td>
<td>0.86</td>
<td>0.647</td>
<td>0.828</td>
<td>H</td>
</tr>
<tr>
<td>Yang 2012</td>
<td>China</td>
<td>WHO</td>
<td>One risk factor</td>
<td>0.63</td>
<td>0.94</td>
<td>0.70</td>
<td>0.87</td>
<td>U</td>
</tr>
<tr>
<td>Tay 2014</td>
<td>China</td>
<td>WHO</td>
<td>Raised BMI</td>
<td>0.72</td>
<td>0.50</td>
<td>0.81</td>
<td>0.39</td>
<td>U</td>
</tr>
<tr>
<td>Gingras 2013</td>
<td>Canada</td>
<td>ADA</td>
<td>GDM, raised WC</td>
<td>0.89</td>
<td>0.33</td>
<td>0.76</td>
<td>0.62</td>
<td>L</td>
</tr>
<tr>
<td>Katredy 2013</td>
<td>Ireland</td>
<td>WHO</td>
<td>GDM</td>
<td>0.19</td>
<td>0.88</td>
<td>0.28</td>
<td>0.80</td>
<td>L</td>
</tr>
<tr>
<td>Picon 2012</td>
<td>Spain</td>
<td>ADA</td>
<td>GDM</td>
<td>0.55</td>
<td>0.56</td>
<td>0.23</td>
<td>0.84</td>
<td>H</td>
</tr>
<tr>
<td>Du 2013</td>
<td>China</td>
<td>ADA</td>
<td>One risk factor</td>
<td>0.55</td>
<td>0.78</td>
<td>0.47</td>
<td>0.83</td>
<td>L</td>
</tr>
<tr>
<td>Kumaravel 2012</td>
<td>UK</td>
<td>ADA</td>
<td>One risk factor</td>
<td>0.63</td>
<td>0.88</td>
<td>0.21</td>
<td>0.98</td>
<td>L</td>
</tr>
<tr>
<td>Kim 2013</td>
<td>US</td>
<td>ADA</td>
<td>GDM</td>
<td>0.75</td>
<td>0.62</td>
<td>0.39</td>
<td>0.81</td>
<td>H</td>
</tr>
<tr>
<td>Sun 2013</td>
<td>China</td>
<td>ADA</td>
<td>Metabolic syndrome</td>
<td>0.31</td>
<td>0.92</td>
<td>0.6</td>
<td>0.76</td>
<td>L</td>
</tr>
<tr>
<td>Hu (repeat†)</td>
<td>China</td>
<td>ADA</td>
<td>One risk factor</td>
<td>0.59</td>
<td>0.70</td>
<td>0.42</td>
<td>0.82</td>
<td>—</td>
</tr>
<tr>
<td>Tanaka (repeat†)</td>
<td>Japan</td>
<td>WHO</td>
<td>National survey</td>
<td>0.97</td>
<td>0.68</td>
<td>0.62</td>
<td>0.98</td>
<td>—</td>
</tr>
<tr>
<td>Modan 1984</td>
<td>Israel</td>
<td>ADA</td>
<td>National data set</td>
<td>0.16</td>
<td>0.99</td>
<td>0.92</td>
<td>0.45</td>
<td>L</td>
</tr>
<tr>
<td>Noctor 2012</td>
<td>Ireland</td>
<td>ADA</td>
<td>GDM</td>
<td>0.56</td>
<td>0.97</td>
<td>0.9</td>
<td>0.84</td>
<td>U</td>
</tr>
<tr>
<td>Hu (repeat†)</td>
<td>China</td>
<td>ADA</td>
<td>One risk factor</td>
<td>0.62</td>
<td>0.69</td>
<td>0.82</td>
<td>0.81</td>
<td>—</td>
</tr>
<tr>
<td>Picon (repeat†)</td>
<td>UK</td>
<td>WHO</td>
<td>Metabolic syndrome</td>
<td>0.32</td>
<td>0.96</td>
<td>0.61</td>
<td>0.93</td>
<td>H</td>
</tr>
</tbody>
</table>

(Continued)
Table 1 | Details of diagnostic accuracy data of for detection of pre-diabetes and QUADAS analysis in studies included in review

<table>
<thead>
<tr>
<th>Country</th>
<th>Reference test criteria</th>
<th>Population information</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
<th>Index sensitivity</th>
<th>Index specificity</th>
<th>QUADAS bias*</th>
<th>QUADAS applicability*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mehmet 2010†</td>
<td>UK</td>
<td>WHO</td>
<td>0.63</td>
<td>0.84</td>
<td>0.86</td>
<td>0.88</td>
<td>U</td>
<td>I</td>
</tr>
<tr>
<td>Holt 2003†</td>
<td>UK</td>
<td>WHO</td>
<td>0.43</td>
<td>1</td>
<td>1</td>
<td>0.94</td>
<td>U</td>
<td>L</td>
</tr>
<tr>
<td>McClean 2010†</td>
<td>UK</td>
<td>WHO</td>
<td>0.51</td>
<td>0.92</td>
<td>0.73</td>
<td>0.94</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Kwong 2009†</td>
<td>Canada</td>
<td>WHO</td>
<td>0.2</td>
<td>0.87</td>
<td>0.05</td>
<td>0.97</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Reimblatt 2006†</td>
<td>Canada</td>
<td>WHO</td>
<td>0.84</td>
<td>0.61</td>
<td>0.14</td>
<td>0.98</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Agarwall 2003†</td>
<td>UAE</td>
<td>WHO</td>
<td>0.33</td>
<td>0.85</td>
<td>0.18</td>
<td>0.92</td>
<td>U</td>
<td>L</td>
</tr>
<tr>
<td>Costa 200FF†</td>
<td>Spain</td>
<td>WHO</td>
<td>0.25</td>
<td>0.89</td>
<td>0.08</td>
<td>0.97</td>
<td>U</td>
<td>L</td>
</tr>
<tr>
<td>Conway 1999†</td>
<td>US</td>
<td>ADA</td>
<td>0.45</td>
<td>0.88</td>
<td>0.36</td>
<td>0.92</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Mudaige 2014†</td>
<td>UK</td>
<td>WHO</td>
<td>0.48</td>
<td>0.92</td>
<td>0.35</td>
<td>0.95</td>
<td>U</td>
<td>L</td>
</tr>
<tr>
<td>Venkataraman 2014†</td>
<td>UK</td>
<td>WHO</td>
<td>0.22</td>
<td>0.93</td>
<td>0.14</td>
<td>0.96</td>
<td>U</td>
<td>L</td>
</tr>
<tr>
<td>Russell 2013†</td>
<td>UK</td>
<td>WHO</td>
<td>0.13</td>
<td>0.96</td>
<td>0.2</td>
<td>0.93</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Joseph 2013†</td>
<td>UK</td>
<td>WHO</td>
<td>0.38</td>
<td>0.87</td>
<td>0.45</td>
<td>0.93</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Ko 2001†</td>
<td>China</td>
<td>WHO</td>
<td>0.63</td>
<td>0.92</td>
<td>0.17</td>
<td>0.99</td>
<td>L</td>
<td>U</td>
</tr>
<tr>
<td>Kakad 2010†</td>
<td>UK</td>
<td>WHO</td>
<td>0.36</td>
<td>0.86</td>
<td>0.21</td>
<td>0.93</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Cypryk 2004†</td>
<td>Poland</td>
<td>WHO</td>
<td>0.75</td>
<td>0.83</td>
<td>0.12</td>
<td>0.99</td>
<td>U</td>
<td>L</td>
</tr>
</tbody>
</table>

ADA=American Diabetes Association; OGTT=oral glucose tolerance test; FPG=fasting plasma glucose; IGT=impaired glucose tolerance; GDM=gestational diabetes mellitus.

*U=unclear risk of bias; L=low risk of bias; H=high risk of bias.
†In table already, data values for that test combination specified.

Fig 3 ROC curve studies using FPG as index test and IGT as reference standard. Area of ellipse is proportional to prevalence.

Fig 2 ROC curve for studies using IGT as index test and OGTT as reference standard. Area of ellipse is proportional to prevalence.
Table 2 | Prevalence analysis of three tests used to identify people with pre-diabetes

<table>
<thead>
<tr>
<th>Author</th>
<th>Country</th>
<th>Population demographics</th>
<th>Diagnostic criteria for HbA1c</th>
<th>Diagnostic criteria for OGTT</th>
<th>Total population tested</th>
<th>% Tested population “pre-diabetic”</th>
<th>% Pre-diabetic population with abnormal HbA1c alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incani28</td>
<td>Italy</td>
<td>BMI &gt;35 or DM risk factors</td>
<td>ADA</td>
<td>ADA</td>
<td>1054</td>
<td>49</td>
<td>17</td>
</tr>
<tr>
<td>James75</td>
<td>USA</td>
<td>Mixed ethnicity. NHANES survey</td>
<td>ADA</td>
<td>ADA</td>
<td>5627</td>
<td>37</td>
<td>8</td>
</tr>
<tr>
<td>Zhang76</td>
<td>China</td>
<td>Chinese. GHS population survey</td>
<td>ADA</td>
<td>ADA</td>
<td>3590</td>
<td>66</td>
<td>29</td>
</tr>
<tr>
<td>Benages77</td>
<td>Spain</td>
<td>Mixed ethnicity. GDM</td>
<td>ADA</td>
<td>ADA</td>
<td>141</td>
<td>42</td>
<td>25</td>
</tr>
<tr>
<td>Mostafa78</td>
<td>UK</td>
<td>WE and SA ethnicity</td>
<td>WHO</td>
<td>IEC</td>
<td>8696</td>
<td>27</td>
<td>48</td>
</tr>
</tbody>
</table>

ADA=American Diabetes Association; OGTT=oral glucose tolerance test; WE=west European; SA=South Asian.

(95% confidence interval 23 to 67). Lifestyle interventions lasting three to six years showed a 37% (28% to 46%) reduction in relative risk, equating to 151 (129 to 172) out of 1000 people in the lifestyle intervention group developing diabetes compared with 239 of 1000 in the control group (NNT 12, 10 to 15). The overall relative risk reduction of developing diabetes after lifestyle interventions was 36%. Because of the small number of follow-up studies it is difficult to assess the reduction in risk of diabetes after the completion of lifestyle interventions. Our estimates show that relative risk reduction of developing diabetes fell to 20% (8% to 31%) in the period after the trial (fig 7).

Meta-analysis evaluating the impact of metformin (fig 8) showed a relative risk reduction of 26% (95% confidence interval 16% to 35%) while participants were taking this drug, translating to 218 (95% confidence interval 192 to 248) out of 1000 developing diabetes compared with 295 of 1000 not receiving this drug (NNT 14 (95% confidence interval 10 to 22)). The benefits of metformin were assessed at the end of the trial periods once the participants had been taking the drug for a prespecified length of time. There were no follow-up studies examining for persistence of benefit once metformin had been discontinued, but the US DPP study did show some improvements in reduction in incidence of diabetes with long term metformin use.130

The main sources of potential bias (as estimated by Cochane risk of bias tool) were selection bias (lack of allocation concealment) and attrition bias (where authors used per protocol analysis instead of an intention to treat analysis to assess changes in outcome measures), potentially leading to overestimation of the benefits of the intervention. To provide the most comprehensive synthesis of relevant studies we did not pre-specify a minimum threshold of methodological quality for included studies. However, we performed a sensitivity analysis removing the studies at high risk of bias to test whether the exclusions of some trials changed the overall findings. Omission of these did not significantly change the overall results (for example, removal of the 2006 study by Ramachandran and colleagues106 did not significantly alter the relative risk reduction).

Using the GRADE approach, we assessed the evidence to be of moderate quality for progression to type 2 diabetes with metformin versus control, low quality for lifestyle interventions of one to two years and three to six years’ duration versus control, and very low qual-
Fig 6 | Relative reduction in risk of diabetes at end of lifestyle trials. A=random sequence generation (selection bias); B=allocation concealment (selection bias); C=blinding of outcome assessment (detection bias); D= incomplete outcome data (attrition bias); E=selective reporting (reporting bias)

Fig 7 | Relative reduction in risk of diabetes at follow-up after intervention. A=random sequence generation (selection bias); B=allocation concealment (selection bias); C=blinding of outcome assessment (detection bias); D= incomplete outcome data (attrition bias); E=selective reporting (reporting bias)
Most interventions were inflexible, reducing four main findings. Firstly, the diagnostic accuracy of tests used to detect pre-diabetes in screening are of tests used to detect pre-diabetes in screening.

Risk (oral glucose tolerance test) is not widely used in practice. A further downgrade was because of the statistical heterogeneity in two out of the four outcomes (lifestyle interventions with a three to six year follow-up (I²=45%; downgraded once) and follow-up after intervention (I²=82%; downgraded twice)). This high degree of heterogeneity is probably because of differences in sample size and length and intensity of interventions included in this analysis, but the small number of trials contributing to the follow-up analysis after intervention limited our ability to explore this using subgroup analysis. Seven papers described at least one element of patient and participant involvement. Most interventions were inflexible, with a one-size-fits-all approach.

Gestational diabetes
Nine trials assessed lifestyle interventions in women with a history of gestational diabetes (see tables B and D, appendix 3). These focused on diet, exercise, and increased uptake of breast feeding. None showed a significant reduction in incidence of diabetes between the intervention and control groups. Attrition rates were high in these trials. Only three trials had sufficient data to be included in the meta-analysis.

Withdrawal and attrition rates
Sixteen studies had the necessary data available to assess withdrawal and attrition rates. Of the pre-diabetic population identified, only 27% went on to complete the trial (the rest were either not eligible, declined to participate, or withdrew from the intervention (fig 9)). Therefore, relative risk reductions calculated from intervention trials reflect risk improvements seen in a limited proportion of the total pre-diabetic population.

Discussion
Principal findings
This systematic review, commissioned by local policymakers who wanted to identify an effective “screen and treat” strategy for prevention of type 2 diabetes in an area of high prevalence, included 99 studies and produced four main findings. Firstly, the diagnostic accuracy of tests used to detect pre-diabetes in screening programmes is low. The most commonly used test (HbA₁c) is neither sensitive nor specific; the fasting glucose test is specific but not sensitive. Low sensitivity results in a high number of people with false negative results, resulting in a large number being falsely reassured. Secondly, the diagnostic tests identify different pre-diabetic population groups with limited overlap. If the American Diabetes Association criteria are used instead of the WHO ones, the prevalence of those with a diagnosis of pre-diabetes doubles. Thirdly, both individually targeted lifestyle interventions and metformin have some efficacy in preventing or delaying the onset of type 2 diabetes, though the protective effect of the former is greatest in longer interventions (three to six years) and attenuates with time from intervention. We have only moderate to very low confidence in these estimates, however, because study quality was often low. Finally, in women with a history of gestational diabetes, the evidence base for lifestyle interventions in preventing progression to type 2 diabetes is currently weak.

### Table: Risk ratio (random effects, M-H, 95% CI)

<table>
<thead>
<tr>
<th>Study</th>
<th>No of events/total</th>
<th>Risk ratio (random effects, M-H, 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li 1999</td>
<td>3/42 6/43</td>
<td>1 0.51 (0.14 to 1.91)</td>
</tr>
<tr>
<td>Lu 2011</td>
<td>0/50 8/104</td>
<td>&lt;1 0.12 (0.01 to 2.06)</td>
</tr>
<tr>
<td>Ramachandran 2006</td>
<td>54/133 75/136</td>
<td>25 0.74 (0.57 to 0.95)</td>
</tr>
<tr>
<td>US DPP 2002</td>
<td>233/1073 313/1082</td>
<td>74 0.75 (0.65 to 0.87)</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>290/1298 402/1365</td>
<td>100 0.74 (0.65 to 0.84)</td>
</tr>
</tbody>
</table>

Test for heterogeneity: χ²=0.00, df=3, P=0.59, I²=0%
Test for overall effect: z=4.62, P<0.001
Most intervention trials included in this study used the oral glucose tolerance test to identify their study population. In practice, however, this test is not widely used. It is time consuming, requires fasting and ingestion of a sugary drink (which many people find unpleasant), and, because of variability within an individual, needs to be done twice. HbA1c is estimated on a single non-fasting blood test but varies by ethnicity, leading to overestimation and underestimation of the result, and could be inaccurate in the presence of haemoglobinopathy. The fasting plasma glucose test is a single blood test but requires the person to have fasted for several hours so is impractical for mass screening.

Accuracy of tests depends on cut-off points. By using the International Expert Committee and WHO criteria for defining pre-diabetes, HbA1c correctly identifies only half the individuals with an abnormal result on an oral glucose tolerance test but also assigns the label of pre-diabetes to large numbers of individuals with a normal test result. Different diagnostic criteria result in a different estimate of the prevalence of pre-diabetes; this will have implications for which (and how many) individuals are eligible for lifestyle interventions. Furthermore, people identified with HbA1c might not have the same glycaemic abnormality as those entered into trials on the basis of an oral glucose tolerance test and might respond differently to interventions.

Systematic reviews assessing progression from at risk states to diabetes have shown that those at most risk of developing diabetes had both impaired fasting glucose and impaired glucose tolerance; HbA1c showed a lower progression rate, similar to impaired fasting glucose alone. Those with a history of gestational diabetes have the highest rates of progression to diabetes, with a sevenfold increased risk after the first diagnosis and a 70% cumulative incidence at 10 years.

Of the 50 intervention trials included in this review, 34 used surrogate endpoints (most commonly, weight loss) as their primary outcome. While most found significant changes in these endpoints, authors rarely commented critically on the sustainability or clinical importance of these. Weight reduction has been shown to correlate poorly with the reduction in incidence of diabetes in some populations. The trials in our sample that did show a significant reduction in the definitive endpoint of incidence of diabetes lasted between three and six years and were intensive in nature with individuals closely monitored.

While reduced incidence of diabetes seems to be possible if the interventions are intensive, the relative risk reductions seen in trials apply only to those who enrol and adhere to the intervention. Given the number of people who will not meet eligibility criteria or who decline or do not complete the intervention (fig 9), there is no scientific basis for extrapolating percentage risk reductions seen in trials to an equivalent reduction in incidence of diabetes across an entire community. Poor enrolment and completion of lifestyle interventions will limit the impact national prevention programmes will have on the overall burden of disease.

**Comparison with other systematic reviews**

This systematic review is the first to combine the analysis of diagnostic accuracy with efficacy of interventions to give an overall estimate of how screen and treat policies could play out in populations, focusing on the endpoint of progression to type 2 diabetes. Edwardson and colleagues reviewed the effectiveness of risk scores and lifestyle interventions but did not assess their accuracy and the implications of their use. Other systematic reviewers performed a more in-depth analysis of improvement in surrogate endpoints such as weight loss and improvements in glycaemic markers, A review carried out by the Institute for Clinical and Economic Review, however, raised concerns regarding the clinical importance and sustainability of improvements of these surrogate markers.

Other systematic reviews have found similar relative risk reductions in incidence of diabetes with lifestyle interventions and metformin in study populations. Previous meta-analyses showed a higher relative risk reduction when they included only the most tightly controlled trials with stringent population enrolment criteria. In contrast, Public Health England’s meta-analysis of translational studies identified a lower relative risk reduction because of the inclusion of pragmatic trials and observational studies and showed high levels of statistical heterogeneity between primary studies. One systematic review assessed UK based community and national interventions whose participants were the most deprived, vulnerable, and socially excluded (groups often omitted from clinical trials). They found that the effects of the interventions were small in these groups, with no evidence of long term reduction in incidence of diabetes.

Labelling people as having “pre-diabetes” has important personal implications (medicalisation, intrusive testing, and stigma) for people who might never go on to develop diabetes. Other scholars have voiced similar concerns to those raised in this systematic review with regards to the danger of inaccurate classification and/or overdiagnosis with tests for pre-diabetes, effectiveness of lifestyle interventions in the real world, and the limited impact of screen and treat policies in the absence of a complementary population based approach.

Whether these interventions reduce longer term cardiovascular morbidity and mortality remains unclear. A meta-analysis and systematic review undertaken by Hopper and colleagues agreed with our findings that lifestyle interventions can reduce the relative risk of developing diabetes. While these interventions did result in a reduction in incidence of cardiovascular events, this did not translate into a significant reduction in all cause or cardiovascular mortality. Long term follow-up studies undertaken by the Chinese Da Qixing study and the Finnish Diabetes Prevention Study found that there was no significant difference between intervention and control groups in first cardiovascular events or cardiovascular morbidity, though the study was not powered to detect such a difference.
Meaning and implications for policy makers, clinicians, and academics

This review was requested by a local clinical commissioning group in an inner London borough where the local diabetes prevention programme has largely consisted of a community prescription initiative offered to people classified as having pre-diabetes with a BMI of 27 or above, a history of gestational diabetes, or a QRISK >20%. Intensive interventions lasting years, such as those included in this systematic review, are not an option given its limited budget.

Our findings indicate that in settings such as this, screen and treat policies for pre-diabetes will benefit individuals who are “true positives” and have sufficient personal, family, and community resources to enable them to attend and comply with preventive interventions. Incentivised diabetes prevention programmes will also pick up people with undiagnosed diabetes (an estimated 210% of those screened[28,29,116,126]), who can be offered timely management. A considerable proportion of people at high risk of developing type 2 diabetes, however, will go on to develop the condition despite such programmes. These include people who test “false negative” and those who, despite testing positive and being offered a lifestyle intervention, lack the personal resources and social connections to support and sustain lifestyle change.

Because of the low accuracy of screening tests and the limited reach of intervention programmes, policymakers might want to consider supplementing screening and treat policies with population based approaches aimed at entire communities. WHO, for example, proposes “multisectoral action that simultaneously addresses different sectors that contribute to the production, distribution and marketing of food, while concurrently shaping an environment that facilitates and promotes adequate levels of physical activity.”[153]

Strengths and limitations

This is the first systematic review to assess both the diagnostic accuracy of screening tests for pre-diabetes and the efficacy of interventions in those classified through screening as having pre-diabetes. Furthermore, it is a comprehensive review synthesising a large volume of international literature, including translations from languages other than English. It was inspired by a question by front line policymakers and focused on producing a practical answer to that question. As such, and unlike much secondary and primary research, it fulfils the important criterion of “usefulness.”[154]

The main limitation of the review was the number of exclusions because of incomplete data available in published studies. Despite efforts to contact authors, we were unable to obtain the data needed to contribute to the meta-analysis in 18 potentially eligible papers. In the prevalence analysis, only five out of 28 papers compared all three diagnostic tests, so these findings should be interpreted with caution. A high proportion of studies that assessed the diagnostic accuracy of fasting plasma glucose did so in participants with a history of gestational diabetes—a bias that could influence the generalisability of this analysis.

Only half of the intervention studies were included in the meta-analysis because the lengths of the trial or intervention were too short to be able to capture incidence of diabetes. Additionally, the analysis of the reduction in the risk of diabetes at various follow-up periods was limited because of the small number of primary studies that performed follow-up analyses. We recommend that primary studies of diabetes prevention programmes should be resourced to undertake long term follow-up to assess for sustained benefits including incidence of diabetes, cardiovascular morbidity, and mortality.

Intervention studies that used risk scores to identify their population instead of blood tests do exist[155] but were outside the scope of this systematic review. Further synthesis of interventions using wider population eligibility criteria could provide additional insights into the benefits of these.

Future work

On the basis of the findings of this review, we suggest three avenues for further research. The first is pragmatic real world effectiveness and cost effectiveness studies of interventions for pre-diabetes that have already been shown to be efficacious in trials.[149,156] Studies of the translational gap between evidence from randomised trials and real world uptake and impact is always important[157] but particularly so when the “real world” seems unlikely to be able to replicate the conditions (for example, health literacy, language fluency, and comorbidities of target population; intensity and duration of intervention; completeness of follow-up) that characterised the trials with the most positive results.[149,158] These real world studies should deal with the impact on behaviour of individuals who test positive for pre-diabetes (only a third of whom would be predicted to engage with interventions; fig 9) and the costs (to both participants and the health service). More specifically, effectiveness and cost effectiveness studies should explore the implications of screen and treat programmes for both commissioners and providers—including the opportunity costs of spending a limited budget on a programme for which a variable proportion of the pre-diabetic population would be eligible for and engage with depending on locality.

The second avenue for further research is the evaluation of population level and/or health system interventions. Individual lifestyle choices are constructed by sociocultural, political, and economic influences, which might be more effectively dealt with by using population-wide strategies such as protection of green spaces, increased walkability of the environment, affordable leisure activities, improved food labelling, independent regulation of food nutritional standards, regulation on food advertising, affordable fruit and vegetables, and school based programmes. Such systematic structural approaches dealing with “upstream” influences on the pathogenesis of diabetes require well supported public health teams working alongside local
governments to improve the health of communities and could be vital components of a multifaceted long term primary prevention strategy.

Currently, only a tiny fraction of the literature on diabetes prevention is informed by an appreciation of the social complexity underlying pathogenesis of diabetes. The 2016 Foresight Report on Obesity was a model of good practice in teasing out the complex interactions between genetic, physiological, psychological, sociocultural, economic, and political determinants of obesity; it provided a strong and consistent message that short term “behaviour change” interventions were unlikely to succeed in isolation. A comparable initiative for type 2 diabetes could add richness to our current understanding of the condition and help to inform the design of evidence based strategies aimed at influencing its “upstream” determinants.

Conclusion

As the prevalence of type 2 diabetes rises inexorably in high, middle, and low income countries alike, controversy continues to surround the questions of who is “at risk” and what preventive interventions to offer them. A screen and treat policy will be effective only if a test exists that correctly identifies those at high risk (sensitivity) while also excluding those at low risk (specificity); and an intervention exists that is acceptable to, and also efficacious in, those at high risk. This review has shown that of the two screening tests for pre-diabetes that are available and acceptable to patients and clinicians, fasting glucose is specific but not sensitive and HbA1c is neither sensitive nor specific. Trial evidence suggests that lifestyle interventions have a potential role in reducing individual progression to diabetes and could benefit those high risk individuals who have the motivation and social support to achieve sustained lifestyle change. Given that this is likely to be a limited proportion of the population identified with pre-diabetes, however, substantial research resources should be directed at the evaluation of upstream interventions aimed at the entire population.

Contributors: EB conceptualised the review, assisted with developing the search strategy and ran the search, scanned all titles and abstracts, extracted quantitative data on all the papers, checked citations, performed the prevalence analysis, performed the meta-analysis of the intervention studies, undertook the QUADAS, risk of bias, and CONSORT assessment, and co-wrote and revised drafts of the paper. SR conceptualised the review, independently reviewed the data extraction process from the search results and methods from the intervention papers, adapted the QUADAS and risk of bias tool verifying the methods, and checked a sample of this assessment. JO advised on the analysis of the quantitative data and carried out the diagnostic accuracy bivariate meta-analysis. RN advised on the quality assessment of the literature and undertook the GRADE assessment. RN also reviewed drafts of the paper and assisted with graphically representing the risk of bias tool using RevMan. SV conceptualised the study, framed the question, and manages the project steering group. TG is the academic supervisor for the project, conceptualised the study, advised on systematic review methods, and co-wrote and revised drafts of the paper. TG is guarantor. All authors, external and internal, had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding: This study was funded by grants from the Newham Clinical Commissioning Group and University College London Partners, a National Institute for Health Research fellowship for EB, National Institute for Health Research senior investigator award for TG, and by internal funding for staff time from the Nuffield Department of Primary Care Health Sciences, University of Oxford. The funders had no input into the selection or analysis of data or the content of the final manuscript.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: No additional data available.

Transparency: The lead author (the manuscript’s guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

We thank Helen Ewell, BMA librarian, for help with the literature search, and Geoffrey Wong, Manja Cvetkovic, and Zoya Georgieva for help with translation of non-English papers. Thanks to the Newham Clinical Commissioning Group and University College Partners for their support of this project.


34 Vlak EM, Admiraal WM, Busschers WB, et al. Screening South Asians for type 2 diabetes and pre-diabetes: (1) comparing oral glucose tolerance and hemoglobin A1c tests results and (2) comparing the two sets of metabolic profiles of individuals diagnosed with these two tests. *BMJ Endocr Disord* 2013;13.8. doi:10.1136/1472-6823-13.8.


Appendix 1: Search strategies

Appendix 2: Adapted QUADAS2 Tool to assess quality of diagnostic accuracy studies

Appendix 3: Interventions to reduce the risk of developing diabetes methods and results tables

Appendix 4: Diagnostic accuracy meta-analysis and prevalence analysis supplementary results